

Ajuste de observaciones  
para topógrafos

Ajuste de observaciones para topógrafos

© Javier A. Múgica de Rivera, 2006–2015  
Todos los derechos reservados

Se permite libremente la copia de sus páginas  
en papel con fines no comerciales.

No se permite su almacenamiento,  
copia o distribución en soporte digital.

ISBN: 978-84-608-3691-9

Primera edición, 2006

Segunda edición, 2015

Impreso por *Publiberia*

Impreso en España – Printed in Spain

# Ajuste de observaciones para topógrafos

*Javier A. Música de Rivera*

SEGUNDA EDICIÓN  
2015



# Índice

Prólogo . . . . .	XIII
1. Linealización . . . . .	1
Una ecuación	1
Un sistema de ecuaciones	4
2. Errores aleatorios . . . . .	7
Una variable. Funciones de densidad y de distribución	7
Dos variables independientes	12
Correlación entre dos variables	14
3. Estimación . . . . .	21
Concepto de estimador. Valores reales, observados y ajustados. Residuos	21
Medida de los tres ángulos de un triángulo	23
Función $g$ de estimación de los valores reales	26
Más ejemplos de función $g$	28
Proceso general para la obtención de la función $g$	30

Estimación de una magnitud de tipo escala	32
Distribución de probabilidad a priori de los valores reales	35
Intervalos de confianza	37
Apéndice al capítulo tercero	38
4. Algunas distribuciones . . . . .	41
5. Estimadores . . . . .	45
Distribuciones uniforme y triangular	45
Estimadores de máxima probabilidad	47
Estimadores centrados	48
Magnitudes «grandes» y «pequeñas»	50
Parametrización de las magnitudes observadas	51
Estimador de máxima probabilidad para variables independientes. Ejemplos	52
6. El teorema central del límite . . . . .	55
7. El estimador mínimo cuadrático . . . . .	59
Hipótesis de aplicación	59
Planteamiento y linealización de las ecuaciones de residuo	60
Coeficientes de A y coeficientes de L	64
Ecuaciones normales y resolución	65
8. Cuadrilátero . . . . .	67
Lados y diagonales	67
Ángulos	68
Ángulos y lados. Pesos	69

# Índice

VII

Observaciones no independientes	72
Método de las ecuaciones de condición	74
Caracterización $A^T V = 0$ de los valores ajustados	78
Sobre los métodos paramétrico y de condición	79
9. Ángulos y distancias . . . . .	81
Ecuación de distancia	81
Ecuación de lectura	82
Coordenadas fijas para definir el sistema. Ecuaciones de azimut y coordenada	83
Un ejemplo	84
Suma de los residuos de lectura de una estación	87
Resumen de fórmulas	88
10. Regiones de confianza . . . . .	91
Transmisión de varianzas	91
Función normal multidimensional	92
Función $g$ de los parámetros. Matriz $\Sigma_{xx}$	95
Elipses de confianza	97
Elipsoides de confianza	100
Función $g$ de las magnitudes observadas. Matriz $\Sigma_{ll}$	101
Función $f$ de los valores ajustados de los parámetros. Ma- triz $\Sigma_{xx}$	102
Función $f$ de los valores ajustados de los parámetros. (bis)	102
Función $f$ de los valores ajustados de las observaciones. Matriz $\Sigma_{ll}$	103
Reflexión sobre las distribuciones directas e inversas	104

Función $g$ de los errores. Matriz $\Sigma_{\varepsilon\varepsilon}$	104
Función $f$ de los residuos. Matriz $\Sigma_{vv}$	105
Análisis de $\Sigma_{ll} + \Sigma_{vv} = \Sigma_{\ell\ell}$	105
Esperanza del valor ajustado $l$ en función del error cometido	106
Diseño de las observaciones	108
Redundancias parciales	109
Correlaciones	112
<b>11. Altimetría y otros . . . . .</b>	<b>117</b>
Ajuste altimétrico	117
Ecuaciones de distancia geométrica y ángulo cenital	118
Error de centrado	119
Estaciones excéntricas	120
Alturas de instrumento y de mira	121
Error de refracción	122
Error de verticalidad de la estación	122
Parámetros adicionales: precisiones y redundancias	123
<b>12. Transformación de semejanza plana . . . . .</b>	<b>125</b>
Notación y convenciones generales	125
Convenciones para los parámetros de semejanza	126
Planteamiento mínimo cuadrático	127
Traslaciones	128
Escala y giro: planteamiento	129
Varianza única combinada para cada punto	130
Primeras fórmulas para escala y giro	132

Fórmulas simétricas para escala y giro	133
Análisis de las fórmulas obtenidas	134
Cómo promediar valores angulares	136
Giro: fórmula definitiva	137
Solución completa	138
Interpretación del parámetro $T$ y su precisión	139
Precisión de los parámetros de la transformación	141
Precisión del transformado de un punto	143
13. Otras transformaciones de coordenadas . . . . .	145
Semejanza espacial. Rotación	145
Cálculo de los parámetros de la rotación	146
Construcción de una matriz $R_{\Delta}$ exacta	149
Descomposición planimétrica / altimétrica	150
Precisiones	151
Afinidad	151
Descomposición de una afinidad	153
Solución rigurosa en casos extremos	153
Transformaciones en general	154
14. Cambio de los parámetros fijos, I: Puntual . . . . .	157
Precisión respecto a un punto	157
Dos puntos fijos	159
Un punto y orientación	166
Posición libre	172
Elección de los grados de libertad a aplicar	172

15. Cambio de los parámetros fijos, II: Global . . . . .	175
Introducción	175
Traslación promedio	176
Escala promedio	177
Giro promedio	180
Combinación de las distintas libertades	181
Interpretaciones de las transformaciones anteriores	182
Resumen de fórmulas	185
Demostración de óptimo para la traslación	187
Independencia de las tres optimizaciones	191
Principio de demostración de óptimo para escala y giro	193
Demostración de óptimo general	195
Extensión a otras transformaciones	201
16. Estimación de la varianza . . . . .	203
Introducción	203
Función $g(\sigma_0^2)$	204
Estimador de máxima probabilidad y centrado	207
Intervalos de confianza para la desviación típica	208
Estimación centrada geométrica	209
Precisión en la estimación de $\sigma_0$	210
Modificación de las regiones de confianza para las magnitudes estimadas	210
Combinación de varias estimaciones de la varianza	213
Aclaración sobre la estimación $\sigma_0^2 = \sum v/r$	214
Estimación de varias varianzas en un ajuste	215

17. Atajos . . . . .	217
Agrupación exacta de residuos, I. Transformación de coordenadas	217
II. Circunferencia	219
Agrupación modificada de residuos, I. Circunferencia	220
II. Coplanaridad	221
Falsa linealización, I. Transformación proyectiva	222
II. Intersección directa fotogramétrica	224
Linealización forzada, I. Rotación 3D	225
II. Coplanaridad	228
18. Interpretación geométrica . . . . .	233
La solución mínimo cuadrática	233
Funciones $g(E)$ , $g(X_r)$ y $g(L_r)$	236
Contribución de cada error a la solución ajustada	237
Matrices $\Sigma_{ll}$ y $\Sigma_{vv}$	239
Redundancias parciales	239
Óptimos para la definición global de traslación, escala, etc.	241
19. Estimadores para otras distribuciones, I . . . . .	245
Obtención de valores siguiendo la distribución de los errores	245
El parámetro $\tau$	248
Agrupación de los datos en intervalos	248
Ajuste de una función de distribución	250
Función $g$ y estimador de máxima probabilidad	253
Funciones $\psi'$	255

Solución alternativa. Función de peso	256
Estudio de la función $\psi$	257
Descentralidad de la estimación de máxima probabilidad	261
Diseño de un estimador más centrado	264
Simulación del caso directo	267
Definición directa de la matriz de pesos	269
20. Estimadores para otras distribuciones, II . . . . .	271
Regiones de confianza. Introducción	271
Obtención de $T_{xx}$ . Función $p_\tau$	271
Distribución a emplear para las regiones de confianza	274
Diseño de la función $p_\tau$	275
Casos límite para $p_\tau$	278
Procedimiento para la estimación de $\tau_0$	279
Medida de la dispersión para comparar $\mathcal{F}_{vr}$ y $\mathcal{U}$	281
Cálculo de la constante de centrado $k_c$	284
Sobre el carácter centrado de la estimación de $\tau_0$	286
Estimación conjunta de $\tau_0$ y los parámetros	287
Cuánto se cuenta cada residuo	287
Obtención de la distribución de los errores	291
Agrupación de observaciones ligadas	291
Aplicación previa de otra función de pesos para obtener valores iniciales	292
Apéndice: <i>Sobre el método de los mínimos cuadrados</i> . . . . .	295
Índice de símbolos . . . . .	299

## Prólogo

Cuando siendo estudiante de topografía alguna vez quise consultar alguna cuestión sobre el ajuste de observaciones en libros que trataban del tema, encontré grandes dificultades para entender lo que en ellos se explicaba. Parcos en explicaciones, los conceptos se escondían detrás de los símbolos, dando por supuesta una habilidad en su manejo y unos conocimientos de los que no disponía, y que no son los habituales en estudiantes de topografía y aun en geodestas. A buen seguro leyendo uno de los libros desde su comienzo y dedicando tiempo se pueda ir comprendiendo todo su contenido, pero tal bagaje de conocimiento estadístico y teoría matemática desborda por completo lo necesario para labores topográficas y técnicas afines. Fue esta circunstancia la que un poco más tarde, cuando a base de práctica conseguí ir entendiendo los conceptos que con los libros no había conseguido aclarar, me animó a escribir un libro pensado para topógrafos. Hace ya algunos años de aquello, y ahora que por fin he encontrado tiempo he aquí el libro. Este no es un libro de estadística, y he procurado no incluir teoría innecesaria. Tal vez personas con conocimiento en la materia echen de menos un capítulo puro acerca de estadística de carácter introductorio, pero dicha ausencia es intencionada; por una parte supongo que quien lee este libro ya sabe lo que es una media y tiene alguna idea de lo que es una varianza, y por otra parte pocos más que esos son los conocimientos necesarios, y si alguna vez surge alguno en su momento se introduce. Sí he querido sin embargo escribir un libro en el que todas las fórmulas se deducen, ningún concepto o propiedad aparece de la nada, y conceptos y

fórmulas se relacionan y surgen unos de otros. Probablemente el estudiante de topografía que se enfrenta por primera vez al ajuste de observaciones encuentre que en algunas partes el libro se vuelve demasiado arduo y cuesta ir comprendiendo cada paso. Creo que esto es inevitable en la correcta explicación de algunos conceptos, como la obtención de intervalos de confianza o la estimación de las varianzas. No obstante no es necesaria la comprensión de todo lo anterior para poder entender lo siguiente, y así la parte final del capítulo tercero y los capítulos 10, 14 y 15 pueden quedar vagamente entendidos en una primera lectura. Considero sin embargo que una vez pasado el tiempo y asentados los conceptos más sencillos se debe volver sobre las partes más complicadas para terminar de comprenderlo todo, a excepción tal vez de los tres últimos capítulos, algunas partes de los cuales el topógrafo puede no necesitar nunca.

Inicialmente pretendía incluir muchos más ejemplos de estimación mínimo cuadrática, pero el libro resultó tener mayor extensión de lo que en un principio esperaba, y consideré más adecuado escribir a parte un libro de ejercicios. Se han omitido algunas demostraciones, bien porque a mi entender revisten únicamente interés matemático, muchas consistentes únicamente en el manejo de expresiones, o bien porque tienen una dificultad que consideré excesiva. Para el lector que quiera conocer y entender todos los pasos que se han dejado sin demostrar se han incluido esas demostraciones como problemas teóricos en el libro de ejercicios.

*Ribadeo, junio de 2006*

# 1 Linealización

## Una ecuación

En la resolución de problemas de ingeniería aparecen a menudo ecuaciones cuya solución no se puede expresar como combinación finita de funciones elementales evaluadas en números enteros, o dicho coloquialmente, no se puede dar su valor exacto. No es necesario buscar ecuaciones complicadas. El siguiente ejemplo es uno de esos casos:

$$x - \sin x = 1.$$

Por otra parte, tampoco estamos interesados en el valor exacto, porque en última instancia el resultado de la ecuación dependerá de unos datos obtenidos con una cierta precisión, y a partir de un cierto decimal las cifras del resultado son erróneas. Habrá que calcular la solución con suficientes cifras decimales para no comprometer la precisión y ninguna más, pues ya no serían significativas.

Así como expresar el valor exacto es imposible, calcular un valor aproximado es sencillo, no hay más que ir dando valores a la función. Por convenio escribiremos siempre las ecuaciones de la forma  $f(x) = 0$ . En el ejemplo anterior,

$$x - \sin x - 1 = 0.$$

En este ejemplo, si  $x = 1,8$  tenemos que  $f(x) \approx -0,17$ , y si  $x = 2$  entonces  $f(x) \approx 0,09$ . La solución está pues entre 1,8 y 2. Podemos probar  $x = 1,9$ . Si el resultado es positivo entonces la solución estará

## 1 Linealización

entre 1,8 y 1,9, y si es negativo entre 1,9 y 2. Repitiendo este proceso unas cuantas veces obtenemos:

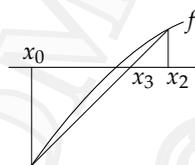
$$\begin{aligned}f(1,9) &< 0, \\f(1,95) &> 0, \\f(1,925) &< 0, \\f(1,9375) &> 0, \\f(1,93125) &< 0.\end{aligned}$$

Tomando la media de los últimos valores de  $x$  para los cuales  $f$  toma signos contrarios (que en este caso son los dos últimos) obtenemos  $x \approx 1,934375$ .

Este método no es malo. En cada iteración se mejora a la mitad la precisión, por término medio, lo que significa que cada 10 iteraciones se obtienen 3 decimales más. Pero se puede mejorar: Sea  $x_0 = 1,9375$ ,  $x_1 = 1,93125$  y  $x_2 = 1,934375$ .  $f(x_2)$  es menor que cero, por lo que el siguiente valor,  $x_3$ , sería  $(x_0 + x_2)/2$ . Pero si atendemos al valor de  $f$  y no sólo a su signo vemos que

$$\begin{aligned}f(1,9375) &\approx 0,0040, \\f(1,934375) &\approx -0,00025.\end{aligned}$$

Por lo que podemos suponer que  $x$  estará mucho más próximo a  $x_2$  que a  $x_0$ . La idea es calcular  $x_3$  como se muestra en la gráfica:



El cálculo es sencillo y el lector debería ser capaz de encontrarlo. Se obtiene  $x_3 \approx 1,934563$ , y  $f(x_3) \approx -2,6 \cdot 10^{-7}$ .

El procedimiento se puede refinar para mejorar lo más posible la convergencia, dando lugar a un método muy bueno, pero que tiene el inconveniente de que no se puede extender fácilmente a sistemas de muchas ecuaciones con otras tantas incógnitas.

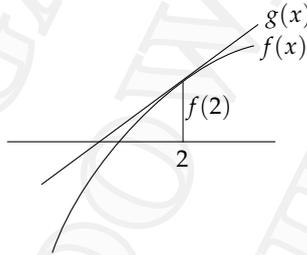
## 1 Linealización

Otra idea consiste en tomar la recta tangente a la función en un punto  $x_0$  próximo a la solución. Siguiendo con el ejemplo anterior,

$$f'(x) = 1 - \cos x;$$

y tomando como punto cercano a la solución  $x_0 = 2$  se obtiene la recta tangente

$$\begin{aligned}g(x) &= f'(2)(x - 2) + f(2), \\g(x) &\approx 1,416(x - 2) + 0,0907, \\g(\Delta x) &\approx 1,416\Delta x + 0,0907.\end{aligned}$$



Se resuelve la ecuación  $g(\Delta x) = 0$ , que es inmediata, y el valor  $x_0 + \Delta x$  se toma como nuevo valor aproximado. En el ejemplo que estamos siguiendo se obtienen sucesivamente los siguientes valores:

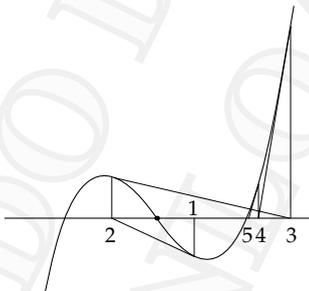
$$\begin{aligned}2, \\1,9360, \\1,934563874, \\1,934563211.\end{aligned}$$

Con este método, cuando se está cerca de la solución, en cada iteración se obtienen *grosso modo* el doble de decimales que en la anterior.

Éste es el método de linealización. A diferencia del anterior puede que no se llegue a ninguna solución, o que se llegue a una pero no sea la que deseamos, como se muestra en el siguiente ejemplo, en el que se parte de un punto demasiado alejado de la solución que se pretende obtener (el punto grueso) y finalmente se obtiene otro valor.

Sin embargo en la práctica esto no supone un problema, ya que rara vez se parte de un punto tan alejado de la solución.

## 1 Linealización



### Un sistema de ecuaciones

La extensión del método de linealización a más dimensiones es sencilla. Con dos variables todavía se puede visualizar. En un sistema de dos ecuaciones con dos incógnitas cada ecuación representa una superficie, que corta el plano  $xy$  en una línea, curva en general. La solución del sistema es el punto de corte de las dos superficies y el plano  $xy$ . La linealización consiste en tomar en el punto aproximado a la solución el plano tangente a cada superficie, y el punto de corte de estos planos con  $xy$  es el nuevo valor aproximado.

El procedimiento con un número cualquiera de variables es el siguiente: Sea un sistema

$$f_1(x_1, x_2, \dots, x_n) = 0$$

$$f_2(x_1, x_2, \dots, x_n) = 0$$

$$\vdots$$

$$f_n(x_1, x_2, \dots, x_n) = 0$$

de  $n$  ecuaciones con  $n$  incógnitas.

Sustituimos las funciones  $f$  por funciones lineales tangentes a ellas (que coinciden en el valor de la función y en el de la primera derivada respecto de cada una de las variables) en un punto cercano a la solución  $(x_1, x_2, \dots, x_n)_0$ :

## 1 Linealización

$$(f_1)_0 + \left(\frac{\partial f_1}{\partial x_1}\right)_0 \Delta x_1 + \left(\frac{\partial f_1}{\partial x_2}\right)_0 \Delta x_2 + \cdots + \left(\frac{\partial f_1}{\partial x_n}\right)_0 \Delta x_n = 0$$

$$(f_2)_0 + \left(\frac{\partial f_2}{\partial x_1}\right)_0 \Delta x_1 + \left(\frac{\partial f_2}{\partial x_2}\right)_0 \Delta x_2 + \cdots + \left(\frac{\partial f_2}{\partial x_n}\right)_0 \Delta x_n = 0$$

⋮

$$(f_n)_0 + \left(\frac{\partial f_n}{\partial x_1}\right)_0 \Delta x_1 + \left(\frac{\partial f_n}{\partial x_2}\right)_0 \Delta x_2 + \cdots + \left(\frac{\partial f_n}{\partial x_n}\right)_0 \Delta x_n = 0$$

El subíndice 0 en la función y en las derivadas indica que se evalúan en el punto aproximado. Por ejemplo,  $(\partial f_1/\partial x_1)_0$  es el valor de la función  $\partial f_1/\partial x_1$  en el punto  $(x_1, x_2, \dots, x_n)_0$ , y por lo tanto un número.

$\Delta x_1$  es el incremento de la variable  $x_1$  respecto al punto aproximado, y lo mismo para  $x_2, \dots, x_n$ . Se obtiene pues un sistema de  $n$  ecuaciones lineales en las incógnitas  $\Delta x_1, \Delta x_2, \dots, \Delta x_n$ .

El sistema se resuelve y el punto solución se toma como nuevo punto de linealización, y se itera el proceso. Salvo que empecemos muy lejos de la solución será suficiente con una o dos iteraciones, ya que la convergencia es muy rápida.



## 2 Errores aleatorios

### Una variable. Funciones de densidad y de distribución

Una medida es un proceso cuyo resultado final es un valor, que en nuestro caso será siempre un número. También por medida nos referiremos al número, propiamente el resultado de la medida, cuando no haya lugar a confusión. La palabra observación se emplea como sinónimo de medida.

El concepto de medida hay que entenderlo tal como lo hemos definido, en toda su amplitud. Así por ejemplo, una medida es la superficie de una parcela, la intensidad de corriente en un circuito, la temperatura del aire... pero también resultados complejos que dependen a su vez de otras medidas, como un coeficiente de redondez, la temperatura estimada en el centro de la Tierra, la edad de un accidente geológico...

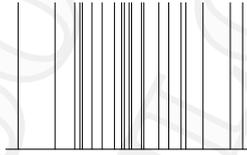
Una medida siempre lleva aparejado un error, que es la diferencia entre el valor obtenido y el valor real. Los valores reales, a menudo desconocidos, los denotaremos con letras como las siguientes:  $abl\alpha$ , mientras que para los valores observados emplearemos el tipo de letra normal para las variables (salvo la  $l$ , que será  $\ell$  para distinguirla mejor):  $abl\alpha$ . El error es un valor real, desconocido, que representamos con la letra  $\varepsilon$ :

$$\varepsilon = l - \ell.$$

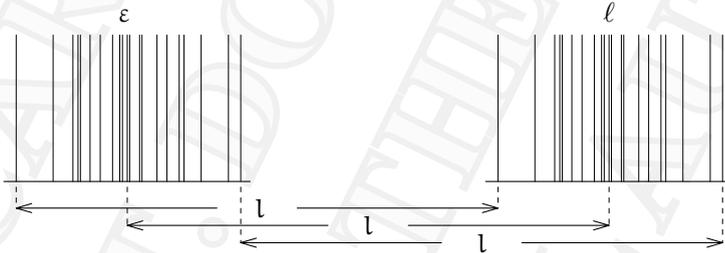
Si repetimos una medida varias veces obtendremos diferentes valores:  $\ell_1, \ell_2, \dots$ . En la siguiente gráfica aparece representado un ejemplo

## 2 Errores aleatorios

de forma esquemática. Cada valor se corresponde con una raya vertical.



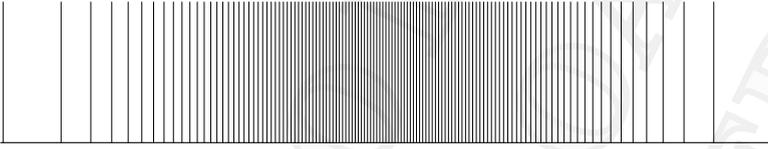
Los errores son desconocidos, pero se diferencian de los valores observados en una cantidad constante; es decir, si un valor  $\ell_1$  es 5 unidades mayor que otro  $\ell_2$ , su correspondiente error  $\varepsilon_1$  será también 5 unidades mayor que  $\varepsilon_2$ . De modo que la distribución de los errores es la misma que la de los valores observados desplazada una cantidad constante: el valor  $\ell$  desconocido.



Si repetimos la medida más y más veces, y suponiendo que fuésemos capaces de realizar la medida siempre en las mismas condiciones (condiciones ambientales, estado del aparato, estado del observador...), la distribución de los valores observados se aproximaría cada vez más a una cierta distribución  $f$ . La figura 2.3 muestra un ejemplo de esto.

La función  $f$  recibe el nombre de función de densidad. Antes de definirla de manera precisa conviene llamar la atención sobre una cuestión: El ejemplo anterior se obtuvo partiendo de la distribución  $f$  y generando de manera aleatoria 120 observaciones. Es muy común pensar que, si las observaciones siguen la distribución  $f$ , entonces una muestra de 120 observaciones tendrá el siguiente aspecto o uno muy parecido.

## 2 Errores aleatorios



Que los errores sigan una cierta distribución no significa que las medidas se vayan ajustando a esa distribución de manera casi perfecta; simplemente, que si se toman *muchas* observaciones la forma de la distribución obtenida se va aproximando a la distribución teórica. Para obtener algo parecido a la última gráfica habría que tomar realmente muchas observaciones.

Una vez aclarado este aspecto definimos ya la función de densidad. Su definición se hace de manera indirecta, estableciendo una propiedad que debe cumplir: Sean  $a, b$  dos valores, que pueden ser números reales o  $\pm\infty$ ; entonces

$$\int_a^b f(t) dt = P(a \leq \ell \leq b), \quad (2.1)$$

en donde el segundo miembro significa la probabilidad de que la observación  $\ell$  obtenida esté en el intervalo  $[a, b]$ .

Es decir, que el área bajo la curva  $f$  en un intervalo es la probabilidad de que un valor generado aleatoriamente siguiendo la distribución supuesta pertenezca a dicho intervalo.

De esta definición de la función de densidad se obtiene una pro-



Fig. 2.3

## 2 Errores aleatorios

propiedad muy importante:

$$\int_{-\infty}^{+\infty} f(t) dt = 1. \quad (2.2)$$

Y finalmente la función  $f$  es

$$f(x) = \lim_{h \rightarrow 0} \frac{\int_{x-h/2}^{x+h/2} f(t) dt}{h}, \quad (2.3)$$

cuando este límite exista.

La definición de la función  $f$  entraña el cálculo de un límite, pero más que el valor  $f(x)$  lo que realmente importa es la propiedad (2.1).

Más sencilla es la definición de la función de distribución,  $F$ :

$$F(x) = P(-\infty < \ell \leq x). \quad (2.4)$$

Por lo tanto,

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Y además  $F$  es creciente:  $b > a \Rightarrow F(b) \geq F(a)$ .

De (2.1) y (2.4) tenemos que

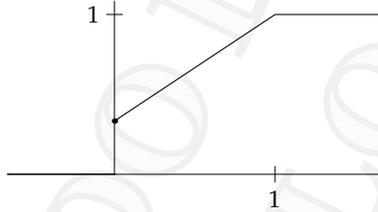
$$F(x) = \int_{-\infty}^x f(t) dt; \quad (2.5)$$

es decir, que la función de distribución es una integral de la función densidad  $o$ , lo que es lo mismo, que la función de densidad es la derivada de la función de distribución.

A parte de casos patológicos en los que el límite (2.3) pueda no existir, hay casos sencillos en los que puede valer  $\infty$ . Esto sucede cuando en un punto concreto se acumula una probabilidad  $p > 0$ .

Supongamos por ejemplo la variable aleatoria: «tiempo de espera en un semáforo». Supongamos además que el semáforo se encuentra en verde  $1/3$  del tiempo y  $2/3$  en rojo. Entonces existe una probabilidad  $p = 1/3$  de que el tiempo de espera sea exactamente 0. Si el tiempo que permanece en rojo es la unidad, entonces la función  $F$  es la de la figura 2.5.

## 2 Errores aleatorios



Si  $x < 0$  entonces  $F(x) = 0$ , ya que el tiempo de espera no puede ser negativo. En cambio,  $F(0) = P(t \leq 0) = P(t = 0) = 1/3$ . A partir de  $x = 0$  la probabilidad de esperar un tiempo  $t \leq x$  aumenta de manera continua hasta  $x = 1$ , y a partir de ahí  $F(x) = 1$ , ya que  $P(t \leq 1) = 1$ .

Esta distribución tiene una densidad infinita en  $x = 0$ , resultado de una probabilidad  $p = 1/3 > 0$  acumulada en un único punto. Si existen varios puntos con una probabilidad mayor que cero la función  $f$  valdrá infinito en todos ellos, y si sólo conocemos esa función no podemos saber cuál es la probabilidad acumulada en cada uno de los puntos. Sin embargo la función  $F$  sí que sigue siendo válida, y se sigue cumpliendo  $P(a < x \leq b) = F(b) - F(a)$ .

Aun en el caso de que  $F$  sea continua puede haber puntos en los que  $f$  sea infinita. Recordemos que  $f$  es la derivada de  $F$ , y una función puede ser continua en un punto y tener derivada infinita.

Si  $F$  es continua en un punto entonces la probabilidad acumulada en ese punto es 0. La mayoría de los errores de observación siguen distribuciones continuas (desconocidas casi siempre, pero continuas), y por lo tanto la probabilidad de que el error cometido sea un valor concreto  $a$  es exactamente cero.

Es habitual escuchar al respecto que «bueno, eso es la teoría; pero en la práctica sabemos que la probabilidad de que el error sea un valor concreto, por ejemplo 3, no es cero. Existe una probabilidad, aunque sea muy pequeña, de que el error sea exactamente 3». Esta falacia se debe a un error en el concepto de «exactamente». Se está pensando en los valores con un número finito de decimales, no sólo los valores observados, sino también los reales, lo cual es imposible. Se puede argumentar, refiriéndonos por ejemplo a medidas de distan-

## 2 Errores aleatorios

cias en topografía, que nunca se considerarán valores con decimales más allá de la centésima de milímetro, incluso los valores reales, y que entonces sí existe una probabilidad, muy pequeña pero mayor que cero al fin y al cabo, de que el error sea exactamente 3. Ciertamente, pero eso no contradice la teoría, ya que si sólo consideramos posibles un conjunto discreto de valores, entonces la distribución está formada por muchos puntitos con una probabilidad muy pequeña acumulada en cada uno de ellos, y en particular habrá un probabilidad mayor que cero en el valor 3.

Esta disertación es importante para resaltar que no se debe achacar cualquier diferencia entre un modelo teórico y la realidad a que la teoría no se ajusta exactamente a la práctica y en algún punto es preciso saltársela. Las frases de tipo «Bueno, pero en la práctica...» suelen esconder una debilidad en la comprensión del modelo teórico, una excusa para evitar buscar un modelo que se ajuste mejor a la realidad que el que se está empleando, o, muy frecuentemente, un error previo en la aplicación del modelo que más adelante se pone de manifiesto con un resultado imposible, y que es entonces cuando se echa la culpa al modelo. Toda la teoría de la estimación, toda, se basa en última instancia en la suposición de que los valores observados siguen una cierta distribución, y la teoría que se deriva de ella (de la distribución) es, en términos matemáticos, exacta. El problema radica pues en conocer la distribución de las observaciones, asunto harto difícil, y que es el único punto en el que el modelo se separa de la realidad, aunque no suele ser ésta la causa de resultados absurdos sino una incorrecta aplicación del modelo.

### Dos variables independientes

El paso siguiente al estudio de la distribución de una medida es el caso en el que son dos las magnitudes a medir, de valores reales  $l$  y  $m$ . Suponemos que las observaciones se realizan a pares, es decir, que por cada medida de  $l$  se obtiene una de  $m$ , generando una pareja de valores  $(l, m)$ . Repitiendo el par de medidas muchas veces se obtendrían los pares  $(l, m)_1, (l, m)_2, \dots$  que podemos representar sobre un plano (fig. 2.6).

Este ejemplo representa un caso en el que la variable  $m$  sigue una distribución con una densidad constante entre dos valores extremos,

## 2 Errores aleatorios

y la variable  $\ell$  la distribución  $f$  de unas páginas atrás. Se ha representado los errores en lugar de los valores observados. La gráfica de estos últimos sería idéntica pero desplazada una cantidad  $\ell$  en el eje  $\ell$  y  $m$  en el eje  $m$ . Si las dos variables siguen la distribución  $f$  la gráfica es la de la figura 2.7.

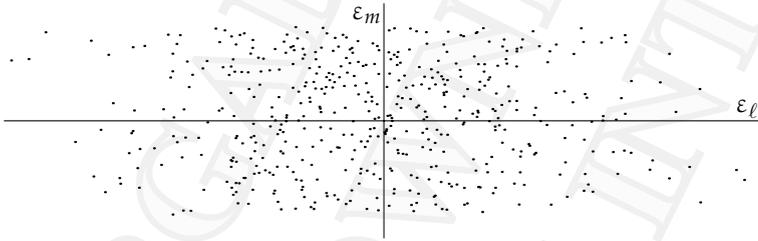


Fig. 2.6

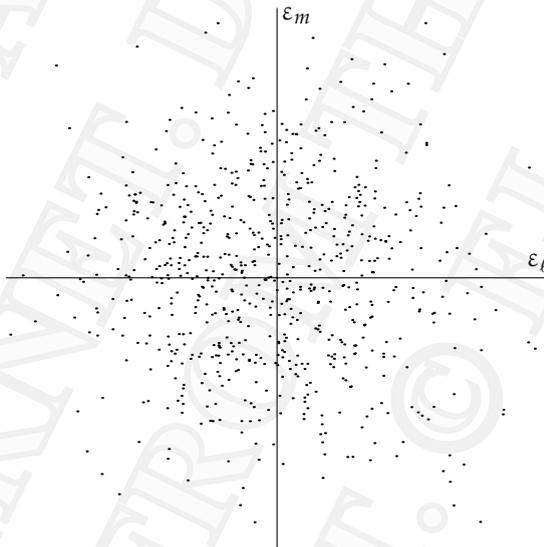


Fig. 2.7

## 2 Errores aleatorios

La función de densidad en dos dimensiones se define de manera análoga a como se hizo en una dimensión: La integral de la función densidad en un área es la probabilidad de que la pareja  $(l, m)$  se encuentre en dicha área.

La función de distribución es

$$F(x, y) = P(l \leq x, m \leq y).$$

La relación entre las funciones de distribución y densidad es más complicada que con una única variable.

### Correlación entre dos variables

En los dos ejemplos anteriores se ha supuesto que las variables  $\ell$  y  $m$  son independientes. En esos casos el hecho de considerar el par  $(\ell, m)$  no aporta más información que cada una de las variables por separado. Sin embargo, a veces los errores en las dos medidas no son independientes, porque en ellas interviene algún elemento común a ambas. Se dice que están correladas. La correlación puede ser positiva o negativa.

Supongamos por ejemplo que se miden simultáneamente, o con poca diferencia entre las medidas, una distancia  $d$  y un tiempo  $t$ . Supongamos que en los resultados que proporcionan los aparatos de medida influye la temperatura, como suele suceder, y supongamos por último que en ambos casos a mayor temperatura mayor valor observado. En las medidas de  $d$  y  $t$  se producirán unos errores, y en consecuencia se obtendrán  $d \neq d$  y  $t \neq t$ . Sea  $T_0$  la temperatura a la que ambos aparatos dan una medida correcta en ausencia de otros errores, y  $T_0 + \Delta T$  la temperatura a la que se realizaron las medidas. Entonces en las medidas de  $d$  y  $t$  se producen unos errores debidos a la temperatura  $\varepsilon_{d_1}$  y  $\varepsilon_{t_1}$  respectivamente, con la propiedad de que cuanto mayor sea  $\varepsilon_{d_1}$  mayor es  $\varepsilon_{t_1}$  y viceversa. Sean  $\varepsilon_{d_0}$  y  $\varepsilon_{t_0}$  los errores debidos a los demás factores, que suponemos independientes del error de temperatura e independientes entre sí, es decir, que no existen más causas comunes de error que la temperatura. Se cumple

$$\begin{aligned} \varepsilon_d &= \varepsilon_{d_1} + \varepsilon_{d_0}, & \varepsilon_t &= \varepsilon_{t_1} + \varepsilon_{t_0}, \\ d &= d + \varepsilon_{d_1} + \varepsilon_{d_0}, & t &= t + \varepsilon_{t_1} + \varepsilon_{t_0}. \end{aligned}$$

## 2 Errores aleatorios

$\varepsilon_{d_0}$  podrá ser en valor absoluto mayor o menor que  $\varepsilon_{d_1}$ , por lo que el signo de  $\varepsilon_d$  no será necesariamente el mismo que el de  $\varepsilon_{d_1}$ , y análogamente para  $\varepsilon_t$ . Sin embargo, sea un conjunto (grande) de observaciones de  $d$  en las que  $\varepsilon_{d_1}$  tiene un valor  $a$ , y otro conjunto en el que  $\varepsilon_{d_1} = b < a$ ; entonces en el primer conjunto la proporción de casos en los que  $\varepsilon_d$  es positivo será lógicamente mayor que en el segundo. Expresado en términos de probabilidad diríamos que cuanto mayor es  $\varepsilon_{d_1}$  mayor es la probabilidad de que  $\varepsilon_d$  sea positivo. En términos más generales: sea  $e$  un valor cualquiera, entonces cuanto mayor es  $\varepsilon_{d_1}$  mayor es la probabilidad de  $\varepsilon > e$ . El mismo razonamiento nos lleva a la conclusión recíproca: para un valor cualquiera  $e$ , cuando mayor es  $\varepsilon_d$  mayor es la probabilidad de que  $\varepsilon_{d_1}$  sea (haya sido) mayor que  $e$ . Esto significa que los errores  $\varepsilon_{d_1}$  y  $\varepsilon_d$  están correlados positivamente.

La conclusión a la que hemos llegado es en cierto modo evidente, pues lo único que hemos demostrado es que un error está positivamente correlado con una cualquiera de sus componentes independiente de las demás. Pero el mismo argumento aplicado dos veces nos permite concluir que  $\varepsilon_d$  y  $\varepsilon_t$  están positivamente correlados. Así,

1. Para todo  $e$ , cuanto mayor es  $\varepsilon_d$  mayor es la probabilidad de  $\varepsilon_{d_1} > e$ .
2. A mayor  $\varepsilon_{d_1}$  mayor  $\varepsilon_{t_1}$  (por hipótesis).
3. Para todo  $e$ , cuanto mayor es  $\varepsilon_{t_1}$  mayor es la probabilidad de  $\varepsilon_t > e$ .

Uniendo los tres puntos (aunque más adelante matizaremos esta cuestión) se obtiene: Para todo  $e$ , cuanto mayor es  $\varepsilon_d$  mayor es la probabilidad de  $\varepsilon_t > e$ , y viceversa; es decir, que los errores en la medida de  $d$  y en la medida de  $t$  están correlados positivamente. Puesto que los errores y los valores observados sólo se diferencian en un desplazamiento, donde decimos error podemos decir valor observado. La gráfica de los pares  $(\varepsilon_d, \varepsilon_t)$  podría ser como la figura 2.8.

Si en una de las dos medidas se cumpliera que a mayor temperatura menor valor medido entonces la correlación sería negativa, y el conjunto de pares  $(\varepsilon_d, \varepsilon_t)$  sería una forma alargada en la dirección de los cuadrantes segundo y cuarto. Y si la temperatura para la cual el error  $\varepsilon_{d_1} = 0$  no es la misma que la que hace cero  $\varepsilon_{t_1}$  entonces el «eje» de la nube de puntos no pasaría por el origen.

## 2 Errores aleatorios

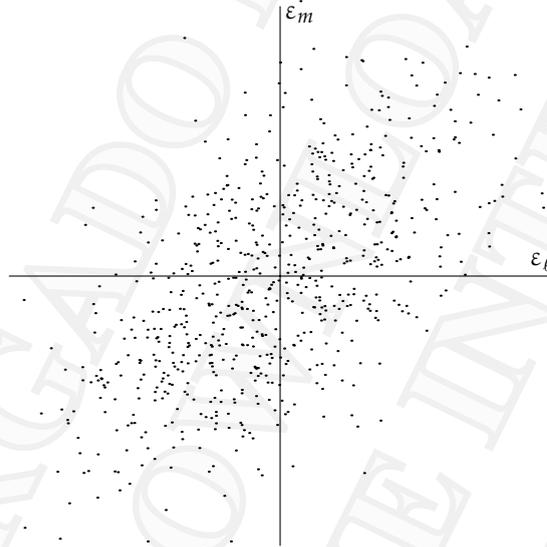


Fig. 2.8

Hemos visto que la correlación está planteada en términos de probabilidades. El conocimiento de  $\varepsilon_d$  no nos permite conocer  $\varepsilon_t$ , simplemente añade información al conocimiento de la variable aleatoria  $\varepsilon_t$ . Más exactamente, la función de distribución (o de densidad) de  $\varepsilon_t$  es función de  $\varepsilon_d$ ; y más exactamente todavía:

*Decimos que dos variables aleatorias  $\ell$  y  $m$  están correladas si la función de densidad  $f(\ell, m)$  no se puede escribir de la forma  $g(\ell)h(m)$ .*

Cuando una variable determina totalmente la otra se dice que hay dependencia funcional, y entonces  $m = m(\ell)$  y  $\ell = \ell(m)$ , y el conjunto de pares  $(m, \ell)$  es una curva. La dependencia funcional es un caso particular de correlación. Manteniendo la nomenclatura de la correlación, decimos que la relación funcional es positiva si la función es creciente, es decir, si a mayor  $\ell$  mayor  $m$ , y negativa en caso contrario. También se puede definir de manera precisa la correlación positiva o negativa en un punto, aunque no lo haremos aquí. Tanto la dependencia funcional como la correlación pueden ser positivas en unos

## 2 Errores aleatorios

intervalos o conjuntos de valores y negativas en otros.

En muchas ocasiones un error depende funcionalmente de uno o varios factores. En el ejemplo anterior de la distancia y el tiempo,  $\varepsilon_{d_1} = \varepsilon_{d_1}(T)$  y  $\varepsilon_{t_1} = \varepsilon_{t_1}(T)$ .

(2.6) Si existe dependencia funcional entre  $a$  y  $b$  y entre  $b$  y  $c$ , entonces obviamente también existe dependencia funcional entre  $a$  y  $c$ . Aplicado al ejemplo, hay dependencia funcional entre  $\varepsilon_{d_1}$  y  $\varepsilon_{t_1}$ .

(2.7) Cuando existe dependencia funcional entre  $a$  y  $b$  y correlación entre  $a$  y  $c$ , entonces también existe correlación entre  $b$  y  $c$ , y si  $a$  y  $c$  no están correladas tampoco lo estarán  $b$  y  $c$ . La dependencia funcional simplemente transforma un valor en otro, y no modifica la condición de correlación entre dos variables.

(2.8) Por el contrario, si  $a$  y  $b$  están correladas así como  $b$  y  $c$ , no podemos asegurar que  $a$  y  $c$  estén correladas. Esta no-propiedad se pone de manifiesto con un ejemplo sencillo: Sean  $\varepsilon_1, \varepsilon_2$  dos errores independientes, y  $\varepsilon = \varepsilon_1 + \varepsilon_2$ . Entonces  $\varepsilon_1$  está correlado con  $\varepsilon$  y  $\varepsilon_2$  con  $\varepsilon$ , y sin embargo  $\varepsilon_1$  y  $\varepsilon_2$  son independientes.

Volvamos entonces al ejemplo de los errores  $\varepsilon_d$  y  $\varepsilon_t$ . Recordamos las tres relaciones, escritas más brevemente:

1. Entre  $\varepsilon_d$  y  $\varepsilon_{d_1}$  existe correlación positiva.
2. Entre  $\varepsilon_{d_1}$  y  $\varepsilon_{t_1}$  existe dependencia funcional positiva.
3. Entre  $\varepsilon_{t_1}$  y  $\varepsilon_t$  existe correlación positiva.

El paso intermedio 2 podemos eliminarlo gracias a la propiedad (2.7), pues es una dependencia funcional, y queda:

1. Entre  $\varepsilon_d$  y  $\varepsilon_{t_1}$  existe correlación positiva.
2. Entre  $\varepsilon_{t_1}$  y  $\varepsilon_t$  existe correlación positiva.

Como acabamos de ver esto no permite concluir la correlación entre  $\varepsilon_d$  y  $\varepsilon_t$ . La información que falta es que para pasar de  $\varepsilon_{t_1}$  a  $\varepsilon_t$  se suma un elemento,  $\varepsilon_{t_0}$ , independiente tanto de  $\varepsilon_d$  como de  $\varepsilon_{t_1}$ , por lo que la correlación positiva entre  $\varepsilon_d$  y  $\varepsilon_{t_1}$  se transmite a  $\varepsilon_t$ . Con esto completamos el razonamiento acerca de la correlación de las medidas  $d$  y  $t$ . Enunciando de manera general, la propiedad es la siguiente:

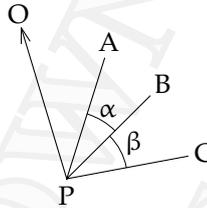
(2.9) Sean  $\ell = \ell_1 + \ell_2$  y  $m = m_1 + m_2$ , en donde  $\ell_1, \ell_2, m_1, m_2$  son variables aleatorias, tales que  $\ell_2$  y  $m_2$  son independientes entre sí y respecto a  $\ell_1$  y  $m_1$ , y entre  $\ell_1$  y  $m_1$  existe una correlación con un cierto

## 2 Errores aleatorios

signo. Entonces  $l$  y  $m$  son variables aleatorias con una correlación del mismo signo.

El ejemplo de la distancia y el tiempo es un caso particular en el que la correlación es una dependencia funcional. La propiedad también incluye el caso trivial en que  $l_1$  y  $m_2$  son independientes y entonces  $l$  y  $m$  también lo son.

Otro ejemplo de correlación se da en topografía en la medida de dos ángulos contiguos.



Se miden de manera directa los ángulos OPA, OPB y OPC, obteniendo unos valores  $L_A$ ,  $L_B$  y  $L_C$ . Los valores observados de los ángulos  $\alpha$  y  $\beta$  son entonces

$$\alpha = L_B - L_A, \quad \beta = L_C - L_B;$$

y sus errores

$$\begin{aligned} \varepsilon_\alpha &= \varepsilon_{L_B} - \varepsilon_{L_A}, \\ \varepsilon_\beta &= \varepsilon_{L_C} - \varepsilon_{L_B}. \end{aligned}$$

Suponemos que los errores en las medidas  $L_A, L_B, L_C$  son independientes. Las expresiones de los errores  $\varepsilon_\alpha$  y  $\varepsilon_\beta$  son exactamente las del teorema (2.9), con  $l_1 = \varepsilon_{L_B}$  y  $l_2 = -\varepsilon_{L_B}$ . Por lo tanto las medidas  $\alpha$  y  $\beta$  están correladas negativamente. La gráfica de los pares  $(\varepsilon_\alpha, \varepsilon_\beta)$  será una nube alargada hacia los cuadrantes segundo y cuarto. Si las distribuciones de los errores de  $L_A, \dots, L_C$  son iguales e iguales a la distribución  $f$  de más arriba, la pareja de errores  $(2, -2)$  es más probable (el punto tiene una densidad de probabilidad mayor) que la pareja  $(1, 1)$ , aunque ésta tenga unos valores menores.

El recíproco del teorema (2.9) es muy interesante desde un punto de vista práctico: Si dos observaciones están correladas es porque, de entre las componentes del error de la primera, existe al menos una

## 2 Errores aleatorios

que está correlada con una de las componentes del error de la segunda. Cuando llegamos en la descomposición hasta los factores primarios del error, entonces existe un factor que es común a ambos errores. En el ejemplo de la distancia y el tiempo es el factor  $\Delta T$ ; en el ejemplo de los ángulos, la medida  $L_B$ .

Si el número de medidas es mayor que dos entonces tendremos variables aleatorias de 3 o más dimensiones, formadas por  $n$ -uplas  $(l_1, l_2, \dots, l_n)$ . El conjunto de muchas  $n$ -uplas será una nube de puntos  $n$ -dimensional.

Sólo tiene sentido considerar variables aleatorias  $n$ -dimensionales si las medidas están correladas. Si no es así, es más sencillo y no resta ninguna información trabajar con cada variable por separado.



### 3 Estimación

#### **Concepto de estimador. Valores reales, observados y ajustados. Residuos**

Dado que las medidas contienen un cierto error, en muchas ocasiones, para obtener la precisión requerida o simplemente por seguridad, se repite la medida de una magnitud varias veces, o se toman más medidas de las necesarias para resolver el problema (por ejemplo, si para determinar la forma de un triángulo medimos los tres lados y los tres ángulos). La repetición de medidas puede entenderse como un caso particular de la obtención de más medidas de las necesarias.

En estos casos se pueden seguir criterios varios para tomar la solución definitiva. Cada criterio recibe el nombre de estimador. Por ejemplo, en el caso de medir varias veces una misma magnitud, el criterio o estimador seguido puede ser la media aritmética de los valores observados, la mediana, la media aritmética previa eliminación de los valores extremos, o cualquier otro que se nos pueda ocurrir. En el caso del triángulo en el que se han medido ángulos y lados, será un criterio que tenga en cuenta tanto las medidas de los ángulos como las de los lados.

La aplicación de un criterio determinado buscará obtener unos valores que se ajusten lo más posible a los valores reales. Los valores obtenidos se llaman valores ajustados, y son estimaciones de los valores reales. Los valores ajustados los representaremos con el siguiente tipo de letra:  $abl\alpha$ . Tenemos entonces para cada medida tres valores:

### 3 Estimación

valor real  $l$ , valor observado  $\ell$  y valor ajustado o estimado  $l$ .

Los valores observados, al estar afectados por unos errores, no serán coherentes como valores reales. Por ejemplo, volviendo al caso del triángulo, si los ángulos medidos son  $A, B, C$  y los lados  $a, b, c$ , no habrá ningún triángulo cuyos ángulos sean  $A, B, C$  y cuyos lados sean  $a, b, c$ . No se cumplirá  $A + B + C = 180^\circ$ ; tampoco la proporcionalidad entre lados y senos de ángulos, ni cualquiera otra relación que se pueda establecer entre los lados y ángulos de un triángulo. En cambio los valores ajustados sí que tienen que ser coherentes, ya que el conjunto de valores ajustados es una posible solución.

Cada valor ajustado implica una estimación del error:  $\varepsilon = \ell - l$ . La cantidad  $\varepsilon$  juega un papel central en toda la teoría de la estimación y recibe el nombre de residuo. Se representa por la letra  $v$ . Por lo tanto  $v = \varepsilon = \ell - l$ . Es habitual definir el residuo como el valor ajustado menos el observado,  $l - \ell$ , lo que da lugar a no pocas complicaciones cuando se representan gráficamente o se trabaja con distribuciones asimétricas. Resulta mucho más natural definir un valor ajustado del error, y no de menos-el-error.

Una vez elegido un estimador, es decir, un método, los valores ajustados son también variables aleatorias. Por ejemplo, supongamos una observación cuya función de densidad es uniforme entre los valores  $l - 0,5$  y  $l + 0,5$  (figura 3.1), que se realizan dos observaciones y que se elige como estimador la media aritmética. Entonces el valor ajustado  $l$  queda determinado mediante el siguiente proceso:

1. Obtener una medida  $\ell_1$ .
2. Obtener otra medida  $\ell_2$ .
3. Sumar  $\ell_1 + \ell_2$ .
4. Dividir entre dos.

Por lo tanto, si repetimos estos cuatro pasos muchas veces obtendremos valores  $l_1, l_2, \dots$ . De modo que el valor ajustado es una variable aleatoria. Se puede obtener teóricamente su distribución, y es la que se muestra en la figura 3.2. La parte inferior es el resultado experimental de aplicar los cuatro pasos muchas veces y trazar una raya vertical para cada valor  $l$  obtenido. Si el lector tiene algún medio para generar valores aleatorios con distribución uniforme es muy aconsejable que realice el experimento por sí mismo.

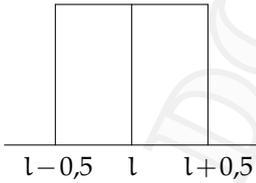


Fig. 3.1

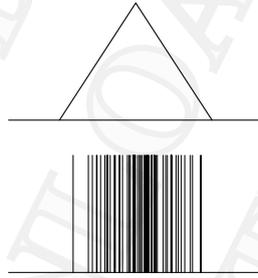


Fig. 3.2

Si en lugar de dos observaciones tomamos tres, entonces la media sigue la distribución de la figura 3.3. Si en lugar de tomar la media empleamos la mediana (en el caso de tres valores el valor del medio), el valor ajustado  $l$  sigue la distribución de la figura 3.4.



Fig. 3.3



Fig. 3.4

Puede verse que en este supuesto, es decir, distribución uniforme centrada en el valor real y tres medidas tomadas, es mejor la media que la mediana.

### Medida de los tres ángulos de un triángulo

Veamos un ejemplo más. En un triángulo se han medido los tres ángulos, obteniendo  $\ell_1, \ell_2, \ell_3$ , y se trata de estimar el valor verdadero de cada uno de ellos. Para ello sumamos los tres valores observados, y el error en la suma, que en este caso sí es conocido ya que sabemos el valor real,  $\ell_1 + \ell_2 + \ell_3 = 180^\circ$ , lo dividimos entre tres y lo restamos al valor observado de cada ángulo. Suponemos para los valores observados la misma distribución que en el ejemplo anterior.

### 3 Estimación

El estimador en este caso es el siguiente:

1.  $\varepsilon = (\ell_1 + \ell_2 + \ell_3) - 180$ ,
2.  $v_1 = v_2 = v_3 = \varepsilon/3$ ,
3.  $l_1 = \ell_1 - v_1$ ,
4.  $l_2 = \ell_2 - v_2$ ,
5.  $l_3 = \ell_3 - v_3$ .

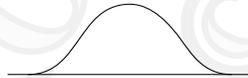
Para obtener la distribución de cada valor ajustado expresamos dicho valor en función de las observaciones originales, cuya distribución conocemos:

$$\begin{aligned}l_1 &= \ell_1 - v_1 = \ell_1 - ((\ell_1 + \ell_2 + \ell_3) - 180)/3 \\ &= \ell_1 + \varepsilon_1 - (\ell_1 + \ell_2 + \ell_3 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 - 180)/3 \\ &= \ell_1 + \varepsilon_1 - (\varepsilon_1 + \varepsilon_2 + \varepsilon_3)/3 = \ell_1 + \frac{2}{3}\varepsilon_1 - \frac{1}{3}\varepsilon_2 - \frac{1}{3}\varepsilon_3.\end{aligned}$$

Y las correspondientes igualdades para  $l_2$  y  $l_3$ :

$$\begin{aligned}l_1 &= \ell_1 + \frac{2}{3}\varepsilon_1 - \frac{1}{3}\varepsilon_2 - \frac{1}{3}\varepsilon_3, \\ l_2 &= \ell_2 - \frac{1}{3}\varepsilon_1 + \frac{2}{3}\varepsilon_2 - \frac{1}{3}\varepsilon_3, \\ l_3 &= \ell_3 - \frac{1}{3}\varepsilon_1 - \frac{1}{3}\varepsilon_2 + \frac{2}{3}\varepsilon_3.\end{aligned}$$

Las variables  $\varepsilon_1$ ,  $\varepsilon_2$  y  $\varepsilon_3$  son independientes, por lo que no resulta muy difícil obtener la distribución de cada uno de los  $l$  conociendo la de los primeros. La transmisión de distribuciones no la explicaremos de manera sistemática porque superaría el objetivo fijado para este libro en cuanto a la teoría sobre probabilidad y estadística. El resultado es la distribución siguiente. Para cada  $l$  está centrada en el valor real  $l$ :



Cada pareja  $l_1, l_2$ ,  $l_2, l_3$ ,  $l_3, l_1$  está correlada negativamente, como puede apreciarse en el resultado experimental de la figura 3.6.

En lo que respecta a las tres variables en conjunto, tienen dependencia funcional, ya que  $l_1 + l_2 + l_3 = 180$ , lo que permite expresar

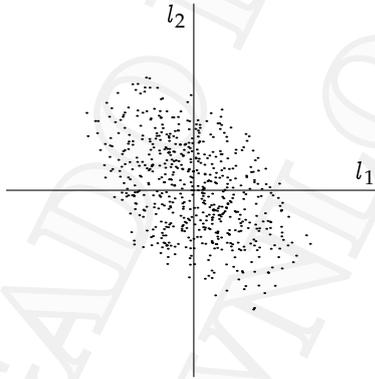


Fig. 3.6

cada uno en función de los otros dos. Siempre que se realicen más medidas de las necesarias para determinar el problema los valores ajustados estarán ligados mediante relaciones funcionales, tantas como exceso de medidas haya. Este exceso de medidas se llama redundancia. Así, si se miden los tres ángulos de un triángulo hay una redundancia. En el ejemplo anterior en el que se mide una misma magnitud tres veces hay dos redundancias, y las relaciones funcionales son  $l_1 = l_2 = l_3$ .

Además de los valores ajustados, cualquier magnitud que dependa de las observaciones será una variable aleatoria que sigue una cierta distribución. Por ejemplo, podemos estar interesados en el máximo valor de tres observaciones  $l_1, l_2, l_3$  de una misma magnitud. En este caso la función de distribución es muy fácil de obtener:

$$P(\max\{l_1, l_2, l_3\} \leq x) = P(l_1 \leq x, l_2 \leq x, l_3 \leq x) = F(x)F(x)F(x).$$

Si al máximo lo llamamos  $M$  y a su función de distribución  $F_M$ ,

$$F_M(x) = (F(x))^3,$$

y su función de densidad

$$f_M(x) = F'_M(x) = 3(F(x))^2 f(x).$$

### 3 Estimación

La distribución de un valor ajustado o estimado queda definida de la siguiente manera:

1. Se realiza un conjunto de medidas  $l_1, l_2, \dots, l_m$  de unas magnitudes  $l_1, l_2, \dots, l_m$ .
2. Las medidas  $l_1, l_2, \dots, l_m$  siguen una distribución que se conoce (o se supone).
3. El valor ajustado es una función de las medidas:  $x = f(l_1, l_2, l_3)$ .

La función del tercer apartado es lo que recibe el nombre de estimador. Estos son algunos de los estimadores que hemos visto hasta el momento:

$$l_1 = (l_1 + l_2 + l_3)/3;$$

$$l_1 = \frac{2}{3}l_1 - \frac{1}{3}l_2 - \frac{1}{3}l_3 + 60;$$

$$M = \text{máx}\{l_1, l_2, l_3\}.$$

En muchas ocasiones los valores ajustados pretenden ser una estimación de los valores reales. En esos casos los estimadores seleccionados deben proporcionar necesariamente un conjunto de valores  $l_1, l_2, \dots, l_m$  coherente.

#### **Función $g$ de estimación de los valores reales**

Vamos a ver un último ejemplo algo artificial pero muy ilustrativo. Sea un valor real  $l$ , que para concretar las ideas suponemos una longitud. Se mide  $l$  con una regla con división de centímetros, y no se intenta estimar fracciones de centímetro. La longitud a medir está muy bien definida, tiene un valor de algo más de 20 cm y se ve claramente que está más próxima a 20 que a 21. En estas circunstancias la distribución de  $l$  es una con toda la probabilidad acumulada en el valor 20. La de  $\varepsilon$  es, como siempre, la misma desplazada la cantidad desconocida  $-l$ . Por lo tanto siempre que repitamos la medida el valor obtenido será  $l = 20$ , y el error será el mismo, aunque no lo conozcamos. El valor ajustado será lógicamente  $l = 20$ , también constante.

Sin embargo, el hecho de obtener  $l = 20$  con probabilidad 1 no significa que el valor real sea 20. Puede haber sido 20,1, 20,2, 20,236... Si

además desconocemos el hecho de que el valor real es mayor que 20, entonces también podría ser  $l = 19,98, 19,6\dots$  Podemos incluso decir que la probabilidad de que el valor real haya sido 20,2 es la misma que la de que haya sido 19,7. Estamos introduciendo el concepto de probabilidad en el valor  $l$ , que es una cantidad fija. Se puede hacer de manera correcta de la siguiente manera: Hasta ahora hemos considerado la distribución que sigue el valor  $\ell$  cuando la magnitud a medir tiene el valor real  $l$ . Podemos considerar cuál sería la distribución de  $\ell$  si el valor real fuese otro valor  $x \neq l$ , que denotaremos por  $f_x(\ell)$ , y esto lo podemos hacer para todo  $x$ .

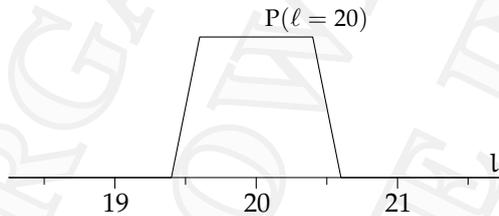


Fig. 3.7: Probabilidad de  $\ell = 20$  para cada  $l$

La gráfica 3.7 muestra la probabilidad de  $\ell = 20$  para cada valor de  $l$ . La probabilidad es 1 cuando 20 es claramente el valor más cercano a  $l$ , y es 0 cuando el valor más cercano es claramente otro. Si  $l$  está próxima a un medio centímetro se considera que puede haber vacilación a la hora de asignar  $\ell$ . En concreto se han tomado como límites para esa vacilación 0,4 y 0,6, siendo más alta la probabilidad de tomar el valor por exceso cuando mayor es  $l$ .

La gráfica 3.7 no es una función de densidad. Es simplemente una función que toma como valor una probabilidad, y por lo tanto habrá de estar entre 0 y 1, nada más. Sin embargo se transforma fácilmente en la función de densidad de « $l$  sabiendo que  $\ell = 20$ », que llamaremos  $g(l)$ . A priori (es decir, antes de conocer que el valor observado ha sido 20), no hay nada que nos haga suponer que el valor real tiene más probabilidad de ser 19,9 que de ser 19,42, por ejemplo. Sin embargo la función de la figura 3.7 indica que de las veces en que  $l = 19,42$  sólo en una proporción de 0,1 obtendremos  $\ell = 20$ , mientras que si  $l = 19,9$  el resultado siempre será  $\ell = 20$ . Por lo tanto, si hemos obtenido  $\ell = 20$  es más probable que el valor original fuese 19,9 que que

### 3 Estimación

fuese 19,42. Exactamente es  $1/0,1 = 10$  veces más probable. Ahora sí estamos hablando de densidad de probabilidad, pues la probabilidad de que  $l$  sea un valor concreto es cero. En definitiva, la función de densidad de  $l$  sabiendo que  $\ell = 20$  es *proporcional* a la fig. 3.7, es decir, a la probabilidad de  $\ell = 20$  para cada  $l$ .

La constante de proporcionalidad se halla sabiendo que, al ser una función de densidad, el área bajo la curva tiene que ser 1 (propiedad (2.1)). En este ejemplo concreto ha dado la casualidad de que el área ya es 1, así que la constante es 1 y la función se queda como está (fig. 3.8).

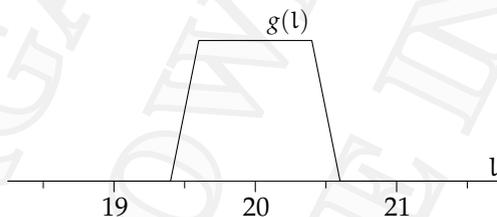


Fig. 3.8: Densidad de probabilidad de  $l$  sabiendo que  $\ell = 20$

#### Más ejemplos de función $g$

Si la distribución es uniforme entre dos extremos, por ejemplo  $l - 2, l + 2$ , y obtenemos una medida  $\ell$ , entonces es evidente que el valor real puede estar entre  $\ell - 2$  y  $\ell + 2$  con igual probabilidad. Sin embargo, si existe una asimetría en el proceso de medida, y el valor medido sigue una distribución uniforme entre  $l - 1, l + 2$ ; entonces, obtenida una medida  $\ell$ , el valor  $l$  puede estar entre  $\ell - 2$  y  $\ell + 1$ , también con igual probabilidad (fig. 3.9).

Si la distribución de  $\ell$  es uniforme entre  $l - 2$  y  $l + 2$ , pero obtenemos dos medidas en lugar de una, entonces la distribución de posibles  $l$  cambia. Sea por ejemplo  $\ell_2 - \ell_1 = 1,2$ : entonces  $l$  no puede ser menor de  $\ell_1 - 0,8$ , ya que en ese caso habría sido imposible que hubiésemos obtenido  $\ell_2$ . Análogamente tampoco puede ser mayor que  $\ell_2 + 0,8$ . Por lo tanto la distribución de posibles  $l$  es uniforme entre  $\ell_1 - 0,8$  y  $\ell_2 + 0,8$ . La longitud de este intervalo es  $0,8 + 1,2 + 0,8 = 2,8$ , por lo que el valor de la función de densidad en ese intervalo es

### 3 Estimación

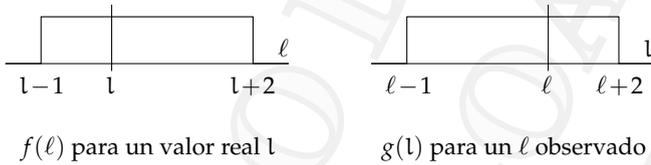


Fig. 3.9

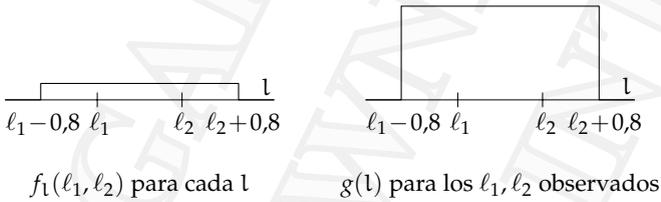


Fig. 3.10

$1/2,8$ , para que se cumpla (2.1). En la figura 3.10 se muestra a la derecha la función  $g(l)$  y a la izquierda la función que, para cada  $l$ , da la densidad del par de medidas  $l_1, l_2$ ; es decir, para cada  $l$  consideramos su función de densidad bidimensional para un par de observaciones, y tomamos el valor que esta función tiene en el punto  $l_1, l_2$ . En el ejemplo de la regla que mide centímetros la función equivalente era la probabilidad de obtener  $l$ . Ahora es una densidad de probabilidad porque la probabilidad de obtener una pareja concreta  $l_1, l_2$  es cero. En los puntos en los que no vale cero esta función vale  $1/16$ , puede el lector intentar averiguar por qué.

Puesto que la función de densidad de  $l$ , que es la que interesa, es el resultado de multiplicar la que da  $f_1(l_1, l_2, \dots, l_m)$  para cada  $l$  por la constante adecuada para que se cumpla (2.1), el valor exacto de ésta es irrelevante, sólo importa la proporcionalidad entre unos valores y otros.

Supongamos ahora que la función  $f$  de densidad de las observaciones es triangular como la de la figura 3.2, entre los valores  $l - 1$  y  $l + 1$ . Realizamos dos medidas y obtenemos  $l_1, l_2$ , que se diferencian en  $1,6$ . Construimos la función que, tomando como variable  $l$ , da la densidad de probabilidad de  $(l_1, l_2)$ , sin más que para cada  $l$  tomar el valor  $f_1(l_1)f_1(l_2)$ . Si  $l$  es menor que  $l_1 + 0,6$  o mayor que  $l_2 - 0,6$

### 3 Estimación

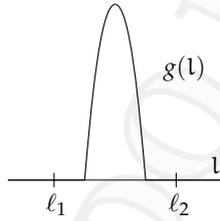


Fig. 3.11

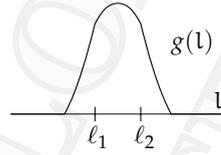


Fig. 3.12

(que son respectivamente  $l_2 - 1$  y  $l_1 + 1$ ), el resultado será cero. Entre ambos valores vale

$$(1 - (l - l_1))(1 - (l_2 - l)).$$

Multiplicando por la constante adecuada para que la probabilidad total sea 1 obtenemos la función de densidad  $g(l)$  para los valores  $(l_1, l_2)$  observados. El resultado es el de la gráfica 3.11.

Si la diferencia entre las dos medidas es 0,6, entonces si  $l$  lo tomamos entre  $l_1 - 0,4$  y  $l_1$  los dos valores  $f_1(l_1), f_1(l_2)$  los obtenemos de la rama derecha de  $f$ , y por lo tanto  $g(l)$  será proporcional a

$$(1 - (l_1 - l))(1 - (l_2 - l)).$$

Si  $l$  está entre  $l_1$  y  $l_2$ , cada observación está a un lado de  $l$ , y

$$g(l) \propto (1 - (l - l_1))(1 - (l_2 - l))^*.$$

Finalmente, si  $l$  está entre  $l_2$  y  $l_2 + 0,4$  entonces

$$g(l) \propto (1 - (l_2 - l))(1 - (l_2 - l)).$$

La función  $g$  es la de la gráfica 3.12.

#### Proceso general para la obtención de la función $g$

El proceso general de obtención de la función  $g$  cuando hay varias magnitudes medidas es el siguiente:

---

\*El símbolo  $\propto$  significa proporcional a.

### 3 Estimación

Hay un conjunto de  $m$  magnitudes cuyos valores reales, desconocidos, son  $l_1, l_2, \dots, l_m$ . Hemos medido las  $m$  magnitudes y obtenido unos valores  $\ell_1, \ell_2, \dots, \ell_m$ .

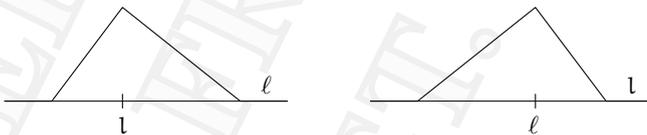
Todas las magnitudes quedan determinadas conociendo  $n$  magnitudes independientes de entre ellas. Entonces el conjunto de posibles  $m$ -uplas  $(l_1, l_2, \dots, l_m)$  es un subconjunto de dimensión  $n$  del espacio  $m$ -dimensional de las  $m$ -uplas (si todas las magnitudes son números reales este espacio será  $\mathbb{R}^m$ ).

Para cada  $(l_1, l_2, \dots, l_m)$  posible suponemos una distribución para los valores observados. Para cada una de estas posibles  $m$ -uplas los valores observados pueden ser cualesquiera, de modo que, para cada  $(l_1, l_2, \dots, l_m)$  concreto, la distribución de los  $(\ell_1, \ell_2, \dots, \ell_m)$  es  $m$ -dimensional. Si las observaciones son independientes estadísticamente (no están correladas) la función de densidad  $f(\ell_1, \ell_2, \dots, \ell_m)$  será simplemente el producto de las funciones de densidad individuales  $f_1(\ell_1)f_2(\ell_2) \dots f_m(\ell_m)$ , pero en general no será así. En todos los ejemplos vistos hasta ahora, al variar el conjunto de valores reales la función  $f$  simplemente se desplaza con ellos. Más adelante veremos un ejemplo en el que esto no es así.

Para cada  $(l_1, l_2, \dots, l_m)$  posible tomamos el valor de la correspondiente  $f_{l_1, l_2, \dots, l_m}(\ell_1, \ell_2, \dots, \ell_m)$  en el punto  $(\ell_1, \ell_2, \dots, \ell_m)$  de valores medidos, obteniendo así una función de las variables  $l_1, l_2, \dots, l_m$ .

Multiplicamos la función anterior por la constante adecuada para que la integral total sea 1, y obtenemos  $g(l_1, l_2, \dots, l_m)$ .

La función  $g$  contiene toda la información que podemos deducir sobre los valores reales a partir de los valores observados. Cuando la función  $f$  es asimétrica la función  $g$  suele presentar la misma asimetría pero en sentido opuesto, como en la figura 3.9. Esto es lógico, ya que si respecto al valor real es más fácil equivocarse por exceso, entonces, obtenido un valor  $\ell$ , es más fácil que el valor real sea menor que  $\ell$  que que sea mayor. Un ejemplo bastante claro es el de la siguiente figura, en el que se considera que se observó un único valor  $\ell$ .



### 3 Estimación

#### Estimación de una magnitud de tipo escala

El último ejemplo de este capítulo es algo más complicado que los anteriores, pero muestra muy bien el carácter inverso de la estimación y las dificultades y paradojas que pueden acompañar al resultado. Sea una función

$$f_2(\ell) = \ell e^{-\ell}, \quad x \geq 0,$$

para un valor real  $l = 2$ .

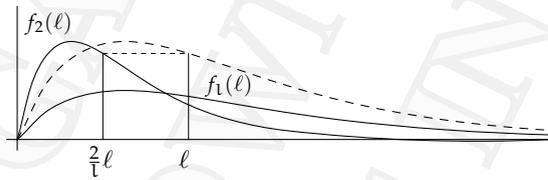


Fig. 3.14

Si el valor  $l$  es otro, la función de densidad en lugar de desplazarse se escala, de manera que el valor  $f_1(\ell)$  es  $f_2(\frac{2}{1}\ell)$ , como se muestra en la figura 3.14. Pero al escalar de esta manera estamos modificando el área total, en proporción  $l/2$ . Entonces es necesario multiplicar toda la función por  $2/l$ , de manera que el área total siga siendo 1. Por lo tanto la función  $f_l$ , para cualquier  $l$ , es

$$f_l(\ell) = \frac{2}{l} f_2\left(\frac{2}{l}\ell\right).$$

Antes de obtener la función  $g(l)$  vamos a hacer un pequeño estudio de la función  $f$ . Nos limitaremos a  $f_2$ , y los resultados obtenidos se pueden aplicar a  $f$  escalando por la constante  $l/2$ .

La esperanza o media es  $\int_0^{\infty} \ell^2 e^{-\ell} d\ell = 2$ ; es decir, el valor real.

La función de distribución  $F_2$  es  $\int_0^{\ell} f_2(t) dt$ .

$$F_2(\ell) = 1 - (\ell + 1)e^{-\ell}.$$

La probabilidad de obtener un valor menor que el real, 2, es

$$F_2(2) = 1 - 3e^{-2} \approx 0,594.$$

Y la mediana,

$$F_2^{-1}(1/2) \approx 1,678.$$

El máximo de  $f_2$  se da en  $f_2'(\ell) = 0$ , que se corresponde con  $\ell = 1$ .

La función  $f_2$  comienza en el punto  $(0,0)$  y va creciendo hasta  $\ell = 1$ , punto a partir del cual empieza a decrecer indefinidamente. La función es muy asimétrica respecto a  $\ell = 1$ , y aunque este sea el valor máximo la mediana está en  $1,678 \dots$  y la media, que coincide con  $l$ , se va hasta 2.

La función  $g(l)$  para un valor  $\ell$  es proporcional a

$$f_1(\ell) = \frac{2}{l} f_2\left(\frac{2}{l}\ell\right) = \frac{4\ell}{l^2} e^{-\frac{2\ell}{l}}.$$

Tenemos que  $\int \frac{4\ell}{l^2} e^{-\frac{2\ell}{l}} d\ell = 2e^{-\frac{2\ell}{l}}$ , y  $\left[2e^{-\frac{2\ell}{l}}\right]_0^\infty = 2$ . Así que, finalmente,

$$g(l) = \frac{2\ell}{l^2} e^{-\frac{2\ell}{l}}, \quad G(l) = e^{-\frac{2\ell}{l}}. \quad (3.1)$$

En la figura 3.15 se muestran las funciones  $f_2(\ell)$  y  $g_2(l)$ , considerando que esta última es  $g(l)$  cuando  $\ell = 2$ .

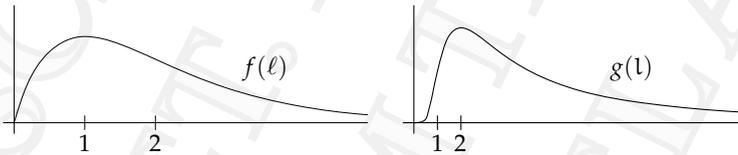


Fig. 3.15

Podríamos haber intentado obtener directamente  $G(l)$  a partir de  $F(\ell)$  mediante el siguiente razonamiento: Al escalar se  $l$  la función  $f$  lo que se mantiene constante es la probabilidad de que  $\ell$  sea menor (o mayor) que cualquier múltiplo de  $l$ . Por ejemplo, la probabilidad de que  $\ell$  sea mayor que  $1,6l$  es independiente de  $l$ , y en concreto vale  $F_2(3,2) \approx 0,475$ . Entonces podemos pensar que, si hemos obtenido un valor  $\ell$ , la probabilidad de que  $l$  sea menor que  $k\ell$ , para cualquier  $k$ , es la probabilidad de que se obtenga un valor mayor que  $\frac{1}{k}l$  (ya que si  $l < k\ell$  entonces  $\ell > \frac{1}{k}l$ ), que es  $1 - F\left(\frac{1}{k}l\right)$ .

### 3 Estimación

Por lo tanto

$$G(k\ell) = 1 - F\left(\frac{1}{k}\ell\right) = 1 - F_2(2/k).$$

$l = k\ell \Rightarrow 1/k = \ell/l$ , de donde

$$G(l) = 1 - F_2\left(\frac{2}{l}\right) = \left(\frac{2}{l} + 1\right)e^{-\frac{2}{l}}, \quad g(l) = \frac{4}{l^3}e^{-\frac{2}{l}}. \quad (3.2)$$

Este resultado es distinto de (3.1). Mirando por ejemplo el valor  $l = \ell$ , vemos que (3.1) da la probabilidad  $G(\ell) \approx 0,135$ , mientras que según (3.2) tenemos  $G(\ell) \approx 0,406$ . Puesto que el factor  $\left(\frac{2}{l} + 1\right)$  es siempre mayor que 1, siempre obtendremos un valor mayor con la segunda expresión que con la primera; la segunda distribución asigna más probabilidad a los valores menores. Esto lo podemos ver también en las funciones  $g$ . A parte de la diferencia en el factor  $2\ell$ , que es una constante, la diferencia entre las dos funciones radica en el factor  $l$  en el denominador, que conlleva valores de  $g(l)$  menores para mayores  $l$ .

El razonamiento seguido en el segundo caso es erróneo. No se cumple  $G(k\ell) = 1 - F\left(\frac{1}{k}\ell\right)$ , aunque parezca paradójico. En la distribución de valores observados, si para un valor  $x$  la probabilidad de que el valor observado sea menor que  $x$  es la misma que la de que sea mayor (por lo tanto  $x$  es la mediana), eso no implica que la probabilidad de obtener un valor concreto  $\ell$  mayor que  $x$  sea la misma que la de obtener un valor concreto menor que  $x$ . Para valores menores que  $x$  la probabilidad  $1/2$  se reparte en el intervalo  $(0, x)$ , mientras que la probabilidad  $1/2$  para valores mayores se reparte en  $(x, \infty)$ . Se verá claramente con el siguiente ejemplo, que es un caso discreto en vez de continuo.

Supongamos que el valor real sólo admite las posibilidades  $l = 10$  y  $l = 20$ . Si  $l = 10$ , sea para cada uno de los valores o conjuntos de valores siguientes la misma probabilidad:

$$1, 2, 3, 4, 5, 6, 7, 8, 9, \{11, \dots, 20\}, \{21, \dots, 30\}, \dots, \{91, \dots, 100\}.$$

Hay nueve valores menores que 10 y nueve grupos mayores que 10, por lo que la probabilidad de obtener un valor menor que el real es la misma que la de obtener uno mayor.

Hay 18 valores o grupos, por lo que la probabilidad de cada uno de ellos es  $1/18$ . En los grupos la probabilidad se distribuye por igual

entre los 10 valores, así que la probabilidad de cada uno de ellos es  $1/180$ .

Supongamos que si el valor real es 20 la distribución de valores observados  $\ell$  es análoga:

$$\{1, 2\}, \{3, 4\}, \dots, \{17, 18\}, \{21, \dots, 40\}, \{41, \dots, 60\}, \dots, \{181, \dots, 200\}.$$

También hay la misma probabilidad de obtener una medida menor que la real que una mayor, y la probabilidad de cada grupo o pareja es igualmente  $1/18$ . La probabilidad de cada valor menor que 20 es  $1/36$ , y es  $1/360$  para los mayores.

Supongamos que medimos y obtenemos el valor 12. Si  $l = 10$ , la probabilidad de obtener el número 12 es  $1/180$ , mientras que si  $l = 20$  la probabilidad es  $1/36$ . Supongamos un conjunto de 3600 medidas. Si no sabemos nada acerca de los valores reales, es decir, si suponemos que en principio todos tienen la misma probabilidad, entonces por término medio 1800 de esas medidas corresponderán a un  $l = 10$  y 1800 a  $l = 20$ . De las 1800 observaciones en las que  $l = 10$ , habrá  $1800/180 = 10$  en las que se obtenga  $\ell = 12$ , y de las 1800 con  $l = 20$  se obtendrá  $\ell = 12$  en  $1800/36 = 50$  medidas.

Por lo tanto, de cada 60 veces en las que se obtenga un valor medido igual a 12 sólo en 10 ocasiones será  $l = 10$ , y en las otras 50  $l = 20$ . Así, aunque la probabilidad de obtener un valor mayor que  $l$  es la misma que la de obtener uno menor, obtenido un valor  $\ell = 12$  (y cualquiera desde 11 hasta 18), la probabilidad de  $l < \ell$  es sólo  $1/6$ , mientras que la de  $l > \ell$  es  $5/6$ .

Considerando más valores  $l$  posibles y realizando cada vez divisiones más finas del conjunto de posibles  $\ell$ , podemos aproximarnos cada vez más al caso continuo, siempre cumpliéndose que  $G(l)$  es menor que  $1 - F(\ell)$ .

### Distribución de probabilidad a priori de los valores reales

Hasta ahora hemos supuesto siempre que no conocemos nada a priori acerca de la mayor probabilidad de unos valores reales u otros. Por ejemplo, en el ejemplo discreto que acabamos de estudiar hemos supuesto que de 3600 observaciones 1800 por término medio corresponderán a cada valor real. Si hubiésemos supuesto, pongamos por caso, que por cada dos veces que el valor real es igual a 10 sólo una es

### 3 Estimación

igual a 20, entonces de las 3600 observaciones en 2400 ocasiones por término medio el valor real será 10, y en las otras 1200 será 20. Bajo este supuesto se obtendrá el valor 12 en 46,6 ocasiones, de las cuales 13,3 se corresponderán con  $l = 10$  y 33,3 con  $l = 20$ . La proporción es ahora  $2/5$ , que es el resultado de multiplicar la original de  $1/5$  por la proporción supuesta a priori, en este caso  $2/1$ .

Llevado al caso continuo, suponemos para los valores reales una distribución  $h(l_1, l_2, \dots, l_m)$ . Hay que multiplicar la función  $g$  original por esa función  $h$  y por una nueva constante para que la probabilidad total siga siendo 1, y obtenemos así la nueva función  $g$ . Por lo tanto el valor de la función  $g$  en cada punto  $(l_1, \dots, l_m)$  es ahora proporcional a  $h(l_1, \dots, l_m) f_{l_1, \dots, l_m}(l_1, \dots, l_m)$ . En lugar de la función  $h$  podemos emplear cualquier función proporcional a ella (es decir, que no nos tenemos que preocupar por que la integral total sea 1), que denotaremos también por  $h$  aunque no sea propiamente una función de densidad, ya que de todos modos al final tendremos que calcular una constante para ajustar la integral de la función  $g$  a 1.

Volviendo al ejemplo  $f_1(\ell) = 2/l f_2(\frac{2}{l}\ell)$ , vamos a suponer una distribución de los valores reales proporcional a  $h(l) = 1/l$ . Entonces ahora  $g$  es proporcional a

$$\frac{1}{l} \frac{2}{l} f_2\left(\frac{2}{l}\ell\right) = \frac{2}{l^2} \frac{2\ell}{l} e^{-\frac{2\ell}{l}} = \frac{4\ell^2}{l^3} e^{-\frac{2\ell}{l}}.$$

Hallamos la constante de proporcionalidad, que resulta ser 1, y obtenemos

$$g(l) = \frac{4\ell^2}{l^3} e^{-\frac{2\ell}{l}}, \quad G(l) = \left(1 + \frac{2\ell}{l}\right) e^{-\frac{2\ell}{l}}, \quad (3.3)$$

que podemos comparar con (3.1).

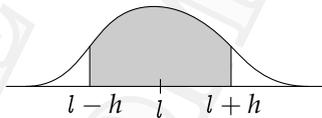
La función  $g$  nueva tiene una asimetría mucho menos marcada. Como muestra, ahora  $G(\ell) \approx 0,406$  frente al 0,135 anterior.

Para funciones del tipo de esta  $f$ , que se escalan con  $l$ , «no saber nada» acerca de la distribución de los valores reales muchas veces significa más bien una función  $h$  como la que hemos supuesto que no emplear ninguna función  $h$ . Esto lo veremos con más detalle cuando estudiemos la estimación de la varianza de una distribución.

### Intervalos de confianza

Como ya apuntamos, la función  $g$  contiene toda la información que podemos dar acerca de los valores reales a partir de la información de que disponemos, que es la suposición una función  $h$  que representa la distribución de todos los valores reales (si no empleamos ninguna estamos haciendo uso implícitamente de una función constante), la suposición de una distribución de los valores observados para cada conjunto de posibles valores reales, y el hecho de haber obtenido unos valores concretos en las observaciones. Sin embargo, cuando alguien nos pregunta por un número no quiere que le respondamos con una función (básicamente porque no sabrá manejar el resultado). El extremo opuesto a dar la función  $g$  es obtener simplemente un valor ajustado. Conviene resaltar que la función  $g(l)$  y el valor ajustado  $l$  son entidades independientes. Partiendo de unas hipótesis (las distribuciones  $f$  y  $h$ ), la función  $g$  es una, única; es el resultado. Sin embargo el valor ajustado es un número, función de los valores observados. Dicho de una manera un poco brusca, pero cierta, es el resultado de sustituir los valores  $(\ell_1, \dots, \ell_m)$  en una fórmula, y podemos generar tantos valores ajustados distintos como estimadores se nos ocurran.

Una información intermedia entre el valor ajustado y la función  $g$  son los intervalos de confianza. Situamos el valor obtenido  $l$  en la gráfica de  $g$ , y buscamos un intervalo en torno a  $l$  tal que la probabilidad contenida por la distribución  $g$  en ese intervalo sea un cierto valor, por ejemplo el 90%.



Por supuesto que hay muchas maneras de obtener tal intervalo, según situemos el punto  $l$  hacia la parte derecha del intervalo, centrado, o hacia la izquierda. Supongamos que lo tomamos en el centro. Entonces obtenemos el valor  $h$  tal que  $(l - h, l + h)$  contiene el 90% de probabilidad, es decir, existe una probabilidad del 90% de que el valor real  $l$  se encuentre en ese intervalo. El resultado se dará de la siguiente manera:

$$l \pm h,$$

### 3 Estimación

en donde  $l$  y  $h$  serán valores concretos y en algún lugar especificamos que el intervalo es el correspondiente al 90 % de probabilidad. Se pueden dar varios intervalos, por ejemplo los correspondientes al 90, 95 y 98 %.

Los intervalos de confianza están muy estrechamente ligados a la función  $g$ . Recordemos que la manera en que se define una función de densidad, y en particular la función  $g$ , es que  $\int_a^b g$  es la probabilidad acumulada en ese intervalo. Al dar algunos intervalos estamos aportando parte de la función  $g$ . Dar infinitos intervalos (y de la manera adecuada) sería equivalente a dar la función  $g$ .

Si hay más valores estimados, los entornos de confianza serán regiones bidimensionales, tridimensionales, y en general  $n$ -dimensionales, aunque los valores estimados se suelen agrupar de 2 en 2, o de 1 en 1, dando respectivamente superficies o intervalos de confianza.

La situación del intervalo o el área respecto al valor ajustado puede tomarse según varios criterios. Uno empleado a menudo es que las densidades de probabilidad (el valor de la función  $g$ ) en los extremos del intervalo (o a lo largo del borde del área) sean iguales. Este es un criterio teórico que da lugar, por ejemplo, a las famosas elipses de error, que explicaremos en capítulos posteriores. Sin embargo parece más lógico tomar el intervalo de manera que los extremos estén igual de alejados del centro, en donde la distancia la medimos en términos de utilidad. Por ejemplo, si el valor calculado es un factor de escala, la precisión de alguna magnitud, o algo similar, entonces tal vez consideremos que el valor  $2x$  está tan alejado de  $x$  como  $\frac{1}{2}x$ , y  $\frac{3}{2}x$  como  $\frac{2}{3}x$ . En estos casos los intervalos de confianza han de ser  $(x/h, hx)$ .

En la gran mayoría de los casos, en los que calcularemos coordenadas u otros valores que no tienen un cero absoluto, y en los que el alejamiento en términos de utilidad es el mismo en todas las direcciones, el mejor intervalo será simplemente  $(x - h, x + h)$ , y en el caso bidimensional un círculo.

#### Apéndice al capítulo tercero

Se puede explicar la función  $g(\ell_1, \dots, \ell_m)$  de manera muy breve a partir de las funciones  $f_{\ell_1, \dots, \ell_m}(\ell_1, \dots, \ell_m)$ , aunque será incomprendible para el que no esté acostumbrado al manejo de distribuciones multidimensionales y probabilidades condicionadas.

### 3 Estimación

Definimos una función  $\mathcal{F}(l_1, \dots, l_m, \ell_1, \dots, \ell_m)$ . La función está definida en los puntos en los que los valores  $(l_1, \dots, l_m)$  son posibles. El conjunto de posibles  $(l_1, \dots, l_m)$  tiene dimensión  $n$ , y por lo tanto el dominio de  $\mathcal{F}$  es  $n+m$ -dimensional. Cada punto de ese dominio es el resultado de una observación.

$\mathcal{F}(l_1, \dots, l_m, \ell_1, \dots, \ell_m) = f_{l_1, \dots, l_m}(\ell_1, \dots, \ell_m)$ , de manera que  $\mathcal{F}$  es proporcional a la función de densidad de los  $(l_1, \dots, l_m, \ell_1, \dots, \ell_m)$ . No es posible obtener la constante de proporcionalidad porque la integral de  $\mathcal{F}$  es infinito, ya que para cada  $(l_1, \dots, l_m)$ ,  $\int \mathcal{F} = \int f = 1$ , en donde las integrales están extendidas a todo el dominio de valores  $(\ell_1, \dots, \ell_m)$  y la función  $f$  es la que corresponde para los  $(l_1, \dots, l_m)$  concretos.

El corte de la función  $\mathcal{F}$  por unos valores  $(\ell_1, \dots, \ell_m)$  es, salvo por una constante, la probabilidad de  $(l_1, \dots, l_m)$  sabiendo que (condicionada)  $(\ell_1, \dots, \ell_m)$  es igual a los  $(\ell_1, \dots, \ell_m)$  concretos. Calculamos la constante de manera que la integral sea 1 y obtenemos así  $g(\ell_1, \dots, \ell_m)$ .

Para poder obtener una función  $g$  es necesario que, fijados unos  $(\ell_1, \dots, \ell_m)$ , la integral  $\int \mathcal{F}$  para los  $(l_1, \dots, l_m)$  posibles sea finita. En la práctica se cumple siempre, ya que  $\mathcal{F}(l_1, \dots, l_m, \ell_1, \dots, \ell_m)$  se aproxima rápidamente a cero cuando  $(\ell_1, \dots, \ell_m)$  se aleja de  $(l_1, \dots, l_m)$ .

Si conocemos a priori la distribución de los  $(l_1, \dots, l_m)$  ( $n$ -dimensional), que llamaremos  $h$ , entonces

$$\mathcal{F}(l_1, \dots, l_m, \ell_1, \dots, \ell_m) = h(l_1, \dots, l_m) f_{l_1, \dots, l_m}(\ell_1, \dots, \ell_m).$$

En esta situación pueden darse dos casos: que la función  $h$  sea auténticamente una función de densidad, en cuyo caso  $\mathcal{F}$  sí que es la función de densidad del conjunto de valores  $(l_1, \dots, l_m, \ell_1, \dots, \ell_m)$ ; o que  $h$  sea una función de integral infinito que consideramos proporcional a la función de densidad de los valores reales, como por ejemplo el caso anterior de  $h = 1/l$ , y entonces tampoco se puede calcular la constante de proporcionalidad.



## 4 Algunas distribuciones

Vamos a mostrar ya de manera concreta las distribuciones más importantes en el campo del ajuste de observaciones en topografía.

En primer lugar la distribución uniforme, que ya se vio en los capítulos precedentes. Falta mencionar que la desviación típica de una distribución uniforme de base 1 es  $\sigma = 1/\sqrt{12}$ .

Si se suman dos cantidades que siguen una distribución uniforme el resultado es una cantidad que sigue una distribución triangular, y si se suman tres el resultado es una distribución formada por tres ramas de parábola.

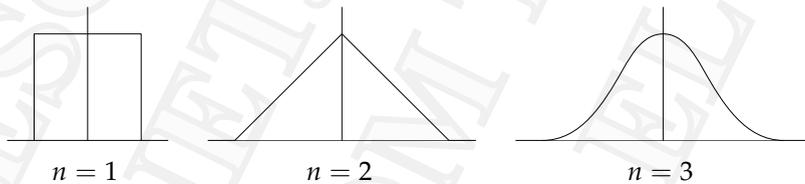


Fig. 4.1: Suma de  $n$  distribuciones uniformes

Si se suma un número cada vez mayor de distribuciones uniformes iguales la variable resultante se aproxima cada vez más a una distribución particular, llamada distribución normal. Una distribución normal de media 0 y desviación típica  $\sigma$  es la siguiente:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}}. \quad (4.1)$$

## 4 Algunas distribuciones

La constante  $(\sqrt{2\pi})^{-1}$  se incluye para ajustar la probabilidad total a 1. Si la media es otro valor  $\mu$  en lugar de 0, simplemente se sustituye  $\ell$  por  $\ell - \mu$ . Se suele emplear la notación  $N(\mu, \sigma)$  para referirse a una distribución normal de media  $\mu$  y desviación típica  $\sigma$ , o  $(\mu, \sigma)$  si sólo queremos referir los dos valores.

Esta convergencia a la distribución normal no aparece únicamente en la distribución uniforme, sino que sucede para cualquier distribución: si se suman distribuciones iguales el resultado se aproxima cada vez más a una distribución normal. Esta característica sorprendente, un poco más generalizada, es lo que recibe el nombre de teorema central del límite. Sobre este teorema v. cap. 6.

Si una variable  $\ell$  sigue una distribución  $A$  lo representaremos mediante  $\ell \sim A$ .

Si  $\ell \sim N(\mu, \sigma)$  entonces  $k\ell \sim N(k\mu, k\sigma)$ . Esta propiedad la cumple cualquier familia de distribuciones. Sin embargo, lo que es una propiedad característica de la distribución normal es que si se suman dos variables con distribuciones normales el resultado es otra distribución normal, como por otra parte cabría esperar de una distribución que se obtiene como límite al sumar distribuciones. La desviación típica de la distribución resultante es la composición cuadrática de las desviaciones típicas originales; es decir,  $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ , y en general, si  $\ell_1, \dots, \ell_n$  son variables aleatorias independientes que siguen distribuciones normales  $N(\mu_1, \sigma_1), \dots, N(\mu_n, \sigma_n)$ , entonces la cantidad

$$\ell = k_1\ell_1 + k_2\ell_2 + \dots + k_n\ell_n$$

sigue una distribución normal  $(\mu, \sigma)$  en donde

$$\begin{aligned}\mu &= k_1\mu_1 + k_2\mu_2 + \dots + k_n\mu_n, \\ \sigma^2 &= k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \dots + k_n^2\sigma_n^2.\end{aligned}$$

Un caso particular y muy importante de esta propiedad es la media de  $n$  variables que siguen una misma distribución normal  $N(\mu, \sigma)$ . Aplicando las fórmulas obtenemos directamente:

$$\bar{\ell} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right). \quad (4.2)$$

También suele presentarse el caso de la suma de  $n$  variables de distribución  $N(0, \sigma)$ :

$$\sum \ell \sim N(0, \sqrt{n}\sigma). \quad (4.3)$$

## 4 Algunas distribuciones

Si  $x$  sigue una distribución  $N(0,1)$ , la variable  $x^2$  sigue una distribución que recibe el nombre de  $\chi_1^2$ . La varianza de la distribución normal es, por definición, la esperanza de  $x^2$ , de modo que la media de  $\chi_1^2$  es 1. La varianza también es 1. Si sumamos  $n$  variables de tipo  $\chi_1^2$  obtenemos una variable que sigue una distribución  $\chi_n^2$ , cuya media es  $n$  al igual que su varianza.

Como estamos sumando distribuciones del mismo tipo, en virtud del teorema central del límite esta suma tiende a una distribución normal, y puesto que la media y la varianza de una  $\chi_n^2$  son iguales a  $n$ , resulta que

$$\lim_{n \rightarrow \infty} \chi_n^2 = N(n, \sqrt{n}).$$

Esta distribución,  $\chi^2$ , será la que sigan las varianzas estimadas. También surge cuando consideramos distancias a un punto. Si respecto a un punto los valores  $\Delta x$  y  $\Delta y$  siguen distribuciones normales  $(0,1)$ , entonces  $d^2 = \Delta x^2 + \Delta y^2$  sigue, según acabamos de ver, una distribución  $\chi_2^2$ . La distribución  $\chi_2^2$  es la más sencilla de las  $\chi^2$ :

$$\chi_2^2(x) = \frac{1}{2}e^{-x/2}, \quad X_2^2(x) = 1 - e^{-x/2}. \quad (4.4)$$

También es la más habitual en topografía, en donde se suelen considerar los resultados de precisión de planimetría y altimetría por separado (si se considerasen en conjunto se emplearía más  $\chi_3^2$ ).

Normalmente queremos la distribución de  $d$ , no de  $d^2$ . Si llamamos  $f$  y  $F$  respectivamente a las funciones de densidad y distribución de  $d$ , entonces  $F(d_0) = P(d \leq d_0) = P(d^2 \leq d_0^2) = X(d_0^2)$ . Y  $f$  la obtenemos derivando  $F$ .

$$f(d) = de^{-\frac{d^2}{2}}, \quad F(d) = 1 - e^{-\frac{d^2}{2}}. \quad (4.5)$$

Supongamos por ejemplo que  $\Delta x$  y  $\Delta y$  siguen cada una una distribución  $N(0,3)$ , y queremos saber cuál es la probabilidad de  $d < 6$ . Las variables  $\Delta x/3$  y  $\Delta y/3$  son normales  $(0,1)$ , por lo que  $(\Delta x/3)^2 + (\Delta y/3)^2 = (d/3)^2$  sigue una  $\chi_2^2$ . Así,  $P(d^2 < 6^2) = P((d/3)^2 < 2^2) = X(4)$  ó  $F(2) \approx 0,865$ .

También aparecerá en los problemas de estimación el conjunto de distribuciones  $t_n$  de Student, que se explicará en su momento.



## 5 Estimadores

Hemos visto que para un mismo problema podemos generar un número ilimitado de estimadores. Vamos a ver ahora algunos de los criterios habitualmente seguidos en la elección de los mismos.

Un criterio muy empleado es tomar la media aritmética, cuando se trata de varias medidas de una misma magnitud. La razón no suele ser otra que la intuición. Si queremos basarnos en un criterio riguroso tenemos que atender a la distribución. Empezaremos por la distribución más sencilla, la distribución uniforme.

### Distribuciones uniforme y triangular

Supongamos para las observaciones una distribución uniforme entre los valores  $l - h$  y  $l + h$ , para todos los  $l$ . Supongamos los valores observados ordenados de menor a mayor:  $l_1, l_2, \dots, l_m$ . Entonces está claro que  $l$  no puede ser menor que  $l_m - h$  ni mayor que  $l_1 + h$ . Entre estos dos valores la función  $g(l)$  es constante, ya que, para cada posible  $l$ , la probabilidad de haber obtenido los valores  $l_1, \dots, l_m$  es la misma. Del intervalo de posibles valores parece que lo mejor es tomar el punto medio, que es  $(l_1 + l_m)/2$ . Un poco más adelante veremos que, en un sentido, efectivamente es el mejor.

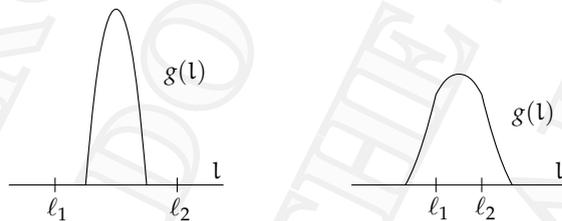
En el caso de distribución uniforme la media aritmética no es un buen estimador. Puede incluso dar lugar a un valor imposible. Supongamos por ejemplo que las medidas son  $l_1, l_2, l_3$ , con  $l_2, l_3$  iguales a  $l_1 + 1,6h$  y  $l_1 + 1,7h$  respectivamente. La media aritmética da como

## 5 Estimadores

resultado  $l = \ell_1 + 1,1h$ , lo que es imposible ya que en ese caso no podríamos haber obtenido  $\ell_1$ .

El siguiente paso en complejidad es la distribución triangular, que se puede obtener como suma de dos distribuciones uniformes. Recordemos que sumar dos distribuciones significa obtener la distribución de la observación (o en general cualquier variable aleatoria) que es suma de las dos observaciones originales. En este caso concreto lo que estamos diciendo es que si una observación se obtiene como suma de dos cantidades que siguen distribuciones uniformes entonces dicha observación sigue una distribución triangular.

Resulta que la distribución triangular no es tan sencilla. Incluso en el caso más sencillo de sólo dos medidas tomadas la distribución  $g$  depende de cómo de separadas estén esas medidas. Repetimos aquí las figuras 3.11 y 3.12:



Si las observaciones están relativamente separadas entonces la función  $g$  consta de un único tramo. Por el contrario, si están más próximas la función está formada por 3 tramos. En el caso de dos observaciones nada más, estas diferencias en las funciones  $g$  no importan a la hora de seleccionar el estimador, ya que de todas formas va a ser la media de los dos valores observados. Sin embargo, si el número de observaciones es tres o mayor, entonces las distintas formas de la función  $g$  sí que determinan distintos estimadores. Cuanto mayor sea el número de observaciones más complicada puede hacerse la función  $g$ . El número de tramos en su definición es igual al número de observaciones comprendidas entre los valores extremos posibles (que al igual que en la distribución uniforme quedan determinados por los valores máximo y mínimo observados) más uno. El valor máximo de la función  $g$  es el punto  $l$  tal que  $\sum 1/f_i(l)$  para los valores menores que  $l$  es igual a la correspondiente suma para los valores mayores. Si  $\ell_{\text{máx}} - \ell_{\text{mín}}$  está próximo al valor máximo posible, que será lo habi-

tual con muchas observaciones, el máximo de  $g$  se da aproximadamente en la media de los valores más extremos. Si por el contrario todas las observaciones se concentran en poca longitud, lo que puede suceder en algunas ocasiones con pocas observaciones, entonces el máximo se obtiene aproximadamente en la mediana de los valores medidos.

### Estimadores de máxima probabilidad

La búsqueda del máximo de  $g$  da lugar a los estimadores llamados de máxima probabilidad. Pero el criterio de máxima probabilidad es un criterio puntual. Traducido a términos prácticos significa que, si  $l$  es el resultado del estimador de máxima probabilidad, los intervalos de confianza para probabilidades muy pequeñas serán mínimos cuando el centro del intervalo es el valor  $l$ . Pero intervalos de confianza de, por ejemplo, el 5%, no sirven de nada. Por lo tanto, el mejor estimador respecto al criterio de máxima probabilidad es el que haga más pequeño un intervalo de una determinada probabilidad  $p$ , por ejemplo el 96%. Hay que buscar entonces el valor  $l$  para el cual el intervalo de probabilidad  $p$  tiene longitud mínima. Esto que puede parecer muy complicado no lo es tanto cuando buscamos intervalos unidimensionales, al menos teóricamente. El intervalo de menor longitud es aquél para el cual el valor de la función  $g$  en sus extremos es igual.

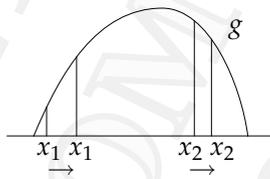


Fig. 5.3

Una manera de obtenerlo es la siguiente. Sea  $x_1$  el límite inferior del intervalo. Entonces en función de la probabilidad  $p$  queda determinado el límite superior,  $x_2$ . Si comenzamos dando a  $x_1$  un valor muy pequeño, entonces  $g(x_1)$  será menor que  $g(x_2)$  (fig. 5.3). A medida que vamos desplazando  $x_1$  hacia valores mayores, el intervalo se

## 5 Estimadores

va haciendo más estrecho, al abarcar una porción de la función  $g$  con valores mayores. Llegará un momento en el que  $g(x_1) = g(x_2)$ , y en ese punto  $x_2 - x_1$  es mínimo. Una vez hallado el intervalo situamos  $l$ . Si queremos que los intervalos sean de la forma  $[l - h, l + h]$  entonces será  $l = (x_1 + x_2)/2$ .

En más dimensiones obtener el intervalo mínimo (que será un área, un volumen, etc.) es más complicado. Siguiendo en una dimensión, si lo que estamos estimando es un factor de escala  $S$  o algo similar, entonces seguramente queremos que los intervalos sean de la forma  $[S/h, hS]$ , de donde  $S = \sqrt{x_1 x_2}$ . Pero además en este caso la longitud de los intervalos debería medirse por  $x_2/x_1$ , y el de mínima longitud ya no se dará cuando  $g(x_1) = g(x_2)$ , sino cuando  $x_1 g(x_1) = x_2 g(x_2)$ . Una manera de tratar los problemas de factores de escala es considerar la variable  $\log S$  en lugar de  $S$ . Respecto a la nueva variable la métrica a emplear es la habitual. De todos modos las diferencias sólo son apreciables cuando la longitud de los intervalos (es decir, la precisión con la que estamos estimando  $S$ ) es comparable al propio valor de  $S$ . En problemas topográficos la precisión de un factor de escala suele ser mucho menor que el propio valor. Sin embargo esto no es así cuando estimemos los valores de dispersión de las distribuciones (en general, la desviación típica), como veremos en capítulos posteriores. En topografía no es habitual dar intervalos de confianza para las desviaciones típicas estimadas, pero se debería.

El estimador de máxima probabilidad propiamente dicho puede interpretarse como el límite de los estimadores de máxima probabilidad para intervalos de probabilidad  $p$  cuando  $p \rightarrow 0$ .

### Estimadores centrados

El criterio de máxima probabilidad puede dar lugar a estimaciones por término medio desviadas del valor real. Es decir, que la media de los valores  $l - l$  no es cero. Cf. p. ej. la función  $g$  de (3.1) (figura 3.15), en la que el valor de máxima probabilidad es de los menores valores que se pueden esperar para  $l$ . Cuando sucede esto se dice que el estimador no es centrado.

El concepto de estimador centrado se puede entender según dos criterios distintos. Por una parte el problema directo. Para cada  $l$ , o en general para cada  $(l_1, \dots, l_m)$ , consideramos la distribución de los

valores estimados  $(l_1, \dots, l_m)$  de acuerdo al estimador seguido. Recordemos del capítulo 3 que los valores estimados son una variable aleatoria, al igual que las observaciones, y cuya distribución se obtiene a partir de la distribución de éstas y de la fórmula que da los valores estimados. Entonces, si se cumple

$$E[l_1, \dots, l_m] = (l_1, \dots, l_m), \quad (5.1)$$

y además esto se cumple para todas las  $m$ -uplas  $(l_1, \dots, l_m)$ , se dice que el estimador es centrado.

El problema inverso parte de las medidas  $(\ell_1, \dots, \ell_m)$  obtenidas. En todos los casos en los que se obtenga ese conjunto de valores observados los valores estimados serán los mismos, ya que son una función de los valores observados. Sean pues  $(l_1, \dots, l_m)$  los valores estimados correspondientes a  $(\ell_1, \dots, \ell_m)$ . Por el contrario los valores reales no serán los mismos, y en concreto su distribución viene dada por la función  $g(l_1, \dots, l_m)$ . Los errores cometidos en la estimación, cambiados de signo, serán  $(-\varepsilon_1, \dots, -\varepsilon_m) = (l_1 - \ell_1, \dots, l_m - \ell_m)^*$ .

Un estimador centrado, desde el punto de vista de los valores obtenidos, será aquél para el cual

$$E[l_1 - \ell_1, \dots, l_m - \ell_m] = (0, \dots, 0). \quad (5.2)$$

Ahora bien,  $(l_1, \dots, l_m)$  son cantidades fijas, por lo que la distribución de  $(l_1 - \ell_1, \dots, l_m - \ell_m)$  es la misma distribución  $g$  desplazada el vector  $(-l_1, \dots, -l_m)$ , de modo que

$$E[l_1 - \ell_1, \dots, l_m - \ell_m] = E[l_1, \dots, l_m] - (l_1, \dots, l_m),$$

en donde  $E[l_1, \dots, l_m]$  es la media de la distribución  $g$ .

Por lo tanto, si queremos que la esperanza sea cero el estimador ha de ser

$$(l_1, \dots, l_m) = E[l_1, \dots, l_m]. \quad (5.3)$$

Es decir,

$$l_1 = \frac{\int l_1 g}{\int g}, \text{ etc.}$$

---

\*Son valores reales desconocidos. No confundir con los residuos, que son valores conocidos  $v = \ell - l$ .

## 5 Estimadores

Las condiciones (5.1) y (5.2) son casi iguales, pero tienen una diferencia fundamental. En la primera los valores  $(l_1, \dots, l_m)$  son fijos, y las variables son  $(l_1, \dots, l_m)$ ; mientras que en la segunda los valores fijos son  $(l_1, \dots, l_m)$ , a través de  $(\ell_1, \dots, \ell_m)$ , y las variables son  $(l_1, \dots, l_m)$ .

Cuando nos refiramos a un estimador centrado sin especificar más entenderemos que el criterio al que se hace referencia es el inverso. Desde el punto de vista directo el estimador de máxima probabilidad de la función  $g$  de (3.1), cuando hay una única observación, es centrado, ya que  $l = \ell$ , y vimos para esa distribución que  $E[\ell] = l$ . Desde el punto de vista inverso está claro que no es centrado.

Determinar si un estimador es o no centrado desde el punto de vista directo es una cuestión simple: partimos de unos valores reales y de una distribución  $f$  para las observaciones; calculamos la distribución del estimador; calculamos su media y comparamos con los valores reales. Obtener un estimador centrado tampoco suele ser muy difícil.

Sin embargo, desde el punto de vista inverso es más complicado. Hay que obtener la función  $g$ , que ya vimos que puede ser difícil. Además depende en cada caso de la distribución  $h(l_1, \dots, l_m)$  que consideremos a priori, y por lo tanto no quedan determinados únicamente por las funciones  $f_{l_1, \dots, l_m}(l_1, \dots, l_m)$ , mientras que los estimadores centrados desde el punto de vista directo sí.

Por lo tanto, los estimadores centrados «directos» están asociados a cada tipo de distribución  $f$ , mientras que los estimadores centrados «inversos» dependen de cada caso.

### Magnitudes «grandes» y «pequeñas»

En topografía, al igual que en muchas otras disciplinas, hay a grandes rasgos dos tipos de magnitudes a estimar. Por un lado valores «grandes»: distancias, coordenadas, ángulos... , la medida de los cuales tiene una variabilidad muy pequeña en comparación con el propio valor, y la distribución  $f$  es simétrica, o podemos considerar que lo es. Además, cuando dependen unas de otras podemos suponer que las distribuciones siguen siendo todas simétricas, aunque estrictamente no sea así. Por ejemplo, si hay una variable  $x$ , y otra variable  $y = \sin x$ , entonces si la distribución de  $x$  es simétrica respecto a su

valor verdadero, la de  $y$  no lo será. Pero la función  $f(x)$  está concentrada en torno a  $x$  en un intervalo tan pequeño que en él la función  $\sin x$  prácticamente coincide con la recta tangente a ella misma en el punto  $x$ , y entonces  $y \cong ax + b$ , y podemos suponer que su distribución es simétrica y centrada en el valor verdadero.

El otro tipo son magnitudes «pequeñas», fundamentalmente residuos y desviaciones típicas u otras medidas de la dispersión. Pueden presentarse asimetrías, especialmente cuando el número de observaciones es pequeño.

En el primer grupo de magnitudes cualquier estimador «lógico» es centrado en sentido directo. En el segundo grupo hay que ser muy cuidadoso con los conceptos y distribuciones que se manejan, y con la interpretación de los resultados obtenidos.

### Parametrización de las magnitudes observadas

El hecho de que el conjunto de posibles  $(l_1, \dots, l_m)$  sea sólo un subconjunto del espacio  $m$ -dimensional, que además no suele ser sencillo, complica mucho la estimación, y en particular la obtención de la función  $g$ . Para evitar este problema se sustituyen muchas veces por un conjunto de  $n$  magnitudes (parámetros)  $(x_1, \dots, x_n)$  independientes entre sí y que determinen totalmente el problema, y de manera que la expresión de las  $l$  en función de  $(x_1, \dots, x_n)$  y viceversa sea sencilla. Parte de las magnitudes  $x$  o incluso todas pueden coincidir con magnitudes  $l$ . Este es el conocido como método paramétrico.

Toda la teoría vista se puede aplicar sin apenas modificación a los parámetros, ya que las relaciones entre magnitudes observadas y magnitudes parámetro son relaciones funcionales; es decir, unos valores coherentes de las magnitudes observadas determinan totalmente los valores de los parámetros y viceversa. Así, para cada  $(x_1, \dots, x_n)$  tenemos la distribución  $f_{x_1, \dots, x_n}(l_1, \dots, l_m) = f_{l_1, \dots, l_m}(l_1, \dots, l_m)$ . Obtenemos la función  $g$  de la misma manera, sólo que ahora se refiere a los parámetros,  $g(x_1, \dots, x_n)$ , en lugar de  $g(l_1, \dots, l_m)$  (en ambos casos el dominio de  $g$  es  $n$ -dimensional). En la función  $g$  podemos obtener intervalos de confianza, y basándonos en ella podemos obtener estimadores para los parámetros. El análisis efectuado acerca de los estimadores centrados tanto en el caso directo como en el inverso es de aplicación a los parámetros de manera idéntica.

## 5 Estimadores

Serán diferentes las distribuciones a priori. Si respecto a las magnitudes  $(l_1, \dots, l_m)$  consideramos que no sabemos nada, es decir, distribución  $h$  proporcional a una distribución uniforme y que por lo tanto podemos ignorar, entonces la distribución de los parámetros ya no equivaldrá a no conocer nada a priori, y habrá que considerar la distribución  $h$  a la hora de obtener la distribución  $g$ . En la práctica las magnitudes  $(l_1, \dots, l_m)$  serán de las que hemos llamado grandes, y entonces se puede suponer una relación lineal entre las  $x$  y las  $l$ , y la distribución de  $(x_1, \dots, x_n)$  también será proporcional a una uniforme y podemos seguir ignorándola.

Si un estimador diseñado para los parámetros da para cada  $(l_1, \dots, l_m)$  unos ciertos valores  $(x_1, \dots, x_n)$  estimados, entonces automáticamente también se obtienen los valores  $(l_1, \dots, l_m)$  correspondientes, y recíprocamente. Por lo tanto todo estimador de los parámetros es también un estimador de las magnitudes observadas, y viceversa.

Si un estimador es de máxima probabilidad o centrado (en cualquiera de los dos sentidos) respecto a las magnitudes  $(l_1, \dots, l_m)$ , no será en general de máxima probabilidad o centrado respecto a  $(x_1, \dots, x_n)$ , debido a que las funciones  $g$  son diferentes, así como las funciones  $h$ . Pero una vez más, si podemos suponer una relación lineal entre los dos conjuntos de magnitudes puede demostrarse que los estimadores de máxima probabilidad para  $(l_1, \dots, l_m)$  lo serán también para  $(x_1, \dots, x_n)$ , y lo mismo para los estimadores centrados, ya sean centrados directos o inversos.

### Estimador de máxima probabilidad para variables independientes. Ejemplos

Cuando la distribución se desplaza siguiendo los valores reales, entonces:

$$f_{l_1, \dots, l_m}(l_1, \dots, l_m) = f_{0, \dots, 0}(\ell_1 - l_1, \dots, \ell_m - l_m) = f_{0, \dots, 0}(\varepsilon_1, \dots, \varepsilon_m). \quad (5.4)$$

Si además las observaciones son independientes y siguen la misma distribución, y denotamos por  $f_0$  la distribución de cada una de ellas para  $l = 0$ , entonces

$$f = f_0(\varepsilon_1) \cdots f_0(\varepsilon_m).$$

Si además no consideramos ninguna función  $h$  entonces el estimador de máxima probabilidad es el que maximiza este producto para los errores estimados, es decir, los residuos:

$$(l_1, \dots, l_m) \text{ tal que } f_0(v_1) \cdots f_0(v_m) \text{ es máximo.} \quad (5.5)$$

Normalmente se trabajará con  $(x_1, \dots, x_n)$  en lugar de  $(l_1, \dots, l_m)$ , pero no conviene olvidar que en general este estimador es de máxima probabilidad para las observaciones, no para los parámetros.

Hay ocasiones en las que las observaciones no cumplen todos estos requisitos pero mediante cambios de variable adecuados se puede conseguir que los cumplan (v. cap. 8).

Para una distribución triangular de base 2, por ejemplo, hay que maximizar

$$\prod (1 - |v|),$$

con la restricción  $1 - |v_i| > 0$  para todo  $i$ .

$$\text{Si } f_0(\varepsilon) = e^{-|\varepsilon|},$$

$$\text{máx} \left( \prod e^{-|v|} \right) \equiv \text{máx} \left( \sum -|v| \right) \equiv \text{mín} \left( \sum |v| \right). \quad (5.6)$$

Y en general, si  $f_0(\varepsilon) = ke^{-\psi(\varepsilon)}$  hay que minimizar  $\sum \psi(v)$ . Así, si  $f_0$  es una distribución normal  $(0, \sigma)$  el estimador de máxima probabilidad es

$$\text{mín} \left( \sum v^2 / (2\sigma^2) \right) \equiv \text{mín} \left( \sum v^2 \right), \quad (5.7)$$

y recibe el nombre de estimador mínimo cuadrático.

Es fácil comprobar que el estimador (5.6) aplicado a un conjunto de medidas de una misma magnitud da como resultado la mediana, mientras que (5.7) genera la media.

La media aritmética es una de entre muchas medias. Según cada problema se podrá aplicar una media u otra en la estimación (media geométrica, media cuadrática...). Al respecto son útiles las desigualdades siguientes:

$$b < \sqrt{ab} < \frac{a+b}{2} < \sqrt{\frac{a^2+b^2}{2}} < a; \quad (5.8)$$

que se cumplen si  $a$  y  $b$  son dos números positivos,  $a > b$ .

## 5 Estimadores

Finalmente podemos resaltar que la mediana es un estimador muy robusto, lo que quiere decir que el resultado se ve muy poco afectado si algunas observaciones, por la razón que sea, han resultado muy erróneas. De hecho es el estimador más robusto, ya que no depende del valor de las observaciones extremas, sólo de la posición de la observación central, que a su vez estará determinada por los valores que tengan las observaciones centrales. Por lo tanto puede ser adecuado emplearlo en lugar del estimador que teóricamente correspondería cuando haya un gran número de observaciones muy erróneas (aunque en ese caso mejor sería dejar de considerar esas observaciones como excepcionales y emplear una distribución  $f$  que se ajuste a la realidad. No es superfluo recordar que no hay que echarle la culpa al modelo, sino a su incorrecta aplicación por parte del aplicador).

## 6 El teorema central del límite

El teorema central del límite (en realidad teorema del límite central) dice, más o menos, que si se van sumando variables aleatorias el resultado tiende a una distribución normal.

Una primera matización surge al pensar en algunos ejemplos. Si medimos a pasos grandes el ancho de una calle, supongamos que de bordillo a bordillo, contando cada paso igual a un metro, entonces podemos encontrar las siguientes causas de error: Error en la apreciación del paso grande medio, irregularidades en los pasos, error de redondeo (diremos que una calle tiene 6 ó  $6\frac{1}{2}$  pasos, pero no 6,17), error en la línea seguida (no será ni recta ni mínima), error en la definición del extremo del bordillo, hundimiento de la calzada al pisar, efecto del viento en los pasos, incorrecta apreciación de la posición relativa del último paso y el bordillo, debida por ejemplo a la refracción atmosférica entre el suelo y nuestros ojos, desgaste del bordillo desde que empezamos a medir hasta que terminamos, y todas las que se nos puedan ocurrir. Las últimas causas generan un error inapreciable, y evidentemente el error, y en consecuencia su distribución, estará determinado prácticamente por las tres primeras, especialmente las dos primeras, y si estas no son normales (la primera y principal seguro que no lo es) el error en la medida no será, ni aproximadamente, una distribución normal.

Por lo tanto habría que decir mejor, que si se van sumando variables aleatorias *de dispersión similar* el resultado tiende a una distribución normal. Esta condición está recogida de manera precisa en el enunciado riguroso del teorema.

## 6 *El teorema central del límite*

Este es un capítulo corto pero fundamental. Hemos visto que la suposición de distribución normal da lugar como estimador de máxima probabilidad al estimador mínimo cuadrático, y que en muchas ocasiones podemos suponer que también es centrado. En la mayoría de los problemas de ingeniería el estimador mínimo cuadrático es el único que se aplica. Se suele acudir al teorema central del límite como justificación, aunque la realidad es que la mayoría de los ingenieros es el único que saben aplicar, y el que todavía se acuerda lo justifica con el teorema central del límite. De todos modos el estimador mínimo cuadrático, aplicado en todos esos casos en los que estrictamente no se debería, no suele ser malo. Proporciona normalmente resultados buenos o cuando menos aceptables, es fácil de aplicar, en particular requiere menos cálculos que otros estimadores (esencial cuando los ordenadores eran más lentos que ahora, y todavía hoy en problemas muy grandes) y, sobre todo, es muy mecánico, lo que lo hace especialmente adecuado para los ingenieros.

Sin embargo existen muchos problemas en los que no es adecuado. En particular, los errores de medidas de coordenadas en fotogramas no siguen una distribución normal, y en bloques grandes las diferencias entre el estimador mínimo cuadrático y otro más adecuado pueden ser apreciables.

El caso opuesto son los errores de nivelación. La distribución del error de una nivelada será una en concreto, y un tramo se obtiene como suma de varias nivelaciones individuales. El teorema central del límite requiere que las distribuciones sean de dispersión similar, y un caso óptimo es que las distribuciones sean iguales. Si en cada nivelada el error de redondeo supone la mayor parte del error total, como suele suceder cuando se anota sólo hasta el milímetro, la distribución del error de cada nivelada será, salvo en los extremos (cuando la medida está casi exactamente en el medio milímetro), una distribución uniforme. Si no es así, esta distribución será más densa en torno al valor verdadero, aproximadamente con forma de campana. Incluso en el caso peor de distribución uniforme, que es completamente distinta de la normal, la composición de distribuciones iguales se aproxima muy rápidamente a la normal.

En muchos procesos de medida el número de causas en que podemos descomponer el error de manera que sean todas apreciables no es muy alto, por lo que el teorema central del límite no es de aplicación, o incluso si son muchas la distribución resultante puede estar

## 6 El teorema central del límite

todavía muy lejos de la normal. El teorema central del límite lo único que dice es que la distribución *tiende* a una normal, pero según cómo sean las distribuciones que se suman puede ser necesario tener que sumar muchas de ellas antes de que el resultado se empiece a parecer a una distribución normal. Además, donde la convergencia es más lenta es en las colas de la distribución, que es lo más característico de cada distribución, y lo que puede dar lugar a diferencias en los estimadores; en la zona central casi todas las distribuciones tienen forma acampanada. A este respecto, las colas de la distribución normal tienen una densidad que se aproxima muy rápidamente a cero en comparación con la mayor parte de las distribuciones que se suelen presentar.



# 7 El estimador mínimo cuadrático

## Hipótesis de aplicación

En el capítulo 5 se vio que bajo ciertas circunstancias el estimador de máxima probabilidad es el que hace mínima la suma de los cuadrados de los errores estimados. Estas condiciones son:

1. La función de densidad de las observaciones se desplaza con los valores reales:  $f_{l_1, \dots, l_m}(l_1, \dots, l_m) = f_{0, \dots, 0}(\varepsilon_1, \dots, \varepsilon_m)$ .
2. Las observaciones son independientes:  $f_{0, \dots, 0}(\varepsilon_1, \dots, \varepsilon_m) = f_{0_1}(\varepsilon_1) \cdots f_{0_m}(\varepsilon_m)$ .
3. Cada una de las observaciones sigue una distribución normal centrada en el valor real:  $f_{0_1} = N(0, \sigma_1)$ ,  $f_{0_2} = N(0, \sigma_2)$ , ...
4.  $\sigma_1 = \sigma_2 = \dots = \sigma_m$ .

Como veremos más adelante, el método de mínimos cuadrados se puede aplicar con apenas variación cuando las observaciones no cumplen estas condiciones pero otro conjunto  $(l'_1, l'_2, \dots, l'_n)$  cuyas componentes son combinaciones lineales de las originales sí las cumple. En particular, si la única condición que no se cumple es la cuarta el cambio de variable es inmediato.

El subconjunto  $n$ -dimensional de posibles  $(l_1, l_2, \dots, l_m)$  será en general complicado, y buscaremos en su lugar un conjunto de  $n$  parámetros  $(x_1, x_2, \dots, x_n)$  de manera que cualquier combinación de valo-

## 7 El estimador mínimo cuadrático

res de ellos sea posible.\*

El método de mínimos cuadrados lo aplicaremos a magnitudes de las que hemos llamado grandes, y podemos sustituir, al menos a efectos de la función  $g$ , las relaciones exactas entre observaciones y parámetros por relaciones lineales. En el estimador mínimo cuadrático esto tiene una ventaja adicional, ya que puede demostrarse que entonces el estimador de máxima probabilidad también es centrado desde el punto de vista directo, y si no consideramos ninguna función  $h$  también desde el punto de vista inverso (v. cap. 10). En ambos casos será centrado respecto a las observaciones y los parámetros; recordamos que si las relaciones son lineales, un estimador centrado en algún sentido respecto a las observaciones lo es también respecto a los parámetros en ese mismo sentido, y viceversa.

Si la dispersión en la distribución de valores observados no permite la aproximación lineal entonces es mejor buscar el estimador que dé un resultado centrado en lugar de aplicar el de máxima probabilidad.

### Planteamiento y linealización de las ecuaciones de residuo

Vamos a concretar todos los conceptos precedentes con un ejemplo. Sea un triángulo ABC en el que se miden los tres lados y la altura  $h_a$  (fig. 7.1). Se trata de obtener unos valores estimados de las dimensiones reales del triángulo. Suponemos que las distribuciones de los valores observados se ajustan a las hipótesis que permiten aplicar el método de mínimos cuadrados. Esta condición se dará por su puesta en todos los problemas de ahora en adelante siempre que no se diga lo contrario.

Supongamos los siguientes valores medidos:

$$\ell_1 = a = 19,095,$$

$$\ell_2 = b = 22,110,$$

---

\*Cualquier combinación en un entorno del valor ajustado. Por ejemplo, si tomamos entre los parámetros los lados  $a, b, c$  de un triángulo se ha de cumplir  $a < b + c$ , por lo que no todos los valores son posibles. No obstante, los valores imposibles estarán tan alejados del valor observado que podemos simplemente ignorarlos. De hecho, al sustituir las relaciones exactas entre magnitudes observadas y magnitudes parámetro por otras lineales estaremos admitiendo que trabajamos sólo en un entorno pequeño del punto de linealización, lo que es una restricción mucho más fuerte.

## 7 El estimador mínimo cuadrático

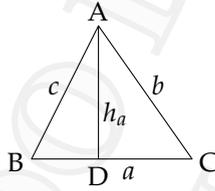


Fig. 7.1

$$\begin{aligned} \ell_3 &= c = 20,690, \\ \ell_4 &= h_a = 19,101. \end{aligned}$$

El número de magnitudes que determinan totalmente el problema es  $n = 3$ . Podemos tomar por ejemplo  $x_1 = h_a$ ,  $x_2 = d = BD$ ,  $x_3 = e = DC$ . Vemos que una de las magnitudes parámetro coincide con una magnitud observada y las otras dos no.

Si los valores reales de los parámetros son  $(x_1, x_2, x_3)$ , tenemos que

$$\begin{aligned} l_1 &= x_2 + x_3, \\ l_2 &= (x_1^2 + x_3^2)^{1/2}, \\ l_3 &= (x_1^2 + x_2^2)^{1/2}, \\ l_4 &= x_1. \end{aligned}$$

Y las mismas relaciones se cumplen entre cualquier conjunto de valores ajustados  $(x_1, x_2, x_3)$  y los correspondientes  $l_1, l_2, l_3, l_4$ .

La solución mínimo cuadrática será la que minimice la suma de los cuadrados de los residuos  $v = \ell - l$ .

$$\begin{aligned} v_1 &= \ell_1 - (x_2 + x_3), \\ v_2 &= \ell_2 - (x_1^2 + x_3^2)^{1/2}, \\ v_3 &= \ell_3 - (x_1^2 + x_2^2)^{1/2}, \\ v_4 &= \ell_4 - x_1. \end{aligned} \tag{7.1}$$

Para obtener el mínimo de la suma de los cuadrados de las cuatro cantidades habría que derivar dicha suma respecto de cada uno de los parámetros, obteniendo un sistema de tres ecuaciones con tres

## 7 El estimador mínimo cuadrático

incógnitas. Ese sistema se puede resolver de manera iterativa linealizando según se explicó en el capítulo primero. Aunque sencillo, resulta muy pesado obtener todas las derivadas. Suele ser mucho más simple linealizar directamente las ecuaciones (7.1). El resultado no es exactamente el mismo, pero la diferencia en los coeficientes obtenidos para los parámetros es del orden del propio residuo (cuando el punto de linealización se aproxima al punto solución) y no afecta a la convergencia.

En primer lugar hay que escoger un punto para linealizar cercano a la solución. Suele ser una tarea sencilla, ya que los valores medidos pueden ser valores conocidos anteriormente (que se vuelven a medir porque habrán sufrido pequeñas variaciones); en problemas geométricos se pueden obtener de un croquis; a veces se han medido directamente de una manera rápida y aproximada; etc.

La figura 7.1 representa el triángulo real, así que el lector puede tomar una regla y obtener los valores aproximados. Sólo tiene que calcular el factor de escala comparando las medidas de la regla con los valores observados. Los valores que obtendrá serán parecidos a los siguientes:  $x_1 \approx 19,10$ ,  $x_2 \approx 7,95$ ,  $x_3 \approx 11,14$ . Para estos valores no emplearemos ningún tipo de letra especial y los denotaremos mediante subíndice cero; así,  $x_{10} = 19,10$ , etc.

Hay que resaltar que, al coincidir la magnitud  $x_1$  con una magnitud observada, se podría haber tomado  $x_{10}$  igual al valor observado, pero no tiene por qué. Además en sucesivas iteraciones ya no será posible, puesto que el nuevo punto de linealización será igual al anterior más los incrementos obtenidos. La forma de la ecuación correspondiente a  $v_4$ , que es la más sencilla que puede existir, es la que se da cuando la magnitud observada coincide con un parámetro.

Se linealiza pues cada ecuación en el punto  $(x_{10}, x_{20}, x_{30})$  y se substituye en ella cada  $(x - x_0)$  por  $\Delta x$ , según se explicó en el capítulo 1. Veamos por ejemplo la segunda ecuación:

$$v_2 = \ell_2 - (x_1^2 + x_3^2)^{1/2} \approx \ell_2 - \left\{ \left( \sqrt{x_1^2 + x_3^2} \right)_0 + \left( \frac{x_1}{\sqrt{x_1^2 + x_3^2}} \right)_0 (x_1 - x_{10}) + \left( \frac{x_3}{\sqrt{x_1^2 + x_3^2}} \right)_0 (x_3 - x_{30}) \right\}.$$

Un subíndice cero aplicado a toda una expresión significa que ésta

## 7 El estimador mínimo cuadrático

se evalúa en el punto aproximado, es decir, que cada aparición de un parámetro toma el valor aproximado.

El segundo término de la ecuación linealizada es siempre el valor que resulta para la magnitud observada si los parámetros toman los valores aproximados; en este ejemplo,  $l_{20} = b_0$ .

$$v_2 \approx \ell_2 - \left\{ l_{20} + \left( \frac{x_1}{b} \right)_0 \Delta x_1 + \left( \frac{x_3}{b} \right)_0 \Delta x_3 \right\}.$$

Los valores  $l_{20}$  y  $b_0$  son los mismos, pero se suelen emplear como lo hemos hecho:  $l_{20}$  para el término independiente que se obtiene al linealizar y  $b_0$ , o en general la letra que empleemos para la magnitud en cuestión, cuando aparece en los coeficientes de los  $\Delta x$ . El término  $\ell_2 - l_{20}$ , valor observado menos valor aproximado (en este orden), se suele denotar también con la letra  $\ell$ , lo que es muy propenso a confusiones. Para evitarlas nosotros emplearemos una letra mayúscula pequeña:  $L_2$ . Por otra parte los coeficientes se denotan por  $a_{ij}$ , en donde el primer subíndice hace referencia al número de ecuación y el segundo al parámetro. Así, para la segunda ecuación tenemos

$$v_2 \approx L_2 - (a_{21}\Delta x_1 + a_{23}\Delta x_3).$$

Si sustituimos  $x_{10}, x_{20}, x_{30}$  por sus valores obtenemos los siguientes números:

$$\begin{aligned} l_{20} = b_0 &= 22,1113, & a_{21} &= 0,864, \\ L_2 &= -0,0013, & a_{23} &= 0,504. \end{aligned}$$

La primera ecuación es lineal, y lo único que hay que hacer es pasar de las variables  $x_2, x_3$  a  $\Delta x_2, \Delta x_3$ .

$$v_1 = \ell_1 - (x_2 + x_3) = \ell_1 - \{(x_2 + x_3)_0 + 1\Delta x_2 + 1\Delta x_3\}.$$

Los unos los hemos escrito para poner de manifiesto el paralelismo con una linealización general.

Una vez linealizadas todas las ecuaciones obtenemos las siguientes relaciones aproximadas:

$$\begin{aligned} v_1 &\approx 0,005 - (\Delta x_2 + \Delta x_3), \\ v_2 &\approx -0,0013 - (0,86\Delta x_1 + 0,50\Delta x_3), \\ v_3 &\approx 0,0015 - (0,92\Delta x_1 + 0,38\Delta x_2), \\ v_4 &\approx 0,001 - \Delta x_1. \end{aligned}$$

## 7 El estimador mínimo cuadrático

Que en el primer y último casos son exactas. Se suele emplear notación matricial:

$$V \approx L - AX_{\Delta}. \quad (7.2)$$

Se emplea normalmente  $X$  para el vector de incógnitas aunque más propiamente sería  $\Delta X$ , pero esto último puede interpretarse erróneamente como dos matrices en lugar de una sola. Alguna vez necesitaremos referirnos al vector  $X$  propiamente dicho, así que para el de incrementos emplearemos  $X_{\Delta}$ . Denotaremos por  $\mathcal{L}$  el vector de valores observados, por  $L_0$  el de valores aproximados, y  $L = \mathcal{L} - L_0$ . En nuestro ejemplo la matriz  $A$  es la siguiente:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0,86 & 0 & 0,50 \\ 0,92 & 0,38 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

### Coefficientes de $A$ y coeficientes de $L$

Es importante entender el significado de los elementos de la matriz  $A$ . Un elemento igual a 0,5, por ejemplo, quiere decir que si el parámetro al que acompaña varía una cierta cantidad, entonces el residuo correspondiente varía la mitad de esa cantidad. Es un error muy común pensar que los elementos de la matriz  $A$  hay que obtenerlos con una exactitud de muchos decimales, lo que además de innecesario es absurdo, ya que la linealización es sólo una aproximación a las relaciones reales. Se puede ver también de otra manera: Los parámetros serán unas magnitudes con unas unidades concretas, por ejemplo milímetros. No podemos saber antes de hacer los cálculos cuál será el valor de los  $\Delta x$ , pero si los valores aproximados están ya cercanos a la solución sí que sabremos aproximadamente en qué intervalo estarán; por ejemplo, estarán en torno a 10 mm o menos. Entonces una variación en el segundo decimal de un coeficiente de la matriz significa una variación en el residuo de  $0,01\Delta x \leq 0,1$ . Este valor normalmente será bastante menor que la precisión de la observación, y si no es así será porque el valor de hasta 10 mm esperado para los  $\Delta x$  es todavía muy grande en relación a la precisión de las observaciones, y el propio error inherente a la linealización hará necesaria otra iteración. Dos cifras significativas es suficiente (si el tamaño

## 7 El estimador mínimo cuadrático

de la matriz es muy grande, para evitar que se acumulen en exceso errores de redondeo se pueden emplear tres).

Por el contrario los valores  $l_0$  tienen que obtenerse con mayor precisión, ya que el término independiente en las relaciones lineales es  $\ell - l_0$ , que será mucho menor que cada uno de los dos. En el ejemplo que estamos siguiendo tenemos entre otros  $L_2 = 22,110 - 22,1113 = -0,0013$ . Dado que  $L = \ell - l_0$ , tomar un valor  $L$  con un pequeño error equivale a haber introducido el valor  $\ell$  con ese mismo error. Por lo tanto, si de acuerdo a la precisión de la observación no tiene sentido, pongamos por caso, distinguir entre el valor observado 18,3 y 18,33, entonces, ya que ambos valores se pueden emplear indistintamente, los podemos emplear indistintamente en particular al obtener  $L$ , y tampoco tiene sentido una segunda cifra decimal en  $L$ . Si por el contrario sí que es distinto a efectos prácticos 18,3 que 18,33, entonces también será necesaria en  $L$  la segunda cifra decimal. Visto de otro modo, si repitiésemos el conjunto de  $m$  observaciones varias veces podríamos realizar para cada una de ellas un ajuste. Los valores aproximados pueden ser los mismos en todos los casos, y en consecuencia también serán los mismos los elementos de la matriz  $A$ , y sólo variarán los valores  $(\ell_1, \dots, \ell_m)$ , y en exactamente la misma medida variará el vector  $(L_1, \dots, L_m)$ , ya que sus valores son las observaciones desplazadas la cantidad constante  $(l_{10}, \dots, l_{m0})$ ; es decir, que los distintos valores del vector  $L$ , determinados directamente por distintos valores de las observaciones, son lo que determina distintos resultados de varios ajustes para el mismo conjunto de magnitudes observadas, y deben introducirse en los cálculos con el mismo número de cifras decimales que consideremos necesario para los valores observados.

### Ecuaciones normales y resolución

El siguiente paso es obtener las derivadas de  $\sum v^2$  respecto de cada uno de los parámetros.

$$\begin{aligned}v_1^2 &\approx (L_1 - (a_{11}\Delta x_1 + a_{12}\Delta x_2 + \dots + a_{1n}\Delta x_n))^2, \\ \frac{\partial v_1^2}{\partial \Delta x_1} &\approx -2(L_1 - (a_{11}\Delta x_1 + \dots + a_{1n}\Delta x_n))a_{11} = \\ &= 2(a_{11}\Delta x_1 + \dots + a_{1n}\Delta x_n - L_1)a_{11}.\end{aligned}$$

## 7 El estimador mínimo cuadrático

Por lo tanto

$$\begin{aligned} \frac{\partial \sum v^2}{\partial \Delta x_1} &= \frac{\partial v_1^2}{\partial \Delta x_1} + \frac{\partial v_2^2}{\partial \Delta x_1} + \dots + \frac{\partial v_m^2}{\partial \Delta x_1} \approx \\ &\approx 2 \sum_{i=0}^m (a_{i1} \Delta x_1 + \dots + a_{in} \Delta x_n - L_i) a_{i1}. \end{aligned}$$

Escribiéndolo en forma matricial e igualando a cero,

$$(a_{11} \ a_{21} \ \dots \ a_{m1})(AX_{\Delta} - L) = 0.$$

Esta es la ecuación que resulta al derivar respecto del primer parámetro. En ella lo único que depende de la variable de derivación es el vector, que es la primera columna de la matriz  $A$ , traspuesta. Derivando respecto a un parámetro cualquiera  $\Delta x_j$  se obtiene una ecuación análoga, con la columna  $j$ -ésima en lugar de la primera. El sistema de  $n$  ecuaciones escrito en forma matricial es el siguiente:

$$A^T (AX_{\Delta} - L) = 0,$$

en donde  $0$  es el vector columna cero de  $n$  elementos. Por lo tanto la solución mínimo cuadrática cumple

$$A^T (AX_{\Delta} - L) \approx 0.$$

La matriz  $A^T A$  se denota por  $N$ , y el sistema recibe el nombre de sistema de ecuaciones normales:

$$NX_{\Delta} \approx A^T L, \quad (7.3)$$

$$X_{\Delta} \approx N^{-1} A^T L. \quad (7.4)$$

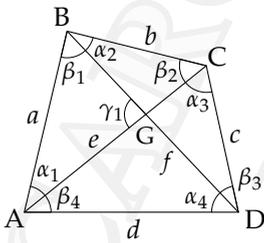
Y puesto que es sólo una solución aproximada puede ser necesario realizar otra iteración.

En el ejemplo del triángulo se llega a la solución

$$\begin{aligned} \Delta x_1 &= -0,0001, & x_1 &= 19,100, \\ \Delta x_2 &= 0,0061, & x_2 &= 7,956, \\ \Delta x_3 &= -0,0013. & x_3 &= 11,139. \end{aligned} \quad \Rightarrow$$

Y en función de estos parámetros ajustados se pueden obtener los valores ajustados  $a, b, c, h_a$  y los residuos  $v_1, \dots, v_4$ .

## 8 Cuadrilátero



### Lados y diagonales

Un cuadrilátero es todavía una figura bastante simple y que sin embargo permite un gran número de variantes teóricas, según cuáles sean los elementos que se escojan como medidas. Empezaremos por considerar que se miden los cuatro lados y las diagonales. Resulta fácil ver que son necesarios 5 parámetros para definir la figura (por ejemplo los segmentos  $d, a, f, e, c$ ). Nosotros tomaremos  $k = AG$ ,  $l = GC$ ,  $p = BG$ ,  $q = GD$  y  $\gamma_1$ . Las relaciones entre observaciones y parámetros ajustados son las siguientes:

$$\begin{aligned} a &= (k^2 + p^2 - 2kp \cos \gamma_1)^{1/2}, & e &= k + l, \\ b &= (p^2 + l^2 + 2pl \cos \gamma_1)^{1/2}, & f &= p + q. \\ c &= (l^2 + q^2 - 2lq \cos \gamma_1)^{1/2}, \\ d &= (q^2 + k^2 + 2qk \cos \gamma_1)^{1/2}, \end{aligned}$$

Las derivadas de las magnitudes  $a, b, c$  y  $d$  respecto de cada uno de los parámetros de los que dependen son similares. A continuación se muestran las de  $a$ :

$$\frac{\partial a}{\partial k} = \frac{k - p \cos \gamma_1}{a} \quad \frac{\partial a}{\partial p} = \frac{p - k \cos \gamma_1}{a} \quad \frac{\partial a}{\partial \gamma_1} = \frac{kp \sin \gamma_1}{a}.$$

## 8 Cuadrilátero

A partir de estas escribimos las aproximaciones lineales de los valores ajustados de las observaciones en función de los parámetros:

$$a \approx a_0 + \left( \frac{k - p \cos \gamma_1}{a} \right)_0 \Delta k + \left( \frac{p - k \cos \gamma_1}{a} \right)_0 \Delta p + \left( \frac{kp \sin \gamma_1}{a} \right)_0 \Delta \gamma_1,$$

etc.

Y las expresiones para los residuos

$$v_a = a - a_0 \approx (a - a_0) - \left\{ \left( \frac{k - p \cos \gamma_1}{a} \right)_0 \Delta k + \left( \frac{p - k \cos \gamma_1}{a} \right)_0 \Delta p + \left( \frac{kp \sin \gamma_1}{a} \right)_0 \Delta \gamma_1 \right\},$$

$$v_e = e - e_0 = (e - e_0) - (\Delta k + \Delta l),$$

$$v_f = f - f_0 = (f - f_0) - (\Delta p + \Delta q).$$

Las de  $v_b, v_c, v_d$  son parecidas a la de  $v_a$ .

Sustituyendo los valores aproximados de los parámetros obtenemos la matriz A, muchos de cuyos elementos valen 0, y el vector L.

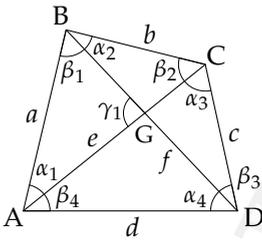
$$V \approx L - \begin{pmatrix} a_{11} & 0 & a_{13} & 0 & a_{15} \\ 0 & a_{22} & a_{23} & 0 & a_{25} \\ 0 & a_{32} & 0 & a_{34} & a_{35} \\ a_{41} & 0 & 0 & a_{44} & a_{45} \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} X_{\Delta}.$$

Y se obtiene  $N = A^T A$  y la solución  $X_{\Delta} \approx N^{-1} A^T L$ .

### Ángulos

La siguiente situación que vamos a estudiar es aquella en la que se miden todos los ángulos  $\alpha$  y  $\beta$ . La anterior elección de parámetros no resulta muy adecuada para este caso. En su lugar emplearemos coordenadas, pero de una manera muy simplificada. El origen de coordenadas lo situamos en el punto A, y podemos suponer que el punto D tiene coordenadas  $(1, 0)$ . Así, los únicos parámetros son

## 8 Cuadrilátero



$(x_b, y_b)$  y  $(x_c, y_c)$ . En este ejemplo, al no medir ninguna distancia, sólo establecemos las relaciones angulares de la figura, no el tamaño, y en consecuencia el número de parámetros disminuye en uno. Las ecuaciones de observación son las siguientes:

$$\alpha_1 = \arctan \frac{y_b}{x_b} - \arctan \frac{y_c}{x_c},$$

$$\alpha_2 = \arctan \frac{x_c - x_b}{y_b - y_c} - \arctan \frac{1 - x_b}{y_b},$$

$$\alpha_3 = \arctan \frac{x_c}{y_c} + \arctan \frac{1 - x_c}{y_c},$$

$$\alpha_4 = \arctan \frac{y_b}{1 - x_b};$$

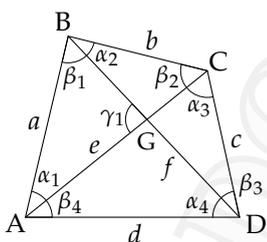
y unas similares para los ángulos  $\beta$ . Las derivadas de estas funciones se verán en el capítulo siguiente, en el que se tratarán de manera general ángulos y distancias, aunque lo único que hace falta conocer es la derivada de la función arcotangente.

### Ángulos y lados. Pesos

Supongamos ahora que medimos ángulos y lados. Los parámetros coordenada permiten expresar de manera sencilla tanto ángulos como distancias. Sin embargo ahora no podemos fijar libremente el tamaño de la figura. En particular no podemos tomar las coordenadas de D como  $(1, 0)$ . En su lugar serán  $(x_d, 0)$ ; es decir, que los parámetros son  $(x_d, x_b, y_b, x_c, y_c)$ . Podemos plantear las relaciones de observación para ángulos y para distancias, pero no podemos aplicar el estimador mínimo cuadrático, ya que las observaciones no cumplen las hipótesis.

En efecto, la aplicación del método exige  $\sigma_1 = \sigma_2 = \dots = \sigma_m$ , y en este problema se han medido tanto ángulos como distancias. Si los ángulos siguen una distribución normal con desviación típica  $\sigma_\alpha$  y las distancias otra con desviación típica  $\sigma_\delta$ , sólo se cumplirá  $\sigma_\alpha = \sigma_\delta$  de casualidad... O si expresamos cada conjunto de magnitudes en unas

## 8 Cuadrilátero



unidades tales que se cumpla exactamente  $\sigma_\alpha = \sigma_\delta$ , y esa es precisamente la solución.

Supongamos por ejemplo que los ángulos están expresados en grados centesimales, y  $\sigma_\alpha = 0^s 0010$ ; y que las longitudes están en metros y  $\sigma_\delta = 0,012$  m. Si multiplicamos todas las distancias por  $\sigma_\alpha / \sigma_\delta \approx 0,083$  obtendremos unos nuevos valores, que podemos interpretar como las mismas medidas

pero en un sistema de unidades anónimo para el cual  $\sigma_\delta = 0,001 = \sigma_\alpha$ . Las observaciones así normalizadas y todas las cantidades que de ellas se derivan son las únicas que se emplean en el ajuste, de modo que los nuevos valores los seguiremos denotando de la forma  $\ell$ , reservando la notación  $\ell'$  para cuando queramos distinguirlos de los originales. Así, tendremos los valores normalizados  $a', b', c', d'$ , para los cuales habrá unos valores ajustados  $a'', b'', c'', d''$  y unos residuos  $v'_a, v'_b, v'_c, v'_d$ , aunque el superíndice ' normalmente lo omitiremos. En las matrices nunca escribiremos el superíndice. Si es necesario distinguir indicaremos si se trata de matrices originales (or.) o normalizadas (no.).

Para obtener los coeficientes de las ecuaciones linealizadas normalizadas simplemente multiplicamos toda la ecuación por  $\sigma_\alpha / \sigma_\delta$ :

$$v_i \approx L_i - (a_{i1}\Delta x_1 + a_{i2}\Delta x_2 + \dots + a_{in}\Delta x_n).$$

⇓

$$v'_i \approx L'_i - (a'_{i1}\Delta x_1 + a'_{i2}\Delta x_2 + \dots + a'_{in}\Delta x_n).$$

En donde  $v'_i = v_i \frac{\sigma_\alpha}{\sigma_\delta}$ ,  $a'_{ij} = a_{ij} \frac{\sigma_\alpha}{\sigma_\delta}$  y  $L'_i = L_i \frac{\sigma_\alpha}{\sigma_\delta}$ .

Si hay varios conjuntos de observaciones con distintas desviaciones típicas en sus distribuciones:  $\sigma_\alpha, \sigma_\beta, \dots, \sigma_\lambda$ , entonces podemos tomar una de ellas como referencia,  $\sigma_\alpha$ , y escalar las demás medidas de manera que las nuevas variables tengan todas desviación típica  $\sigma_\alpha$ , multiplicando por las constantes  $1, \sigma_\alpha / \sigma_\beta, \dots, \sigma_\alpha / \sigma_\lambda$ . También podemos no tomar ninguna como referencia y dividir cada ecuación por la desviación típica de su observación, obteniendo ecuaciones de residuo correspondientes a observaciones siguiendo una distribución  $N(0,1)$ . En general, si tenemos unos residuos  $v_1, v_2, \dots, v_m$ , con sus correspondientes ecuaciones lineales aproximadas, entonces

## 8 Cuadrilátero

$v'_i = v_i \sigma_0 / \sigma_i$ , siendo  $\sigma_0 \neq 0$  una constante arbitraria (la misma para todas las ecuaciones):

$$v'_i = v_i \frac{\sigma_0}{\sigma_i}, \quad a'_{ij} = a_{ij} \frac{\sigma_0}{\sigma_i}, \quad L_i = L'_i \frac{\sigma_0}{\sigma_i}. \quad (8.1)$$

Supongamos que tenemos un conjunto de medidas pertenecientes todas a distribuciones  $N(0, \sigma)$  iguales. Supongamos que una de ellas la realizamos un número  $p$  de veces. Entonces podemos plantear las ecuaciones de residuo, y las  $p$  ecuaciones correspondientes a esas medidas serán iguales salvo por el término  $L$ . La media  $\bar{l}$  de los  $p$  valores sigue una distribución  $N(0, \sigma/\sqrt{p})$ , según se vio en (4.2). Puede demostrarse que el resultado de la estimación es el mismo si sustituimos las  $p$  observaciones por una única cuyo valor sea igual a la media. En ese caso, si tomamos  $\sigma$  como desviación típica de referencia, la ecuación del residuo  $v_{\bar{l}}$  habrá que multiplicarla por  $\sqrt{p}$ .

En general, si cada ecuación de residuo  $v_i$  la multiplicamos por  $\sigma_0/\sigma_i$  es como si todas las observaciones siguiesen distribuciones normales con  $\sigma = 1$  y cada una de ellas se hubiese repetido  $\sigma_0^2/\sigma_i^2$  veces.\*

Por ello estas cantidades reciben el nombre de pesos:

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}, \quad \sqrt{p_i} = \frac{\sigma_0}{\sigma_i}. \quad (8.2)$$

Si una observación tiene un peso igual a 1 es porque su varianza es igual a  $\sigma_0^2$ . Por ello la varianza de referencia recibe también el nombre de varianza unidad.

Si medimos los ocho ángulos de los vértices y los cuatro lados, y tomando como parámetros  $(x_d, x_b, y_b, x_c, y_c)$ , la forma de la matriz  $A^T$  se muestra en la página siguiente. Vemos fácilmente que  $a_{121} = 1$ . Las

---

\*Si las cantidades  $\sigma_0^2/\sigma_i^2$  son todas números enteros el *como si* es una afirmación rigurosa. Si son números racionales entonces podemos multiplicar cada ecuación por un común denominador (lo que equivale a escoger otro  $\sigma_0$ ) y obtenemos unas nuevas cantidades que son todas números enteros. Finalmente, si son irracionales, puesto que los números irracionales se pueden aproximar con un error arbitrariamente pequeño por números racionales, podemos encontrar números racionales que estén en una relación tan próxima como queramos a las cantidades  $1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_m$ , y podemos contemplar el problema como el caso límite en el que las observaciones se repiten respectivamente  $p_1, p_2, \dots, p_m$  veces cuando estos números enteros tienden a ser proporcionales a  $1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_m$ .

## 8 Cuadrilátero

$$\begin{array}{l}
 \alpha_1 \rightarrow \\
 \beta_1 \rightarrow \\
 \alpha_2 \rightarrow \\
 \beta_2 \rightarrow \\
 \alpha_3 \rightarrow \\
 \beta_3 \rightarrow \\
 \alpha_4 \rightarrow \\
 \beta_4 \rightarrow \\
 a \rightarrow \\
 b \rightarrow \\
 c \rightarrow \\
 d \rightarrow
 \end{array}
 \begin{array}{c}
 x_d \quad x_b \quad y_b \quad x_c \quad y_c \\
 \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\
 \left( \begin{array}{ccccc}
 0 & a_{12} & a_{13} & a_{14} & a_{15} \\
 a_{21} & a_{22} & a_{23} & 0 & 0 \\
 a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
 0 & a_{42} & a_{43} & a_{44} & a_{45} \\
 a_{51} & 0 & 0 & a_{54} & a_{55} \\
 a_{61} & a_{62} & a_{63} & a_{64} & a_{65} \\
 a_{71} & a_{72} & a_{73} & 0 & 0 \\
 a_{81} & 0 & 0 & a_{84} & a_{85} \\
 0 & a_{91} & a_{92} & 0 & 0 \\
 0 & a_{102} & a_{103} & a_{104} & a_{105} \\
 a_{111} & 0 & 0 & a_{114} & a_{115} \\
 a_{121} & 0 & 0 & 0 & 0
 \end{array} \right)
 \end{array}$$

derivadas de la distancia entre dos puntos respecto a las coordenadas de los puntos se verán en el capítulo siguiente.

Multiplicamos las ecuaciones de residuo de las distancias por  $\sigma_\alpha / \sigma_d$  (es decir, tomamos  $\sigma_0 = \sigma_\alpha$ ), obtenemos el sistema  $V = L - AX_\Delta$  normalizado y la solución  $X_\Delta = N^{-1}A^T L$ , y si es necesario realizamos más iteraciones.

La matriz  $N$  siempre será  $= A^T A$  (no.) ya que el producto  $A^T A$  (or.) no tiene ningún significado. Lo mismo se aplica a la matriz  $A^T L$ .

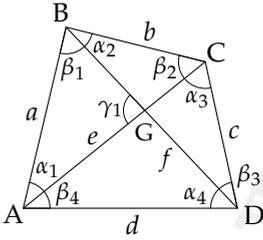
### Observaciones no independientes

Podría darse el caso de que algunas observaciones, de distribuciones con igual o distinta desviación típica, no fuesen independientes (pero se habrá de seguir cumpliendo que la distribución se desplace con los valores reales). Explicaremos cómo se resuelve este problema sin ningún ejemplo práctico porque es muy raro que el topógrafo tenga que aplicarlo alguna vez.

Si las observaciones en cuestión son  $\ell_1, \ell_2, \dots, \ell_k$ , se trata de obtener un conjunto  $\ell'_1, \ell'_2, \dots, \ell'_k$  de variables independientes y de la misma varianza, para tener los correspondientes residuos  $v'_1, v'_2, \dots, v'_k$  con las mismas propiedades. Necesitamos conocer la matriz de covarianzas del conjunto de observaciones:

$$\Sigma_{\ell\ell} = (\sigma_{ij}), \quad \sigma_{ij} = \sigma_{\ell_i \ell_j} = E[(\ell_i - \mu_{\ell_i})(\ell_j - \mu_{\ell_j})].$$

## 8 Cuadrilátero



Las nuevas variables serán combinaciones lineales de las originales. Hay muchos conjuntos de  $k$  combinaciones lineales de las variables  $\ell$  que dan lugar a  $k$  variables tales que  $\sigma_{ij} = 0$  si  $i \neq j$ . Si las nuevas variables tienen varianzas distintas  $\lambda_1, \lambda_2, \dots, \lambda_k$  falta solamente multiplicar cada una de ellas por  $1/\lambda_i$ , que podemos incluir en los coeficientes de la combinación lineal que ha dado lugar a esa variable. Por tanto existen muchas matrices  $C$  tales que las variables  $(\ell') = C(\ell)$  cumplen  $\Sigma_{\ell'\ell'} = I$ , siendo  $(\ell)$  y  $(\ell')$  vectores columna. Si las nuevas variables son independientes (no es suficiente con que  $i \neq j \Rightarrow \sigma_{ij} = 0$ ) y siguen distribuciones normales, entonces puede aplicarse el método de mínimos cuadrados. Más exactamente diríamos que el método de mínimos cuadrados sigue siendo el estimador de máxima probabilidad, ya que aplicarse se puede siempre.

Como se verá en el capítulo 10, las covarianzas de las variables  $\ell'$  las obtenemos a partir de las de  $\ell$  como sigue:

$$\Sigma_{\ell'\ell'} = C\Sigma_{\ell\ell}C^T.$$

Dado que  $\Sigma_{\ell'\ell'} = I$  se tiene entonces

$$\Sigma_{\ell\ell} = C^{-1}C^T^{-1}, \quad \Sigma_{\ell\ell}^{-1} = C^T C.$$

Esto los usaremos en lo que sigue.

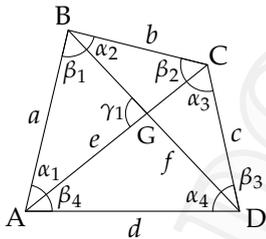
Tenemos que  $V_{(no.)} = CV_{(or.)} = CL - CAX_\Delta$  y por lo tanto

$$A_{(no.)} = CA_{(or.)}, \quad L_{(no.)} = CL_{(or.)}.$$

Pero no es necesario obtener estas matrices ni siquiera la matriz  $C$ , ya que

$$\begin{aligned} N &= (A^T A)_{(no.)} = (A^T C^T C A)_{(or.)} = A^T \Sigma_{\ell\ell}^{-1} A, \\ (A^T L)_{(no.)} &= (A^T C^T C L)_{(or.)} = A^T \Sigma_{\ell\ell}^{-1} L, \\ X_\Delta &= N^{-1} (A^T L)_{(no.)} = N^{-1} A^T \Sigma_{\ell\ell}^{-1} L. \end{aligned} \quad (8.3)$$

## 8 Cuadrilátero



### Método de las ecuaciones de condición

El cuadrilátero también nos vale para ilustrar el método de las ecuaciones de condición. En este método en lugar de expresar las magnitudes  $(l_1, \dots, l_m)$  en función de  $(x_1, \dots, x_n)$  se plantean las ecuaciones que relacionan los  $l_1, \dots, l_m$  entre sí; es decir, las condiciones que deben cumplir un conjunto de valores  $(l_1, \dots, l_m)$  para que puedan ser unos valores ajustados. El número de condiciones es  $r = m - n$ , y esta cantidad recibe el nombre de número de redundancias del problema. La limitación más grave de este método es que las ecuaciones de condición son particulares para cada problema, y pueden ser difíciles de obtener, mientras que el método paramétrico es muy fácil de sistematizar. En particular, en el capítulo siguiente veremos que un gran número de problemas de topografía se pueden resolver tomando como parámetros coordenadas. Para estudiar el método tomaremos como ejemplo el caso en el que se miden los ángulos  $\alpha$  y  $\beta$ .

El número de observaciones es  $m = 8$ , y el número de parámetros que serían necesarios para determinarlas todas es  $n = 4$ , y por lo tanto  $r = 8 - 4 = 4$ . Los ángulos opuestos por el vértice  $\gamma_1$  y  $\gamma_3$  son iguales, de donde

$$\alpha_1 + \beta_1 = \alpha_3 + \beta_3.$$

Por la misma razón  $\alpha_2 + \beta_2 = \alpha_4 + \beta_4$ . Además en el triángulo ABD obtenemos  $\alpha_1 + \beta_1 + \alpha_4 + \beta_4 = 180$ . Las relaciones análogas a esta que se pueden deducir de los triángulos ABC, BCD y ACD son combinaciones de las tres que ya tenemos, por lo que no expresan la cuarta relación que falta.

La manera de obtener la última relación es la siguiente: Aplicando el teorema del seno en el triángulo ABC podemos expresar el lado  $b$  en función del lado  $a$  y  $\alpha_1$  y  $\beta_2$ . A su vez en el triángulo BCD podemos expresar el lado  $c$  en función del lado  $b$  y de los ángulos  $\alpha_2$  y  $\beta_3$ , y en consecuencia en función del lado  $a$  y los ángulos  $\alpha_1$ ,  $\beta_2$ ,  $\alpha_2$  y  $\beta_3$ . La relación que obtenemos es la siguiente:

$$b = \frac{\text{sen } \alpha_1}{\text{sen } \beta_2} a, \quad c = \frac{\text{sen } \alpha_2}{\text{sen } \beta_3} b,$$

## 8 Cuadrilátero

$$c = \frac{\operatorname{sen} \alpha_1 \operatorname{sen} \alpha_2}{\operatorname{sen} \beta_2 \operatorname{sen} \beta_3} a.$$

Realizando el proceso análogo a través de los triángulos ABD, DAC se llega a

$$c = \frac{\operatorname{sen} \beta_4 \operatorname{sen} \beta_1}{\operatorname{sen} \alpha_3 \operatorname{sen} \alpha_4} a,$$

y en consecuencia

$$\frac{\operatorname{sen} \alpha_1 \operatorname{sen} \alpha_2 \operatorname{sen} \alpha_3 \operatorname{sen} \alpha_4}{\operatorname{sen} \beta_1 \operatorname{sen} \beta_2 \operatorname{sen} \beta_3 \operatorname{sen} \beta_4} = 1. \quad (8.4)$$

La ecuación (8.4) recibe el nombre de ecuación de lado. Las ecuaciones de relación se linealizan y se escriben de la forma  $b_{i1} \Delta l_1 + \dots + b_{im} \Delta l_m - d_i = 0$ . Los valores aproximados de las observaciones que se toman como punto de linealización no tienen que formar necesariamente un conjunto coherente; serán unos cualesquiera y los que habrán de ser coherentes serán los  $\Delta l$ , cuyas relaciones a cumplir quedan expresadas mediante las ecuaciones de condición. Puesto que vamos a trabajar con razones trigonométricas y sus derivadas es mejor expresar los ángulos en radianes, y una vez terminado el ajuste pasar los resultados a grados. Se puede trabajar con grados multiplicando las derivadas por la constante adecuada, pero este modo de operar es muy propenso a error. La primera ecuación queda como sigue:

$$\begin{aligned} \alpha_{10} + \Delta \alpha_1 - \alpha_{30} - \Delta \alpha_3 + \beta_{10} + \Delta \beta_1 - \beta_{30} - \Delta \beta_3 &= 0, \\ \Delta \alpha_1 - \Delta \alpha_3 + \Delta \beta_1 - \Delta \beta_3 + (\alpha_1 - \alpha_3 + \beta_1 - \beta_3)_0 &= 0; \end{aligned}$$

y la tercera:

$$\Delta \alpha_1 + \Delta \alpha_4 + \Delta \beta_1 + \Delta \beta_4 + ((\alpha_1 + \alpha_4 + \beta_1 + \beta_4)_0 - 2\pi) = 0.$$

La cuarta ecuación es de la forma  $p/q = 1$ , que podemos escribir como  $p - q = 0$ , y podemos linealizar por separado  $p$  y  $q$ . Sea  $p_0 = (\operatorname{sen} \alpha_1 \operatorname{sen} \alpha_2 \operatorname{sen} \alpha_3 \operatorname{sen} \alpha_4)_0$ , y análogamente  $q_0$ , entonces

$$\begin{aligned} p \approx p_0 + (\cos \alpha_1 \operatorname{sen} \alpha_2 \operatorname{sen} \alpha_3 \operatorname{sen} \alpha_4)_0 \Delta \alpha_1 + \dots \\ + (\operatorname{sen} \alpha_1 \operatorname{sen} \alpha_2 \operatorname{sen} \alpha_3 \cos \alpha_4)_0 \Delta \alpha_4, \end{aligned}$$

## 8 Cuadrilátero

que podemos escribir como

$$p \approx (\cotg \alpha_1 p)_0 \Delta \alpha_1 + (\cotg \alpha_2 p)_0 \Delta \alpha_2 + (\cotg \alpha_3 p)_0 \Delta \alpha_3 + (\cotg \alpha_4 p)_0 \Delta \alpha_4 + p_0,$$

y la cuarta ecuación linealizada queda de la forma

$$(\cotg \alpha_1 p)_0 \Delta \alpha_1 + \dots + (\cotg \alpha_4 p)_0 \Delta \alpha_4 - (\cotg \beta_1 q)_0 \Delta \beta_1 - \dots - (\cotg \beta_4 q)_0 \Delta \beta_4 + (p_0 - q_0) \approx 0.$$

El sistema de ecuaciones queda de la siguiente forma

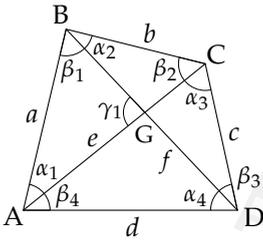
$$\begin{pmatrix} 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} & b_{47} & b_{48} \end{pmatrix} \begin{pmatrix} \Delta \alpha_1 \\ \Delta \alpha_2 \\ \Delta \alpha_3 \\ \Delta \alpha_4 \\ \Delta \beta_1 \\ \Delta \beta_2 \\ \Delta \beta_3 \\ \Delta \beta_4 \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

que denotaremos de manera compacta por  $B\Delta + d \approx 0$ .

La función  $\sum v^2$  tiene una representación bastante sencilla. En una dimensión es una parábola, si son dos el número de observaciones es un paraboloides de revolución, y para un número mayor de observaciones una figura análoga. En cada punto del espacio  $m$ -dimensional el vector gradiente es el vector de derivadas parciales, es decir,  $(2v_1, 2v_2, \dots, 2v_m)$ . Hemos tomado un conjunto de valores aproximados para las magnitudes observadas, que serán distintos de los observados (sólo podremos tomarlos iguales en la primera iteración), y el residuo de una observación es  $v = \ell - l = \ell - (l_0 + \Delta l) = v_0 - \Delta l$ . Entonces el vector gradiente en función de las variables  $\Delta l$  es  $2((v_{10}, \dots, v_{m0}) - (\Delta l_1, \dots, \Delta l_m))$ .

El conjunto de posibles  $(\Delta l_1, \dots, \Delta l_m)$  es un subconjunto  $n$ -dimensional, cuya proyección sobre la gráfica  $\sum v^2 = \sum (v_0 - \Delta l)^2$  es el conjunto de posibles puntos  $(\Delta l_1, \dots, \Delta l_m, \sum v^2)$ , y es también  $n$ -dimensional. Por ejemplo, si  $m = 2$  y  $n = 1$  la gráfica de  $\sum v^2$  es un paraboloides y el conjunto de posibles valores  $(\Delta l_1, \Delta l_2)$  una línea. La proyección vertical de esta línea sobre el paraboloides es una línea sita

## 8 Cuadrilátero



en la superficie, y es el conjunto de posibles  $(\Delta l_1, \Delta l_2, v_1^2 + v_2^2)$ . En el punto en que esta línea alcanza su mínimo la recta tangente a ella es horizontal, paralela a la tangente a la línea de  $(\Delta l_1, \Delta l_2)$  posibles (simplemente está desplazada en vertical), y sigue una dirección perpendicular al gradiente de la superficie en ese punto. En general, cuando  $\sum v^2$  alcanza su mínimo dentro

del conjunto de valores posibles el subespacio tangente al subconjunto  $n$ -dimensional de posibles  $(\Delta l_1, \dots, \Delta l_m)$  es perpendicular al gradiente de la función en el punto.

No trabajamos con el conjunto exacto de posibles  $(\Delta l_1, \dots, \Delta l_m)$ , sino con su aproximación lineal dada por las ecuaciones  $B\Delta + d \approx 0$ . Al ser lineal el conjunto aproximado, él mismo es su propia tangente en cualquier punto, y por tanto el mínimo se dará en un punto en el que el gradiente de  $\sum v^2$  sea perpendicular al subespacio determinado por  $B\Delta + d = 0$ . Cada ecuación de este sistema representa un subespacio  $(m - 1)$ -dimensional (hiperplano), cuya perpendicular es la recta definida por el vector  $(b_{i1}, \dots, b_{im})$ . Los posibles  $\Delta$  son la intersección de los  $r$  hiperplanos, y el conjunto de vectores perpendiculares son las combinaciones lineales de los vectores perpendiculares a cada uno de los hiperplanos (por lo tanto, según se reduce la dimensión de los posibles  $\Delta$  aumenta la de su espacio perpendicular. Así, en tres dimensiones el espacio perpendicular a un plano es una recta, y el perpendicular a una recta es un plano).

Por lo tanto en el mínimo se cumple

$$\begin{pmatrix} v_{10} - \Delta l_1 \\ v_{20} - \Delta l_2 \\ \vdots \\ v_{m0} - \Delta l_m \end{pmatrix} \approx k_1 \begin{pmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{1m} \end{pmatrix} + k_2 \begin{pmatrix} b_{21} \\ b_{22} \\ \vdots \\ b_{2m} \end{pmatrix} + \dots + k_r \begin{pmatrix} b_{r1} \\ b_{r2} \\ \vdots \\ b_{rm} \end{pmatrix}$$

para algunos valores reales  $k_1, \dots, k_m$ . En forma matricial se escribe  $V_0 - \Delta \approx B^T K$ .

Uniendo estas ecuaciones a las de condición obtenemos un sistema de  $m + r$  ecuaciones con  $m + r$  incógnitas:

$$\begin{aligned} V_0 - \Delta &\approx B^T K, \\ B\Delta + d &\approx 0. \end{aligned} \tag{8.5}$$

## 8 Cuadrilátero

En el primer grupo de ecuaciones expresamos  $\Delta$  en función de  $K$ :

$$\Delta \approx V_0 - B^T K. \quad (8.6)$$

Sustituyendo en el segundo, y llamando  $N_B$  a  $BB^T$ , obtenemos

$$K \approx N_B^{-1}(BV_0 + d), \quad (8.7)$$

y sustituyendo en (8.6) obtenemos la solución  $\Delta$ .

Si las observaciones pertenecen a distribuciones con distintos  $\sigma$  trabajamos al igual que antes con variables  $\ell'$ , y el vector de incrementos incógnita queda referido a estas variables:  $\Delta'$ . Las ecuaciones de condición expresan unas relaciones que han de cumplir los  $\Delta l$ . Al sustituir cada  $\Delta l_i$  por  $\Delta l'_i = (\sigma_0/\sigma_i)\Delta l_i$ , el coeficiente que acompaña habrá de dividirse por  $\sigma_0/\sigma_i$ , de modo que si antes un conjunto  $(\Delta l_1, \Delta l_2, \dots, \Delta l_m)$  cumplía las ecuaciones ahora las cumple el conjunto  $(\sigma_0/\sigma_1 \Delta l_1, \sigma_0/\sigma_2 \Delta l_2, \dots, \sigma_0/\sigma_m \Delta l_m)$ . Por lo tanto hay que dividir cada columna de la matriz  $B$  por el correspondiente  $\sigma_0/\sigma_i$ .

$$B_{(or.)}\Delta_{(or.)} + d \approx 0 \quad \Rightarrow \quad B_{(no.)}\Delta_{(no.)} + d \approx 0.$$

La función que hay que minimizar es  $\sum v'^2$ , y la condición de ortogonalidad del gradiente en el punto mínimo es

$$V_0 - \Delta \approx B^T K \quad (no.),$$

siendo  $v'_{i0} = v_{i0} \cdot \sigma_0/\sigma_i$ .

Si las observaciones no son estadísticamente independientes el proceso para obtener la nueva solución es similar al del método paramétrico, quedando la solución expresada en función de  $\Sigma_{\ell\ell}^{-1}$  al igual que antes.

### Caracterización $A^T V = 0$ de los valores ajustados

Volviendo al método paramétrico, si cuando realizamos el ajuste los valores aproximados coinciden con los ajustados el vector de incrementos  $(\Delta x_1, \Delta x_2, \dots, \Delta x_n)$  que obtendremos será 0, puesto que ya estamos en la solución:  $X_\Delta = N^{-1}A^T L = 0$ . En la matriz  $N^{-1}$  ninguna columna es combinación lineal de otras, ya que de no ser

así no sería invertible, y lo es. Entonces para que se anule el vector  $X_{\Delta}$  se ha de cumplir  $A^T L = 0$ . La expresión  $L - AX_{\Delta}$  es la aproximación lineal por la que hemos sustituido las relaciones exactas entre parámetros y magnitudes observadas. Si modificamos los valores aproximados, respecto a los cuales expresamos  $X_{\Delta}$ , tendremos  $X_{\Delta \text{ antiguo}} = X_{\Delta \text{ nuevo}} - \Delta X_0$ , y el vector  $L$  se modifica la cantidad  $-A\Delta X_0$ , pero la matriz  $A$  permanece constante. De modo que la propiedad  $A^T L = 0$  depende del vector  $L$ . Además, si  $X_0$  son los valores ajustados entonces el vector  $L$  son los residuos  $v$ . Así,

(8.8) *Los valores ajustados son aquéllos para los cuales se cumple  $A^T V = 0$ .*

Esta propiedad define completamente la solución mínimo cuadrática. Sin embargo hay  $n$  ecuaciones para  $m$  incógnitas. Las  $r = m - n$  ecuaciones que faltan son las ecuaciones de condición, y combinando los dos sistemas obtenemos uno único formado en una parte por la matriz  $A^T$  y en otra por la matriz  $B$ . Para expresar todas las ecuaciones en función de  $\Delta$  (variables  $\Delta l$ ), escribimos  $v = \ell - l = \ell - (l_0 + \Delta l) = -\Delta l + L$ , de donde

$$A^T V = 0 \quad \Rightarrow \quad A^T \Delta + d = 0,$$

siendo  $d = -A^T L$ . El sistema de  $m$  ecuaciones con  $m$  incógnitas es

$$\begin{pmatrix} A^T \\ B \end{pmatrix} \Delta = -d.$$

### Sobre los métodos paramétrico y de condición

El ejemplo del cuadrilátero en el que se miden los ocho ángulos es muy típico para ilustrar el método de las ecuaciones de condición, y sin embargo vimos que se puede resolver de manera muy sencilla por el método paramétrico, mientras que de la otra forma entraña el planteamiento de la ecuación de lado, que no es en absoluto evidente. En problemas topográficos y fotogramétricos no he encontrado todavía ninguno que se resuelva de manera más sencilla mediante ecuaciones de condición que por el método paramétrico.



## 9 Ángulos y distancias

### Ecuación de distancia

Los problemas en los que se miden ángulos y distancias se pueden resolver fácilmente y de modo sistemático tomando como parámetros coordenadas de puntos.

La distancia entre dos puntos de coordenadas  $(x_1, y_1)$ ,  $(x_2, y_2)$  es

$$S = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Las derivadas son las siguientes:

$$\begin{aligned} \frac{\partial S}{\partial x_1} &= -\frac{x_2 - x_1}{S}, & \frac{\partial S}{\partial x_2} &= \frac{x_2 - x_1}{S}, \\ \frac{\partial S}{\partial y_1} &= -\frac{y_2 - y_1}{S}, & \frac{\partial S}{\partial y_2} &= \frac{y_2 - y_1}{S}. \end{aligned}$$

De modo que

$$v_S \approx (S - S_0) - \left\{ \left( \frac{x_2 - x_1}{S} \right)_0 \Delta x_1 - \left( \frac{y_2 - y_1}{S} \right)_0 \Delta y_1 + \left( \frac{x_2 - x_1}{S} \right)_0 \Delta x_2 + \left( \frac{y_2 - y_1}{S} \right)_0 \Delta y_2 \right\}. \quad (9.1)$$

## 9 Ángulos y distancias

### Ecuación de lectura

En topografía no se suelen observar directamente ángulos, sino el ángulo que una dirección forma respecto a un origen fijo para la estación. Este origen es la desorientación  $\Sigma$ , que suele ser desconocida. Por lo tanto por cada punto habrá en general tres parámetros:  $x, y, \Sigma$ . El ángulo de la dirección respecto al sentido positivo del eje Y, y medido en sentido horario, es el azimut  $\theta$ , y el ángulo respecto al origen de la estación, la lectura L. La desorientación es el azimut del origen de ángulos de la estación (fig. 9.1).

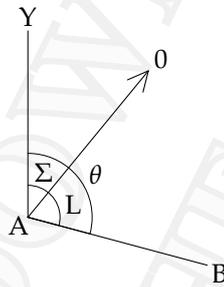


Fig. 9.1

La expresión para L es entonces la siguiente:\*

$$L_1^2 = \theta_1^2 - \Sigma_1 = \arctan \frac{x_2 - x_1}{y_2 - y_1} - \Sigma_1;$$

y las derivadas:

$$\frac{\partial L}{\partial x_1} = \frac{-\frac{1}{y_2 - y_1}}{1 + \left(\frac{x_2 - x_1}{y_2 - y_1}\right)^2} = -\frac{y_2 - y_1}{(y_2 - y_1)^2 + (x_2 - x_1)^2} = -\frac{y_2 - y_1}{S^2},$$

$$\frac{\partial L}{\partial x_2} = \frac{y_2 - y_1}{S^2},$$

$$\frac{\partial L}{\partial y_1} = \frac{+\frac{x_2 - x_1}{(y_2 - y_1)^2}}{1 + \left(\frac{x_2 - x_1}{y_2 - y_1}\right)^2} = \frac{x_2 - x_1}{(y_2 - y_1)^2 + (x_2 - x_1)^2} = \frac{x_2 - x_1}{S^2},$$

\* $L_1^2$  quiere decir lectura del punto 1 al punto 2

$$\frac{\partial L}{\partial y_2} = -\frac{x_2 - x_1}{S^2},$$

$$\frac{\partial L}{\partial \Sigma_1} = -1.$$

De donde la ecuación de residuo es

$$v_L \approx (L - L_0) - \left\{ \left( \frac{y_2 - y_1}{S^2} \right)_0 \Delta x_1 + \left( \frac{x_2 - x_1}{S^2} \right)_0 \Delta y_1 + \left( \frac{y_2 - y_1}{S^2} \right)_0 \Delta x_2 - \left( \frac{x_2 - x_1}{S^2} \right)_0 \Delta y_2 - \Delta \Sigma_1 \right\}, \quad (9.2)$$

siendo  $L_0 = \theta_0 - \Sigma_0$ .

### Coordenadas fijas para definir el sistema. Ecuaciones de azimut y coordenada

En las ecuaciones (9.1) y (9.2) alguno de los parámetros puede no existir. Si el número de puntos del ajuste es  $k$ , entonces el número máximo posible de parámetros es  $n = 3k$ . Sin embargo normalmente habrá que fijar algunos de ellos. Si sólo se han medido ángulos tenemos que fijar la figura en posición, escala y orientación. Una manera de hacerlo es eligiendo las coordenadas de dos puntos, por ejemplo  $P_1 = (0, 0)$ ,  $P_2 = (1, 0)$ . Puede ser que en este sistema algunos puntos tengan coordenadas negativas. Para evitarlo, y para distinguir fácilmente una coordenada  $x$  de una  $y$ , podemos tomar, pongamos por caso,  $P_1 = (450, 1050)$ ,  $P_2 = (460, 1050)$ . También podemos fijar las coordenadas y desorientación de un punto,  $(x_1, y_1, \Sigma_1)$ , y la coordenada  $x$  (o la  $y$ ) de otro punto. No se pueden fijar las desorientaciones de dos puntos.

Si se han medido distancias, bien con ángulos también medidos o bien sin ellos, el parámetro a fijar es uno menos (no podemos escoger el tamaño). Es habitual tomar las coordenadas y desorientación de un punto, aunque también se puede operar como hicimos en el ejemplo del cuadrilátero, en el que fijamos las dos coordenadas de un punto y una de otro.

Puede suceder que la fijación de la figura en posición, orientación y tamaño venga dada por un conjunto de puntos con coordenadas fijas. Por ejemplo, cuando se observa una red partiendo de unos puntos de una red superior.

## 9 Ángulos y distancias

También, aunque más raro, puede haber alguna visual de azimut conocido  $\theta$ . En ese caso la ecuación de residuo es muy sencilla:

$$v_L = L - (\theta - \Sigma) = L - (\theta - (\Sigma_0 + \Delta\Sigma)) = (L - L_0) - \Delta\Sigma, \quad (9.3)$$

siendo  $L_0 = \theta - \Sigma_0$ . En estos casos no se puede fijar libremente la orientación.

Finalmente, puede haber coordenadas observadas directamente, por ejemplo mediante GPS, o bien con valores conocidos de trabajos anteriores y que queramos tener en consideración. Se trata del caso particular en el que una magnitud observada coincide con un parámetro, y entonces la ecuación de residuo es la más sencilla posible:

$$v_x = (x - x_0) - \Delta x, \quad v_y = (y - y_0) - \Delta y. \quad (9.4)$$

Si no hay puntos con coordenadas fijas, es decir, si tenemos que fijar libremente al menos las coordenadas de un punto, entonces un único punto con coordenadas observadas sirve para situar la figura, pero no aporta nada al ajuste, simplemente sus dos coordenadas estarán obligatoriamente entre los parámetros fijos (se pueden incluir en el ajuste, pero los valores ajustados serán lógicamente iguales a los observados y los demás valores ajustados serán exactamente los mismos que si se hubiesen dejado fijas). Situaciones similares se producen si sólo hay una visual con azimut conocido, lo que equivale a medir directamente una desorientación; si además de esta visual sólo hay un punto con coordenadas observadas; o si, no habiendo medido distancias, hay sólo dos puntos con coordenadas observadas. En todos estos casos los parámetros observados no superan los que son necesario fijar arbitrariamente, y lo único que sucede es que esa fijación ya no se puede hacer arbitrariamente, sino mediante los valores observados, pero no aportan nada al ajuste.

### Un ejemplo

A continuación desarrollamos y calculamos un ejemplo, que consiste en la obtención de las coordenadas y orientación de un punto mediante intersección inversa, disponiendo además de coordenadas medidas directamente con una cierta precisión y una visual a una dirección con azimut conocido.

## 9 Ángulos y distancias

Sea P el punto de estación, desde el cual se han observado los puntos 1, 2, 3, 4 y 5, una visual 6 de azimut conocido y unas observaciones  $X_P, Y_P$ . Los valores conocidos son los siguientes:

Punto	X	Y
1	44 830,13	14 211,55
2	46 645,07	14 600,79
3	47 481,19	12 822,10
4	47 070,72	10 131,40
5	45 012,96	10 248,64

$$\theta_P^6 = 398^{\circ}9963$$

Y las observaciones con sus precisiones:

Obs.	Valor	Precisión
L <sup>1</sup>	381 <sup>s</sup> 1860	0 <sup>s</sup> 0070
L <sup>2</sup>	38 <sup>s</sup> 6589	"
L <sup>3</sup>	106 <sup>s</sup> 9395	"
L <sup>4</sup>	193 <sup>s</sup> 7038	"
L <sup>5</sup>	246 <sup>s</sup> 7703	"
L <sup>6</sup>	19 <sup>s</sup> 1215	"
X <sub>P</sub>	46 020,60	0,1
Y <sub>P</sub>	12 515,30	"

Sólo hay observadas las coordenadas de un punto y una dirección con azimut conocido, pero en este caso sí que aportan observaciones para el ajuste, ya que la fijación de la figura ya está dada a través de las coordenadas de los puntos del 1 al 5.

Los únicos parámetros a calcular son  $X_P, Y_P, \Sigma_P$ , al último de los cuales llamaremos simplemente  $\Sigma$ . Como valores aproximados para las coordenadas podemos emplear los propios observados, pero para mostrar que no tiene por qué ser así tomaremos  $X_{P0} = 46\ 020,50$ ,  $Y_{P0} = 12\ 515,30$ . Para la desorientación podemos tomar el que se obtiene como  $\theta^6 - L^6$ :  $\Sigma_0 = 379^{\circ}8748$ . Puesto que vamos a trabajar con razones trigonométricas y sus derivadas pasaremos todos los valores angulares a radianes:

## 9 Ángulos y distancias

Obs.	Valor	Precisión
L <sup>1</sup>	5,987 656	0,000 110
L <sup>2</sup>	0,607 253	"
L <sup>3</sup>	1,679 802	"
L <sup>4</sup>	3,042 692	"
L <sup>5</sup>	3,876 259	"
L <sup>6</sup>	0,300 360	"

$$\theta^6 = 6,267 419$$

$$\Sigma_0 = 5,967 059$$

A continuación se muestran los valores aproximados y los valores L para todas las magnitudes observadas.  $L_0^i = \theta_0^i - \Sigma_0$ , excepto  $L_0^6 = \theta^6 - \Sigma_0$ .

Constante	Valor	L
L <sub>0</sub> <sup>1</sup>	5,987 401	0,000 255
L <sub>0</sub> <sup>2</sup>	0,607 109	0,000 144
L <sub>0</sub> <sup>3</sup>	1,679 894	-0,000 092
L <sub>0</sub> <sup>4</sup>	3,042 754	-0,000 062
L <sub>0</sub> <sup>5</sup>	3,875 993	0,000 266
L <sub>0</sub> <sup>6</sup>	0,300 360	0,000 000
X <sub>P0</sub>	46 020,50	0,10
Y <sub>P0</sub>	12 515,30	0,00

Calculamos los valores  $X_i - X_{P0}$ ,  $Y_i - Y_{P0}$  y  $(S_P^i)_0$  para cada punto, y en función de esos valores obtenemos las derivadas parciales. Las ecuaciones de residuo que se obtienen son las siguientes:

$$v_{L^1} \approx 0,000255 - (2,77 \cdot 10^{-4} \Delta X_P + 3,95 \cdot 10^{-4} \Delta Y_P - \Delta \Sigma),$$

$$v_{L^2} \approx 0,000144 - (-1,32 \cdot 10^{-4} \Delta X_P + 4,40 \cdot 10^{-4} \Delta Y_P - \Delta \Sigma),$$

$$v_{L^3} \approx -0,000092 - (-6,56 \cdot 10^{-4} \Delta X_P + 0,65 \cdot 10^{-4} \Delta Y_P - \Delta \Sigma),$$

$$v_{L^4} \approx -0,000062 - (-1,55 \cdot 10^{-4} \Delta X_P - 3,51 \cdot 10^{-4} \Delta Y_P - \Delta \Sigma),$$

$$v_{L^5} \approx -0,000266 - (1,64 \cdot 10^{-4} \Delta X_P - 3,68 \cdot 10^{-4} \Delta Y_P - \Delta \Sigma),$$

$$v_{L^6} \approx 0 - \Delta \Sigma,$$

$$v_{X_P} \approx 0,10 - \Delta X_P, \quad v_{Y_P} \approx 0 - \Delta Y_P.$$

## 9 Ángulos y distancias

Si tomamos como desviación típica de referencia la de las medidas angulares,  $\sigma_\alpha = 0,00011$ , hay que multiplicar las dos últimas ecuaciones por  $\sigma_\alpha/0,1 = 0,0011$ . Se van obteniendo los siguientes resultados:

$$N = \begin{pmatrix} 17,85 \cdot 10^{-7} & 0,027 \cdot 10^{-7} & 5,02 \cdot 10^{-4} \\ 0,027 \cdot 10^{-7} & 18,22 \cdot 10^{-7} & -1,81 \cdot 10^{-4} \\ 5,02 \cdot 10^{-4} & -1,81 \cdot 10^{-4} & 6 \end{pmatrix},$$

$$N^{-1} = \begin{pmatrix} 57,4 \cdot 10^{-4} & -0,56 \cdot 10^{-4} & -48 \\ -0,56 \cdot 10^{-4} & 55,0 \cdot 10^{-4} & 17 \\ -48 & 17 & 0,171 \end{pmatrix},$$

$$A^T L = \begin{pmatrix} 2,86 \cdot 10^{-7} \\ 0,82 \cdot 10^{-7} \\ -5,11 \cdot 10^{-4} \end{pmatrix}, \quad X_\Delta \approx N^{-1} A^T L = \begin{pmatrix} 0,19 \\ 0,03 \\ -0,000 \ 100 \end{pmatrix}.$$

Y por lo tanto los valores ajustados son  $X = 46 \ 020,69$ ,  $Y = 12 \ 515,33$  y  $\Sigma = 5,966959 = 379^s \ 8684$ . Si sustituimos los valores  $X_\Delta$  en  $AX_\Delta - L$  normalizado obtenemos los residuos  $V_{(no.)}$ , de distribuciones con la misma varianza y que por tanto podemos comparar. Si los sustituimos en  $L - AX_\Delta$  original, o si calculamos con las formulas exactas los valores ajustados de las observaciones, obtenemos los residuos  $V_{(or.)}$ . Los de las lecturas están en radianes, que podemos pasar a grados.

$$\left. \begin{array}{cccc} v'_{L1} = 0,9 & v'_{L2} = 0,5 & v'_{L3} = -0,7 & v'_{L4} = -1,2 \\ v'_{L5} = 1,5 & v'_{L6} = -1,0 & v'_X = -1,0 & v'_Y = -0,4 \end{array} \right\} \times 10^{-4},$$

$$\begin{array}{cccc} v_{L1} = 0^s \ 0057, & v_{L2} = 0^s \ 0034, & v_{L3} = -0^s \ 0045, & v_{L4} = -0^s \ 0077, \\ v_{L5} = 0^s \ 0094, & v_{L6} = -0^s \ 0063, & v_X = -0,09, & v_Y = -0,03. \end{array}$$

### Suma de los residuos de lectura de una estación

En un ajuste con ángulos y distancias, la suma de los residuos de las lecturas de una estación será siempre cero. Esto es consecuencia de que para un conjunto de medidas de una misma magnitud el estimador mínimo cuadrático proporciona como resultado la media aritmética. La media aritmética  $x$  es el valor que hace que la suma de los residuos sea cero. Si escribimos las ecuaciones de residuo para las

## 9 Ángulos y distancias

observaciones, sus derivadas respecto al único parámetro  $x_1$  son todas  $-1$ . Aunque multipliquemos todas las ecuaciones de residuo por una constante el resultado del ajuste sigue siendo el mismo, de manera que, dado un parámetro  $x$  en función del cual se expresan varias magnitudes observadas, y la derivada de cada una de ellas respecto del parámetro es la misma constante  $k$ , entonces el valor ajustado  $x$  es el que hace  $\sum v = 0$ , en donde el sumatorio está extendido a las observaciones que dependen de  $x$ .

En el caso una desorientación  $\Sigma_1$ , este parámetro está presente exactamente en las ecuaciones de residuo de las lecturas que tienen como origen la estación 1, y la derivada es  $-(-1) = 1$ . Intervienen otros parámetros en esas observaciones, parámetros que también aparecen en otras observaciones distintas, pero veamos que eso no influye. En efecto, sea un conjunto cualquiera  $(\Sigma_1, X_1, Y_1, X_2, \dots)$  de valores ajustados para el cual  $\sum_{L_1} v \neq 0$ , en donde el sumatorio se aplica a las observaciones de lectura con origen en 1. Supongamos que manteniendo fijos todos los demás parámetros variamos  $\Sigma_1$ . Entonces, según variamos  $\Sigma_1$  los residuos afectados disminuyen o aumentan todos una cantidad constante. Así, tenemos un parámetro  $\Sigma_1$  tal que al variarlo un conjunto de residuos se modifica en una cantidad constante. Esta es la situación de la media aritmética (exactamente, de la estimación cuyo resultado es la media aritmética), y la solución que minimiza la suma de los cuadrados de los residuos es el valor del parámetro que hace su suma igual a 0. Por lo tanto, sean cuales sean los demás parámetros, el parámetro  $\Sigma_1$  será siempre tal que  $\sum_{L_1} v = 0$ .

Esto quiere decir que el hecho de que en una estación la suma de los residuos de las lecturas sea 0 no significa nada, pues siempre es así. También, si una estación sólo tiene dos lecturas angulares entonces sus residuos serán iguales en valor absoluto y de signo contrario, lo que tampoco significa nada.

### Resumen de fórmulas

Ecuación del residuo de distancia:

$$v_S \approx (S - S_0) - \left\{ \left( \frac{x_2 - x_1}{S} \right)_0 \Delta x_1 - \left( \frac{y_2 - y_1}{S} \right)_0 \Delta y_1 + \left( \frac{x_2 - x_1}{S} \right)_0 \Delta x_2 + \left( \frac{y_2 - y_1}{S} \right)_0 \Delta y_2 \right\}. \quad (9.5)$$

## 9 Ángulos y distancias

Ecuación del residuo de lectura:

$$v_L \approx (L - L_0) - \left\{ \left( \frac{y_2 - y_1}{S^2} \right)_0 \Delta x_1 + \left( \frac{x_2 - x_1}{S^2} \right)_0 \Delta y_1 + \right. \\ \left. + \left( \frac{y_2 - y_1}{S^2} \right)_0 \Delta x_2 - \left( \frac{x_2 - x_1}{S^2} \right)_0 \Delta y_2 - \Delta \Sigma_1 \right\}, \quad (9.6)$$

siendo  $L_0 = \theta_0 - \Sigma_0$ .

Residuo de azimut (residuo de lectura de una visual de azimut conocido):

$$v_L = (L - L_0) - \Sigma_0, \quad \text{siendo } L_0 = \theta - \Sigma_0 \quad (9.7)$$

Ecuación de residuo de coordenada:

$$v_x = (x - x_0) - \Delta x, \quad v_y = (y - y_0) - \Delta y. \quad (9.8)$$



# 10 Regiones de confianza

En este capítulo aplicaremos lo visto sobre la obtención de los intervalos y regiones de confianza al estimador mínimo cuadrático. Estudiaremos sólo el método paramétrico.

## Transmisión de varianzas

Antes de abordar el problema necesitamos unos resultados previos. En primer lugar la *transmisión de varianzas* en el caso lineal; es decir, si

$$Y = AX \quad (10.1)$$

y conocemos  $\Sigma_{xx}$ , obtener  $\Sigma_{yy}$ .  $X$  es el vector de variables aleatorias  $(x_1, x_2, \dots, x_n)$ , e  $Y$  son otras variables aleatorias  $(y_1, y_2, \dots, y_m)$  que se obtienen en función de las  $X$ .

Si  $x$  es una variable aleatoria denotamos por  $\mu_x$  su esperanza (media). Recordamos que la varianza  $\sigma_x^2$  se define como

$$\sigma_x^2 = E[(x - \mu_x)^2].$$

Si  $x_1, x_2$  son dos variables aleatorias la covarianza  $\sigma_{x_1x_2}$  es

$$\sigma_{x_1x_2} = E[(x_1 - \mu_{x_1})(x_2 - \mu_{x_2})].$$

De acuerdo a esta definición  $\sigma_{x_1x_1} = \sigma_{x_1}^2$ .

## 10 Regiones de confianza

En (10.1) podemos escribir  $x = (x - \mu_x) + \mu_x$ , de donde  $Y = A(X - M_x) + AM_x$ . Las variables del vector  $X' = X - M_x$  tienen media 0 y las mismas varianzas que  $X$ . Por otra parte, las varianzas de un conjunto de variables aleatorias no varían si desplazamos toda la distribución una cantidad constante, y en consecuencia  $\Sigma_{yy} = \Sigma_{y'y'}$ , siendo  $Y' = AX'$ . Por lo tanto podemos suponer que en (10.1)  $E[X] = 0$  y  $E[Y] = 0$ , y entonces

$$\sigma_{x_1x_2} = E[x_1x_2], \quad \sigma_{y_1y_2} = E[y_1y_2].$$

Obtenemos  $\sigma_{y_iy_j}$  a partir de las expresiones de  $y_i$  e  $y_j$ :

$$y_i = a_{i1}x_1 + \dots + a_{in}x_n,$$

$$y_j = a_{j1}x_1 + \dots + a_{jn}x_n,$$

$$\sigma_{y_iy_j} = E[y_iy_j] = E[(a_{i1}x_1 + \dots + a_{in}x_n)(a_{j1}x_1 + \dots + a_{jn}x_n)].$$

Desarrollando,

$$\begin{aligned} \sigma_{y_iy_j} &= E[a_{i1}a_{j1}x_1^2 + (a_{i1}a_{j2} + a_{j1}a_{i2})x_1x_2 + a_{i2}a_{j2}x_2^2 + \dots] \\ &= E[a_{i1}a_{j1}x_1^2] + E[(a_{i1}a_{j2} + a_{j1}a_{i2})x_1x_2] + E[a_{i2}a_{j2}x_2^2] + \dots \\ &= a_{i1}a_{j1}\sigma_{x_1}^2 + (a_{i1}a_{j2} + a_{j1}a_{i2})\sigma_{x_1x_2} + a_{i2}a_{j2}\sigma_{x_2}^2 + \dots \\ &= \sum_{\substack{1 \leq k \leq n \\ 1 \leq l \leq n}} a_{ik}a_{jl}\sigma_{x_kx_l} = (a_{i1} \dots a_{in}) \Sigma_{xx} \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jn} \end{pmatrix}. \end{aligned}$$

Por lo tanto si efectuamos  $A\Sigma_{xx}A^T$  obtendremos una matriz que en la fila  $i$ , columna  $j$  tiene el elemento  $\sigma_{y_iy_j}$ ; es decir, la matriz  $\Sigma_{yy}$ :

$$\Sigma_{yy} = A\Sigma_{xx}A^T. \quad (10.2)$$

### **Función normal multidimensional**

Una matriz de varianzas  $\Sigma_{xx}$  es simétrica, ya que  $\sigma_{x_i x_j} = \sigma_{x_j x_i}$ . Cumple otras muchas propiedades, de las cuales la que nos interesa ahora es que puede diagonalizarse, todos sus autovalores son reales y

## 10 Regiones de confianza

positivos y los autovectores son ortogonales entre sí. Por lo tanto si se toman estos de módulo 1 la matriz de diagonalización es ortogonal:

$$\Lambda = Q\Sigma_{xx}Q^{-1},$$

siendo  $\Lambda$  una matriz diagonal de valores positivos y  $Q$  una matriz ortogonal. Sea

$$Y = QX.$$

Entonces  $\Sigma_{yy} = Q\Sigma_{xx}Q^T = Q\Sigma_{xx}Q^{-1} = \Lambda$ .

Esto último significa que si  $i \neq j$  entonces  $\sigma_{y_i y_j} = 0$ . Es importante ver que esto no significa en absoluto que las variables sean independientes. El concepto de independencia, concretado en la definición de la p. 16, es mucho más fuerte. Dos variables pueden tener covarianza 0 y aun así ser dependientes, incluso tener dependencia funcional. Sin embargo el recíproco sí que es cierto: si dos variables son independientes entonces su covarianza es cero.

Si las variables de un conjunto  $(y_1, \dots, y_n)$  siguen distribuciones normales independientes  $N(0, \sqrt{\lambda_i})$ , entonces la distribución conjunta es el producto de las distribuciones individuales:

$$f_y(Y) = \frac{1}{(2\pi)^{n/2} \sqrt{\lambda_1 \dots \lambda_n}} e^{-\frac{1}{2} \left( \frac{y_1^2}{\lambda_1} + \dots + \frac{y_n^2}{\lambda_n} \right)}, \quad (10.3)$$

en donde  $f_y(Y)$  debe leerse como  $f_y(y_1, \dots, y_n)$ . Podemos escribir la misma expresión de la forma

$$f_y(Y) = \frac{1}{(2\pi)^{n/2} \sqrt{|\Lambda|}} e^{-\frac{1}{2} Y^T \Lambda^{-1} Y}$$

siendo  $\Lambda$  la matriz diagonal  $\Sigma_{yy}$ .

Si  $X = A^{-1}Y$  (y por lo tanto  $Y = AX$ ), siendo  $A$  una matriz cuadrada con  $|A| \neq 0$ , puede deducirse que la distribución de las  $X$  en función de la distribución de las  $Y$  es

$$f_x(X) = |A| f_y(Y(X)) = |A| f_y(AX).$$

Si  $f_y$  es la función de más arriba entonces

$$f_x(X) = \frac{|A|}{(2\pi)^{n/2} \sqrt{|\Lambda|}} e^{-\frac{1}{2} (AX)^T \Lambda^{-1} AX}.$$

## 10 Regiones de confianza

Ahora bien, puesto que  $X = A^{-1}Y$ ,

$$\Sigma_{xx} = A^{-1}\Lambda A^{-1T}, \quad \Sigma_{xx}^{-1} = A^T\Lambda^{-1}A, \quad |\Sigma_{xx}| = |A^{-1}|^2 |\Lambda|;$$

de donde

$$f_x(X) = \frac{1}{(2\pi)^{n/2} \sqrt{|\Sigma_{xx}|}} e^{-\frac{1}{2}X^T \Sigma_{xx}^{-1} X}. \quad (10.4)$$

La distribución de cada variable  $x_i$  es una normal. Puede intentar deducirse a partir de (10.4), pero es una cuestión mucho más simple.  $X = A^{-1}Y$ , es decir, que cada  $x_i$  es una combinación lineal de variables normales independientes, y por lo tanto sigue una distribución normal. Su varianza es el elemento  $\sigma_{x_i}^2$  de la matriz  $\Sigma_{xx}$ , por definición. También puede demostrarse que la distribución de un subconjunto cualquiera de variables de  $X$ ,  $(x_{r_1}, \dots, x_{r_k})$ , es igual a la de (10.4) con  $X^T = (x_{r_1}, \dots, x_{r_k})$ .

La distribución (10.4) recibe el nombre de normal multidimensional. Podemos trasladar cada variable una cantidad  $\mu_i$ , y en la formulación simplemente sustituimos  $X$  por  $X - M_x$ . El nuevo vector de medias es  $M_x = (\mu_1, \dots, \mu_n)$ .

La matriz ortogonal  $Q$  de diagonalización de una matriz  $\Sigma_{xx}$  cumple que si  $Y = QX$  entonces  $\Sigma_{yy}$  es diagonal. Existen muchas otras matrices  $A$  tales que  $A\Sigma_{xx}A^T$  es diagonal. A partir de (10.3) y  $X = A^{-1}Y$  obtuvimos (10.4). Aplicando idéntico razonamiento para obtener  $f_y(Y)$  a partir de (10.4), si  $Y = AX$  siendo  $A$  una de las matrices que cumplen que  $A\Sigma_{xx}A^T$  es diagonal, llegamos a (10.3), lo que quiere decir que las variables  $Y = AX$  siguen distribuciones normales y son independientes.

Si de una función de distribución sabemos que es proporcional a cierta función  $g$ , entonces existe una única constante  $k$  tal que  $kg$  es una función de distribución; es la constante que hace la integral total igual a 1. Por lo tanto una función de distribución queda totalmente definida si establecemos su proporcionalidad a otra función. Aplicado a la distribución normal multidimensional, ésta queda totalmente definida mediante la relación

$$f_x(X) \propto e^{-\frac{1}{2}X^T \Sigma_{xx}^{-1} X}. \quad (10.5)$$

Por lo tanto, si un conjunto de  $n$  variables  $X$  sigue una distribución

$f_x(X)$  proporcional a

$$e^{-\frac{1}{2}X^T N X}$$

entonces:

1. Las variables  $X$  siguen una distribución normal multidimensional.
2. La constante que falta es  $\sqrt{|N|}/(2\pi)^{n/2}$  (y por lo tanto la integral de la exponencial es el inverso).
3. La matriz de covarianzas de  $X$  es  $\Sigma_{xx} = N^{-1}$ .
4. Cada variable  $x_i$  sigue una distribución normal (de media 0).
5. Cualquier subconjunto  $(x_{r_1}, \dots, x_{r_k})$  sigue una distribución normal multidimensional.
6. Si  $A\Sigma_{xx}A^T$  es una matriz diagonal entonces  $Y = AX$  son variables aleatorias independientes que siguen distribuciones normales.

El recíproco de estas propiedades no es cierto. Si las variables de un conjunto  $X$  siguen, cada una, una distribución normal, entonces no podemos afirmar que el conjunto siga una distribución normal multidimensional. Aunque además  $\Sigma_{xx}$  sea diagonal tampoco seguirán en general una distribución normal multidimensional, y ni siquiera podemos afirmar que sean independientes.

De todas las matrices  $A$  para las cuales  $Y = AX$  son variables independientes, la matriz ortogonal  $Q$  tiene un significado especial. Es la matriz de un giro, y por lo tanto podemos interpretar las ecuaciones  $Y = QX$  como un giro, lo que quiere decir que cualquier distribución normal multidimensional tiene la misma forma que el producto de distribuciones normales independientes de  $n$  variables, y para obtener dichas variables sólo hay que aplicar un giro.

### **Función $g$ de los parámetros. Matriz $\Sigma_{xx}$**

Armados con estos resultados ya podemos abordar el problema de obtener las regiones de confianza para los valores estimados. Tenemos que obtener la función  $g$  de los parámetros. No suponemos ninguna distribución  $h$  a priori. Aunque la hubiese y fuese conocida,

## 10 Regiones de confianza

al tratar con magnitudes de las «grandes», en el entorno de la solución el valor de  $h$  sería prácticamente constante. Las magnitudes y matrices que aparezcan serán todas normalizadas, nunca originales. Bajo este supuesto

$$\begin{aligned} g(x_1, \dots, x_n) &= f_{x_1, \dots, x_n}(\ell_1, \dots, \ell_m) \propto \exp\left(-\frac{1}{2} \sum_{x_1, \dots, x_n} \frac{\varepsilon^2}{\sigma_0^2}\right) = \\ &= \exp\left(-\frac{1}{2} \frac{1}{\sigma_0^2} (\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta)^T (\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta)\right). \end{aligned}$$

El sumatorio está extendido a los  $\varepsilon$  correspondientes a los valores reales  $(x_1, \dots, x_n)$ , y los elementos del vector  $\mathbf{X}_\Delta$  son  $\Delta x_i = x_i - x_{0i}$ . Los cálculos se simplifican si tomamos como valores aproximados los valores ajustados. En ese caso,  $\Delta x_i = x_i - x_i$ . La función la podemos interpretar indistintamente como de las variables  $x_i$  o de las variables  $\Delta x_i$ . Emplearemos estas últimas.

$$(\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta)^T (\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta) = \mathbf{L}^T \mathbf{L} - \mathbf{L}^T \mathbf{A}\mathbf{X}_\Delta - \mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L} + \mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{A}\mathbf{X}_\Delta$$

Cada uno de los sumandos es un número, y por tanto igual a su traspuesto, así que  $-\mathbf{L}^T \mathbf{A}\mathbf{X}_\Delta - \mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L} = -2\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L}$ . Por otra parte  $\mathbf{L}^T \mathbf{L}$  es una constante (recordamos que las variables son  $\Delta x_i$ ), y entonces

$$\begin{aligned} \exp\left(-\frac{1}{2} \frac{1}{\sigma_0^2} (\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta)^T (\mathbf{L} - \mathbf{A}\mathbf{X}_\Delta)\right) &= \\ &= e^{-\frac{1}{2\sigma_0^2} \mathbf{L}^T \mathbf{L}} \exp\left(-\frac{1}{2} \frac{1}{\sigma_0^2} \left(\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{A}\mathbf{X}_\Delta - 2\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L}\right)\right) \\ &\propto \exp\left(-\frac{1}{2} \frac{1}{\sigma_0^2} \left(\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{A}\mathbf{X}_\Delta - 2\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L}\right)\right). \end{aligned}$$

Sabemos que si los valores aproximados coinciden con los ajustados se cumple  $\mathbf{A}^T \mathbf{L} = 0$  (cf. (10.5)). Entonces en la función  $g$  que estamos obteniendo  $2\mathbf{X}_\Delta^T \mathbf{A}^T \mathbf{L} = 0$ , y dado que  $\mathbf{A}^T \mathbf{A} = \mathbf{N}$  llegamos a

$$g(\Delta x_1, \Delta x_2, \dots, \Delta x_n) \propto e^{-\frac{1}{2} \mathbf{X}_\Delta^T \sigma_0^{-2} \mathbf{N} \mathbf{X}_\Delta}. \quad (10.6)$$

Por lo tanto la función  $g$  es una normal multidimensional, y podemos aplicar todo lo deducido para ella. En primer lugar, las cantidades  $\Delta x_i = x_i - x_i$  del vector  $\mathbf{X}_\Delta$  siguen distribuciones normales

## 10 Regiones de confianza

de media 0, y por lo tanto  $E[x_i - x_i] = 0$ , y el estimador mínimo cuadrático es centrado desde el punto de vista inverso respecto a los parámetros.

En segundo lugar,

$$\Sigma_{xx} = (\sigma_0^{-2} N)^{-1} = \sigma_0^2 N^{-1}. \quad (10.7)$$

Y por último, si  $Q$  es la matriz ortogonal que diagonaliza  $\sigma_0^2 N^{-1}$ , para obtener la probabilidad de una región en el espacio de los valores  $X_\Delta$  podemos, si lo consideramos oportuno, realizar los cálculos en la distribución de  $Y = QX_\Delta$ , que es mucho más sencilla. Estas regiones serán normalmente las regiones de confianza.

Puesto que cada variable sigue una distribución normal, los intervalos de confianza de variables individuales se obtienen en función del área bajo la gráfica de la normal. A continuación se muestra el área en tanto por ciento en el intervalo  $[-k\sigma, k\sigma]$  para distintos valores de  $k$ :

$k$	1	1,64	1,8	2	2,5	3
$p$ (%)	68	90	93	95,5	98,8	99,7

### Elipses de confianza

Si dos variables,  $y_1, y_2$ , son independientes y siguen cada una una distribución normal  $(0, 1)$ , entonces la distribución tiene simetría radial, ya que

$$f(y_1, y_2) \propto e^{-\frac{1}{2}(y_1^2 + y_2^2)} = e^{-\frac{d^2}{2}}.$$

Por lo tanto la función sólo depende de la distancia al origen  $d$ , y en consecuencia tiene simetría radial. Esta propiedad la podemos extender a cualquier número de variables. La variable  $d^2 = y_1^2 + y_2^2$  sigue, por definición, una distribución  $\chi_2^2$ . Entonces el volumen (que representa una probabilidad) que tiene como base un círculo de radio  $k$  es  $\chi_2^2(k^2) = 1 - e^{-k^2/2}$ :

$k$	1	2	2,15	2,5	3
$p$ (%)	39	86,5	90	95,5	98,9

Si multiplicamos  $y_1$  por  $\sigma_1$  la variable pasa a tener una desviación típica igual a  $\sigma_1$ , y el círculo se convierte en una elipse de semiejes

## 10 Regiones de confianza

$k, \sigma_1 k$ . A un valor dentro del círculo corresponde uno dentro de la elipse, y viceversa, de manera que la probabilidad en el interior de uno y otro es la misma. En general, si  $y_1$  e  $y_2$  son independientes y siguen distribuciones normales  $(0, \sigma_1)$  y  $(0, \sigma_2)$ , entonces la elipse de semiejes  $(k\sigma_1, k\sigma_2)$  orientados según los ejes  $y_1$  e  $y_2$  tiene densidad de probabilidad constante, y la probabilidad acumulada por la distribución en su interior es  $X_2^2(k^2)$ . Más en general, para un número cualquiera de variables independientes  $y_1, y_2, \dots, y_n$  que siguen distribuciones normales (de media 0) de desviaciones típicas  $\sigma_1, \sigma_2, \dots, \sigma_n$ , la figura determinada por

$$\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} + \dots + \frac{y_n^2}{\sigma_n^2} = k^2$$

tiene una densidad de probabilidad constante, y la probabilidad de la distribución en su interior es  $X_n^2(k^2)$ .

Cualquier distribución normal multidimensional es una distribución de variables normales independientes girada, y sus regiones de confianza son iguales a las de ésta última pero giradas. Así, en dos dimensiones para obtener la elipse tenemos que encontrar la matriz de giro  $Q$  tal que  $Q\Sigma_{xx}Q^T = \Lambda$  es diagonal. Si  $A$  es una matriz cuadrada  $\text{Tr}(A)$  denota la suma de los elementos de la diagonal principal, y se lee «traza de  $A$ ». Puede demostrarse fácilmente que  $|\Sigma_{xx}| = |\Lambda|$  y  $\text{Tr}(\Sigma_{xx}) = \text{Tr}(\Lambda)$ . Sea  $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ , tenemos entonces

$$\begin{aligned} \lambda_1 \lambda_2 &= \sigma_1^2 \sigma_2^2 - \sigma_{12}^2, \\ \lambda_1 + \lambda_2 &= \sigma_1^2 + \sigma_2^2. \end{aligned}$$

Resolviendo este sistema obtenemos

$$\left. \begin{matrix} \lambda_1 \\ \lambda_2 \end{matrix} \right\} = \frac{\sigma_1^2 + \sigma_2^2 \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\sigma_{12}^2}}{2}. \quad (10.8)$$

En donde los dos valores se obtienen tomando un signo u otro en la raíz. Entonces la elipse de semiejes  $(k\sqrt{\lambda_1}, k\sqrt{\lambda_2})$ , cuyos ejes siguen las direcciones de los ejes coordenados girados  $Q^T$ , encierra una probabilidad  $X_2^2(k^2)$ . El giro es  $Q^T$  y no  $Q$ , ya que si  $Y = QX$  la elipse respecto a las variables  $Y$  tiene sus ejes siguiendo los ejes coordenados, y  $X = Q^TY$ , y entonces respecto a las variables  $X$  está girada  $Q^T$ .

## 10 Regiones de confianza

Sea  $Z$  un punto del eje de la elipse asociado a  $\lambda_1$  de coordenadas  $(x_1, x_2)$ . En el sistema  $y_1, y_2$  es un punto de uno de los ejes de coordenadas, supongamos que de  $y_1$ , y por lo tanto sus coordenadas son  $(y_1, 0)$ . Emplearemos la notación  $Z_X$  para denotar el par  $(x_1, x_2)$  y  $Z_Y$  para  $(y_1, 0)$ . De esta manera  $Z$  designa el punto, sin referirlo a ningún sistema de coordenadas particular.

Se cumple

$$\Lambda Z_Y = \begin{pmatrix} \lambda_1 y_1 \\ 0 \end{pmatrix} = \lambda_1 Z_Y.$$

Pero  $Z_Y = QZ_X$ , de donde

$$\begin{aligned} \Lambda QZ_X &= \lambda_1 QZ_X, \\ Q^T \Lambda QZ_X &= \lambda_1 Z_X, \\ \Sigma_{xx} Z_X &= \lambda_1 Z_X. \end{aligned} \tag{10.9}$$

La última ecuación permite obtener el conjunto de valores  $Z_X$ , es decir, la dirección del eje. Sea  $Z_X = (\cos \theta, \sin \theta)$  (y entonces  $\theta$  es uno de los dos ángulos que el eje forma con el eje  $x_1$ ). Según la ecuación (10.9),

$$\sigma_1^2 \cos \theta + \sigma_{12} \sin \theta = \lambda_1 \cos \theta.$$

Dividiendo toda la ecuación por  $\cos \theta$  obtenemos

$$\tan \theta = \frac{\lambda_1 - \sigma_1^2}{\sigma_{12}}. \tag{10.10}$$

El otro eje forma un ángulo recto con este.

Una elipse de error proporciona mucha información. De hecho proporciona toda la información de dos variables que siguen una distribución normal multidimensional. Sin embargo, si la precisión para el par de variables se va a dar como un único valor, dentro del cual se encuentra el par (punto) con una probabilidad  $p$ , entonces el número que se dé tiene que ser precisamente eso: el radio del círculo que contiene una probabilidad  $p$ . Sucede que la elipse es mucho más fácil de calcular que el círculo, que requiere un cálculo numérico aproximado de la integral, y lo habitual es sustituir ese valor por el semieje mayor de la elipse de probabilidad  $p$ . El círculo que tiene por radio ese semieje contiene más probabilidad que  $p$ , ya que contiene a la elipse,

## 10 Regiones de confianza

por lo que el radio de probabilidad  $p$  es menor. Si la elipse es redondeada la diferencia será pequeña, y si es muy alargada será mayor. Cuando la elipse se aproxima a un segmento, el radio  $r$  de un círculo de probabilidad  $p$  tiende al valor  $h$  para el cual en el intervalo  $[-h, h]$  contiene una probabilidad  $p$  en la distribución normal, pero siempre será mayor.

### Elipsoides de confianza

Si buscamos regiones de confianza para un conjunto de tres variables, estas regiones serán elipsoides. Los pasos para hallar el elipsoide, semiejes y orientación, son los mismos que en dos dimensiones: obtener los valores  $\lambda_1, \lambda_2, \lambda_3$  y a partir de éstos la orientación de cada eje. El proceso es un poco largo, y requiere la solución de una ecuación de tercer grado. Precisamente por eso es muy difícil si no imposible encontrar escritos en algún libro o artículo de manera explícita todos los pasos, y por eso los indicamos a continuación, aunque sin ninguna explicación.

$$\begin{aligned}a &= (\sigma_1^2 + \sigma_2^2 + \sigma_3^2), \\b &= -(\sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_3^2 + \sigma_3^2\sigma_1^2) + (\sigma_{12}^2 + \sigma_{23}^2 + \sigma_{31}^2), \\c &= \sigma_1^2\sigma_2^2\sigma_3^2 - (\sigma_1^2\sigma_{23}^2 + \sigma_2^2\sigma_{31}^2 + \sigma_3^2\sigma_{12}^2) + 2\sigma_{12}\sigma_{23}\sigma_{31}, \\m &= b + a^2/3, \\n &= c + ab/3 + 2a^3/27, \quad n' = n/2, \\r &= (m/3)^3 - n'^2, \\\rho &= (n'^2 + r)^{1/6}, \\\alpha &= \frac{1}{3} \arctan \frac{\sqrt{r}}{n'}, \\\mu_1 &= 2\rho \cos \alpha, \\d &= \sqrt{\mu_1^2 - 4n/\mu_1}, \\\mu_2 &= \frac{1}{2}(-\mu_1 + d), \\\mu_3 &= \frac{1}{2}(-\mu_1 - d), \\\lambda_1 &= \mu_1 + a/3, \quad \lambda_2 = \mu_2 + a/3, \quad \lambda_3 = \mu_3 + a/3.\end{aligned}$$

## 10 Regiones de confianza

Este algoritmo contempla todos los casos posibles salvo  $m = n = 0$ , que equivale a  $\sigma_{12} = \sigma_{23} = \sigma_{31} = 0$  y  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ , en cuyo caso  $\lambda_1 = \lambda_2 = \lambda_3 = \sigma_1^2$ . Cuando se calcula  $\alpha$ , de las dos soluciones posibles ha de tomarse aquélla en la que  $\cos \alpha$  tiene el mismo signo que  $n'$ . Este método tiene además la propiedad de que los valores obtenidos cumplen  $\lambda_1 \geq \lambda_2 \geq \lambda_3$ .

Siguiendo un razonamiento idéntico al que permite obtener las direcciones de los ejes en el caso bidimensional llegamos al sistema de ecuaciones (por ejemplo para  $\lambda_1$ )

$$\Sigma_{xx}Z_X = \lambda_1 Z_X,$$

que cumplen todos los puntos  $Z_X$  del eje. Hay un punto del eje que viene dado por las siguientes expresiones:

$$\begin{aligned} x_1 &= \sigma_{12}\sigma_{23} - (\sigma_2^2 - \lambda_1)\sigma_{31}, \\ x_2 &= \sigma_{31}\sigma_{12} - (\sigma_1^2 - \lambda_1)\sigma_{23}, \\ x_3 &= (\sigma_1^2 - \lambda_1)(\sigma_2^2 - \lambda_1) - \sigma_{12}^2. \end{aligned}$$

Análogamente obtenemos los otros dos semiejes. El tercero, en lugar de obtenerlo así podemos calcularlo como el producto vectorial de los otros dos. Si dos semiejes son iguales sólo tiene sentido calcular la dirección del distinto.

### **Función $g$ de las magnitudes observadas. Matriz $\Sigma_{ll}$**

Los valores reales de las magnitudes observadas en función de los valores reales de los parámetros son aproximadamente  $L_r = AX_\Delta + L_0$ , en donde  $L_r = (l_1, \dots, l_n)$  y  $X_\Delta = (\Delta x_1, \dots, \Delta x_n)$ . Entonces la función  $g$  de las magnitudes observadas también es una normal multidimensional y su matriz de varianzas es

$$\Sigma_{ll} = \sigma_0^2 AN^{-1}A^\top. \quad (10.11)$$

Si tomamos la matriz  $A_{(or.)}$  obtendremos  $\Sigma_{ll}$  para las observaciones originales. Si tomamos  $A_{(no.)}$  obtendremos  $\Sigma_{ll}$  para las observaciones normalizadas.

Podemos obtener las regiones de confianza para un conjunto de variables  $l$  como se explicó en las secciones anteriores, siempre las variables  $l$  seleccionadas sean funcionalmente independientes.

## 10 Regiones de confianza

### **Función $f$ de los valores ajustados de los parámetros. Matriz $\Sigma_{xx}$**

Más sencillo que las funciones  $g$  resulta el obtener la distribución de los valores ajustados. Una vez más la razón es que aquello es un problema inverso, mientras que éste es un problema directo.

El resultado del ajuste no varía si en las ecuaciones de residuo linealizadas modificamos los valores  $x_0$ ; lógicamente el resultado no puede depender del valor aproximado. Entonces, para simplificar el desarrollo, podemos suponer que  $x_{i0} = x_i$ , incluso aun cuando al calcular el ajuste no conocemos los valores reales. Los valores  $l_0$  correspondientes a los valores reales de los parámetros son los valores reales  $l$ , y entonces  $L_i = \ell_i - l_{i0} = \ell_i - l_i = \varepsilon_i$ . Por lo tanto,

$$V = E - AX_{\Delta}, \quad (10.12)$$

siendo  $X_{\Delta} = (\Delta x_i) = (x_i - x_i)$ .

Los valores ajustados son

$$X_{\Delta} = N^{-1}A^T E.$$

Las variables  $\varepsilon$  son independientes y siguen distribuciones normales  $(0, \sigma_0)$ , de donde  $\Sigma_{\varepsilon\varepsilon} = \sigma_0^2 I$ , y entonces los valores ajustados siguen una distribución normal multidimensional, siendo

$$\Sigma_{xx} = \sigma_0^2 N^{-1}A^T(N^{-1}A^T)^T = \sigma_0^2 N^{-1}A^T A N^{-1} = \sigma_0^2 N^{-1}. \quad (10.13)$$

### **Función $f$ de los valores ajustados de los parámetros. (bis)**

Por si el lector no se ha quedado convencido con la posibilidad de tomar como valores aproximados los valores reales efectuaremos la demostración paso a paso, sin atajos. Empezaremos recordando que la distribución de los valores ajustados significa lo siguiente: Los valores ajustados de cualesquiera magnitudes (parámetros, magnitudes observadas... ) se obtienen como el resultado de una fórmula en función de los valores observados. Supuestos unos valores reales  $(x_1, \dots, x_n) \equiv (l_1, \dots, l_n)$ , los valores observados seguirán una cierta distribución  $f$ . Los valores ajustados, al obtenerse en función de los observados, seguirán una distribución que se obtiene en función de  $f$ , es decir, que si repitiésemos las observaciones y el ajuste muchas

## 10 Regiones de confianza

veces para el mismo conjunto de valores reales los valores ajustados tenderían a una cierta distribución, que es la que hallamos.

Los valores ajustados de los parámetros son

$$X_0 + X_\Delta = X_0 + N^{-1}A^T L.$$

Si repetimos el ajuste muchas veces la matriz  $A$  será siempre la misma, y en consecuencia también  $N^{-1}$ . Los valores  $L = \ell - l_0$  debemos expresarlos en función de los valores observados, que ya está hecho, y de los valores aproximados. Sea  $(x_i)$  un conjunto cualquiera de valores de los parámetros y  $(x_i - x_i)$  su diferencia respecto al valor real, que vectorialmente denotaremos por  $\Delta_x$ . Sea  $L_r$  el vector de valores reales de las magnitudes observadas y  $L_x$  el vector que corresponde a los valores  $x$ . Entonces se cumple

$$L_x = A\Delta_x + L_r.$$

Si los valores  $x$  son los valores aproximados entonces

$$\begin{aligned} L_0 &= A\Delta_{x_0} + L_r, \\ L &= \mathcal{L} - L_0 = \mathcal{L} - A\Delta_{x_0} - L_r = -A\Delta_{x_0} + E. \end{aligned}$$

Por lo tanto, si hemos tomados unos valores aproximados  $X_0$  ésta será la expresión de  $L$ .

Los valores ajustados son

$$\begin{aligned} X &= X_0 + X_\Delta = X_0 + N^{-1}A^T L = X_0 + N^{-1}A^T (-A\Delta_{x_0} + E) \\ &= X_0 - N^{-1}NA\Delta_{x_0} + N^{-1}A^T E = X_r + N^{-1}A^T E. \end{aligned} \quad (10.14)$$

Siendo  $X_r$  el vector de valores reales. El resultado no depende de  $X_0$ , como no podía ser de otra manera.  $\Sigma_{\varepsilon\varepsilon} = \sigma_0^2 I$ , y por lo tanto los valores ajustados  $X$  siguen una distribución normal multidimensional centrada en  $X_r$  y con  $\Sigma_{xx} = \sigma_0^2 N^{-1}A^T(N^{-1}A^T)^T = \sigma_0^2 N^{-1}$ .

### **Función $f$ de los valores ajustados de las observaciones. Matriz $\Sigma_{ll}$**

Los valores ajustados de las observaciones son  $L_{aj} = AX + L_0$ , donde  $L_0$  es una constante, y entonces

$$\Sigma_{ll} = AN^{-1}A^T, \quad (10.15)$$

## 10 Regiones de confianza

En donde  $\Sigma_{ll}$  corresponderá a las observaciones originales o normalizadas según hayamos tomado la matriz  $A_{(or.)}$  o  $A_{(no.)}$ .

Al valor exacto  $X_r$  corresponde el valor exacto  $L_r$ , y dado que la distribución de los parámetros ajustados está centrada en los valores verdaderos, y la relación entre  $X$  y  $L_{aj}$  es (podemos suponer que es) lineal, entonces la distribución de los valores ajustados de las observaciones también está centrada en los valores reales.

### Reflexión sobre las distribuciones directas e inversas

Si las observaciones siguen distribuciones normales el estimador mínimo cuadrático es el de máxima probabilidad. Hemos demostrado que si además podemos admitir relaciones lineales entre observaciones y parámetros entonces también es centrado, tanto desde el punto de vista directo como inverso. Además se produce la coincidencia

$$\Sigma_{xx} = \Sigma_{xx}, \quad \Sigma_{ll} = \Sigma_{ll}. \quad (10.16)$$

Esta coincidencia es desafortunada, ya que está muy extendida la creencia de que las regiones de confianza para los parámetros vienen dadas por  $\Sigma_{xx}$ . La función  $g(x_1, \dots, x_n)$  y la función  $f(x_1, \dots, x_n)$  son dos conceptos totalmente distintos que en este estimador en particular casi coinciden.\*

### Función $g$ de los errores. Matriz $\Sigma_{\varepsilon\varepsilon}$

La coincidencia se rompe para los residuos:  $\Sigma_{\varepsilon\varepsilon} \neq \Sigma_{vv}$ . La primera de las dos no la hemos escrito explícitamente, ya que evidentemente es igual que  $\Sigma_{ll}$ . Hacer referencia a un valor real o a un error es lo mismo, ya que dado  $l$  entonces  $\varepsilon = l - l$ , y dado  $\varepsilon$  entonces  $l = l - \varepsilon$ . Por lo tanto las funciones  $g(l_1, \dots, l_m)$  y  $g(\varepsilon_1, \dots, \varepsilon_m)$  son iguales salvo por un desplazamiento y un cambio de signo. La primera está centrada en  $(l_1, \dots, l_m)$ , y entonces la segunda está centrada en  $(v_1, \dots, v_m)$ . Decir que hay una probabilidad  $p$  de que un valor real  $l$

---

\* Están centradas en puntos distintos. La primera en los valores ajustados y la segunda en los valores reales, que desconocemos.

## 10 Regiones de confianza

esté en el intervalo  $[l - h, l + h]$  equivale a decir que hay una probabilidad  $p$  de que el error  $\varepsilon$  esté en  $[v - h, v + h]$  (el  $+h$  de este intervalo se corresponde con el  $-h$  del intervalo anterior y viceversa).

$$\Sigma_{\varepsilon\varepsilon} = \Sigma_{ll} = \sigma_0^2 \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T.$$

### Función $f$ de los residuos. Matriz $\Sigma_{vv}$

Para obtener  $\Sigma_{vv}$  tenemos que expresar los residuos en función de los valores reales y los errores  $E$ , al igual que hicimos para hallar  $\Sigma_{xx}$ . Podemos escribir los valores ajustados de las observaciones en función de los valores ajustados de los parámetros de la forma

$$L_{aj} = \mathbf{A}(\mathbf{X} - \mathbf{X}_r) + L_r.$$

Empleamos para  $\mathbf{X}$  la expresión de (10.14), y obtenemos

$$\mathbf{V} = \mathcal{L} - L_{aj} = \mathbf{E} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T \mathbf{E} = (\mathbf{I} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T) \mathbf{E}.$$

De aquí,

$$\begin{aligned} \Sigma_{vv} &= \sigma_0^2 (\mathbf{I} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T) (\mathbf{I} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T)^T = \dots \\ &= \sigma_0^2 (\mathbf{I} - \mathbf{A} \mathbf{N}^{-1} \mathbf{A}^T) = \sigma_0^2 (\mathbf{I} - \Sigma_{ll}). \end{aligned} \quad (10.17)$$

Y teniendo en cuenta que  $\sigma_0^2 \mathbf{I} = \Sigma_{\ell\ell}$ , podemos escribir esto mismo de las formas equivalentes

$$\begin{aligned} \Sigma_{ll} + \Sigma_{vv} &= \Sigma_{\ell\ell} && \text{(no.)}, \\ \Sigma_{ll} + \Sigma_{vv} &= \Sigma_{\ell\ell} && \text{(or.)}. \end{aligned} \quad (10.18)$$

### Análisis de $\Sigma_{ll} + \Sigma_{vv} = \Sigma_{\ell\ell}$

La interpretación de (10.18) es clara: Supongamos un valor real  $l$  y un valor observado  $\ell$ . Supongamos además que en ese problema en particular, debido a la disposición de las demás observaciones, el ajuste tiende a estimar muy bien el valor de  $l$  (por ejemplo,  $l$  puede ser un ángulo de un triángulo y hemos medido los tres ángulos 20 veces). Entonces la variabilidad de  $l$  será muy pequeña, mientras que

## 10 Regiones de confianza

$v = \ell - l$  tendrá una variabilidad casi igual que la de la propia  $\ell$ . Por el contrario, si el ajuste prácticamente se ajusta, valga la redundancia, a una observación  $\ell$ , entonces esto quiere decir que apenas se puede estimar el error, y por lo tanto  $\sigma_v$  será muy pequeña, mientras que el valor  $l$  será casi igual que  $\ell$  y su desviación típica casi igual que la de la observación.

Sea por ejemplo un octógono en el que se han medido los ocho ángulos y nada más. Supongamos que los errores en las medidas angulares siguen distribuciones iguales  $N(0, \sigma)$ . Es fácil ver que el estimador mínimo cuadrático reparte los errores por igual. Si el error en la suma total, que es conocido, es  $\varepsilon_s$ , entonces  $v_1 = \dots = v_8 = \varepsilon_s/8$ . La variable  $\varepsilon_s$  (cf. (4.3)) sigue una distribución  $N(0, \sqrt{8}\sigma)$ , y

$$v_i \sim N(0, \sqrt{8}/8\sigma) = N(0, 1/\sqrt{8}\sigma), \quad i = 1, \dots, 8.$$

Es decir,  $\sigma_v$  es sólo  $1/\sqrt{8}$  de  $\sigma_\ell$ . Por el contrario

$$l_i = \ell_i - v_i = \ell_i + \frac{7}{8}\varepsilon_i - \frac{1}{8}\sum_{j \neq i} \varepsilon_j,$$

de donde

$$\sigma_l^2 = \left\{ \left(\frac{7}{8}\right)^2 + \left(\frac{1}{8}\right)^2 + \dots + \left(\frac{1}{8}\right)^2 \right\} \sigma^2 = \frac{7 \cdot 7 + 7}{8 \cdot 8} \sigma^2 = \frac{7}{8} \sigma^2;$$

y en efecto se cumple  $\Sigma_{ll} + \Sigma_{vv} = \Sigma_{\ell\ell}$ , que en este caso equivale a

$$\sigma_l^2 + \sigma_v^2 = \sigma_\ell^2.$$

### Esperanza del valor ajustado $l$ en función del error cometido

Parece lógico suponer que el estimador siempre tiende a amortiguar el error, es decir, que si una medida tiene un error  $\varepsilon$  entonces el ajuste tiende a dar un valor ajustado que está entre  $l$  y  $\ell$ , de manera que  $0 < v < \varepsilon$  si  $\varepsilon$  es positivo, y análogamente  $0 < -v < -\varepsilon$  si  $\varepsilon$  es negativo. Antes de estudiar la cuestión conviene distinguir claramente entre «tiende a dar» y «da». Es imposible garantizar que el valor ajustado está entre el valor observado y el valor real (pensar por ejemplo en dos medidas de una misma magnitud, o en un triángulo

en el que se miden los ángulos). Una primera idea de amortiguación puede ser que  $\sigma_l < \sigma_\ell$ , ya que ambos están centrados en el valor real, y  $\sigma_v < \sigma_\ell$ , ya que lo contrario podría indicar una sobrecorrección (que el valor  $l$  sea intermedio entre  $l$  y  $\ell$ ). Esto es así puesto que, tomando los valores de la diagonal en la igualdad  $\Sigma_{ll} + \Sigma_{vv} = \Sigma_{\ell\ell}$ , tenemos que para cada observación se cumple

$$\sigma_l^2 + \sigma_v^2 = \sigma_\ell^2,$$

y las tres cantidades son positivas, de manera que  $\sigma_l < \sigma_\ell$  y  $\sigma_v < \sigma_\ell$ .

Sin embargo esto no es suficiente para concluir que, sea cual sea  $\ell$ , por término medio el valor de  $l$  está entre  $l$  y  $\ell$ . Necesitamos formular la idea de amortiguación de manera precisa. Sabemos que la distribución de un valor ajustado  $l$  está centrada en  $l$ ; es decir, que si repitiésemos muchas y muchas veces un conjunto de observaciones y el ajuste, la media de los valores ajustados  $l$  tendería a  $l$ , o lo que es lo mismo  $E[l] = l$ . Tenemos que estudiar lo que ocurre, no para la media de todos los ajustes en general, sino cuando el error cometido en la observación es un valor concreto  $\varepsilon$ . Por lo tanto si ahora  $\varepsilon$  en una observación, por ejemplo  $\varepsilon_1$ , es fijo, tenemos que encontrar cuánto vale  $E[l_1]$ ; es decir, qué podemos esperar por término medio del valor ajustado cuando el error cometido ha sido  $\varepsilon_1$ . Lo denotaremos por  $E[l_1]_{\varepsilon_1}$ . Lo deseable es que  $E[l_1]_{\varepsilon_1}$  esté entre  $l_1$  y  $\ell_1$  sea cual sea  $\varepsilon_1$ .

En la p. 105 tenemos

$$V = (I - AN^{-1}A^T)E = \frac{1}{\sigma_0^2} \Sigma_{vv} E,$$

de donde

$$v_1 = \frac{\sigma^2_{v_1}}{\sigma_0^2} \varepsilon_1 + \dots,$$

en donde tenemos que considerar fijo  $\varepsilon_1$ . Los sumandos que faltan por escribir son una combinación lineal de las variables independientes  $\varepsilon_2, \varepsilon_3, \dots, \varepsilon_m$ , que siguen distribuciones normales y tienen media 0. Entonces  $v_1$  sigue una distribución normal centrada en  $\sigma^2_{v_1} / \sigma_0^2 \varepsilon_1$ . La constante  $(\sigma^2_{v_1} / \sigma_0^2)_{(no.)} = \sigma^2_{v_1} / \sigma^2_{\ell_1}$  es positiva y menor que 1, lo que quiere decir que  $E[v_1]_{\varepsilon_1}$  está comprendida entre 0 y  $\varepsilon_1$ , como esperábamos.

## 10 Regiones de confianza

Una vez realizado el ajuste, la información acerca de  $\varepsilon$  viene dada por la función  $g(l)$  o su equivalente  $g(\varepsilon)$ , la primera centrada en  $l$  y la segunda en  $v$ , y con orientaciones opuestas. Supongamos  $l < \ell$ ; el estudio es idéntico en caso contrario. Si  $l < (l + \ell)/2$  entonces el valor  $l$  obtenido es mejor que  $\ell$ . La distribución  $g(l)$  está centrada en  $l$ , y en consecuencia la probabilidad de que eso sea así es mayor que  $\frac{1}{2}$ . Por lo tanto existe una probabilidad mayor que  $\frac{1}{2}$  de que el valor ajustado tenga menos error que el observado, y mayor será cuanto más separados estén ambos valores. En los casos en los que  $\sigma_v$  es muy pequeño,  $l$  tiende a ser muy próximo a  $\ell$ , y por lo tanto la probabilidad (por encima de  $\frac{1}{2}$ ) de que el valor ajustado sea mejor que el observado normalmente será muy pequeña. Tengamos en cuenta que en estos casos además  $\sigma_\varepsilon$  será mayor (recordamos que  $\sigma_\varepsilon = \sigma_l = \sigma_l$ ), lo que acentúa la poca probabilidad en el intervalo  $[l, (l + \ell)/2]$ .

### Diseño de las observaciones

Las distribuciones de valores estimados y cuantos estudios acerca de dichas magnitudes se puedan realizar a priori, como el que acabamos de desarrollar relativo a  $E[v_1]_{\varepsilon_1}$ , no aportan ninguna información sobre los valores reales en un ajuste particular. Esta información está contenida, toda, en las funciones  $g$ . Sin embargo son útiles a la hora de diseñar el conjunto de observaciones que se efectuarán para resolver el problema. Entre el conjunto de magnitudes observadas y magnitudes parámetro habrá algunas que revistan particular interés, y que queramos obtener con la mejor precisión posible. Las matrices de varianzas  $\Sigma_{xx}$  y  $\Sigma_{ll}$  nos indican, ya antes de observar, lo que podemos esperar de los valores que se obtendrán en el ajuste. Así, variando la disposición y precisión relativa de las observaciones podemos estudiar el comportamiento de estas matrices, y buscar la configuración que mejor convenga para las magnitudes que nos son de interés. Puede haber incluso magnitudes que no se encuentren ni entre las observaciones ni entre los parámetros, pero podremos escribirlas, en forma lineal aproximada, como  $Y = BX_\Delta + Y_0$ , y obtenemos  $\Sigma_{yy} = B\Sigma_{xx}B^T$  sin mayor complicación.

Cuando diseñando las observaciones obtenemos  $\Sigma_{xx}$ , o más generalmente sin restringirnos al estimador mínimo cuadrático y distribuciones normales,  $f(x_1, \dots, x_n)$ , estamos obteniendo simplemente lo

que podemos esperar de los valores ajustados, no la precisión con la que obtendremos en el ajuste las estimaciones de los valores reales. Eso no lo podemos saber hasta que hallamos realizado las observaciones y obtenido en función de ellas la función  $g$ . El problema es que en el caso particular del estimador mínimo cuadrático sí lo podemos saber de antemano, ya que coincide  $\Sigma_{xx}$  con  $\Sigma_{xx}$ , y se tiende a confundir ambos conceptos. La diferencia se ve claramente con otra distribución distinta, por ejemplo la distribución uniforme. Si realizamos dos medidas de una magnitud, y esas medidas siguen una distribución uniforme centrada en el valor real, entonces el estimador centrado es la media:  $l = (\ell_1 + \ell_2)/2$ . Esta media sigue una distribución  $f(l)$  triangular. Sin embargo cuando observemos puede que tengamos la suerte de obtener cada medida cerca de un extremo de la distribución, y entonces el dominio de la función  $g$  es un intervalo muy estrecho; o puede que tengamos la mala suerte de que las dos medidas estén muy próximas, y entonces la indeterminación en el valor de  $l$  es mayor. Además, sea cual sea la diferencia entre las dos observaciones la función  $g$  será constante entre los valores máximo y mínimo posibles, y no triangular. En el caso de que las observaciones sigan distribuciones normales, la función  $g$  también depende de las medidas, como no podía ser de otra manera pues en eso consiste la estimación, pero depende sólo en lo que respecta a su valor central,  $(x_1, \dots, x_2)$  (o bien  $(l_1, \dots, l_2)$  para las observaciones), pero no en su forma. Volviendo a unas distribuciones cualesquiera, rizando el rizo podríamos realizar estimaciones a priori de cómo será la función  $g$ , pero supone complicarse mucho (obtendríamos una distribución cuya variable es una función en lugar de un único valor), el resultado estaría muy relacionado con la función  $f$ , y de todas formas la información que queremos al diseñar es la que contiene la función  $f$ , que nos da probabilidades de errores en la estimación. Así, si para un parámetro concreto tenemos que la probabilidad de la distribución  $f(x)$  en un intervalo  $[x - h, x + h]$  es  $p$ , esto quiere decir que la probabilidad de que el valor estimado para ese parámetro como resultado del ajuste tenga un error menor que  $h$  es  $p$ .

### Redundancias parciales

De todo lo que podemos obtener a priori acerca de las distribuciones de los valores ajustados, los cocientes  $\sigma^2_{l_i} / \sigma^2_{\ell_i}$  y su complemento

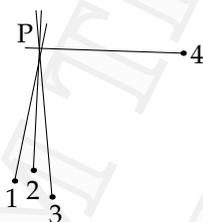
## 10 Regiones de confianza

a uno  $\sigma^2_{v_i} / \sigma^2_{\ell_i}$  hemos visto que tienen un significado particular. Los segundos son  $E[v_i] / \varepsilon_i$ , y los denotaremos por  $r_i$ . Para el estimador mínimo cuadrático no dependen de  $\varepsilon_i$ , pero en general serán funciones de  $\varepsilon_i$ :

$$r_i(\varepsilon_i) = \frac{E[v_i] \varepsilon_i}{\varepsilon_i}. \quad (10.19)$$

Hemos demostrado ya que  $0 < r_i < 1$  (si  $r_i = 1$  estamos observando un parámetro fijo, y la observación sobra, y si  $r_i = 0$  la observación obliga a fijar un parámetro que de no existir la observación fijaríamos libremente, y también sobra en el ajuste).  $r_i$  será pequeño cuando el ajuste no pueda estimar muy bien el error de la observación, y tienda a adaptarse al valor observado  $\ell_i$ . Esto a su vez ocurrirá cuando la observación tenga poca comprobación mediante las demás observaciones, o expresado de una manera un poco más precisa, si excluyésemos la observación del ajuste, entonces la precisión con la que quedaría determinada esa magnitud como resultado del ajuste sería muy mala.

Por ejemplo, sea la determinación de las coordenadas de un punto mediante 4 visuales desde puntos con coordenadas conocidas como se indica en la figura.



Supongamos para simplificar que las desorientaciones también son conocidas. El ajuste se ajustará a la visual 4, ya que puede hacerlo sin apenas variar el residuo en las demás visuales. Por lo tanto  $\sigma^2_{v_4}$  será muy pequeño, y también  $r_4$ . Visto de otra manera, si suprimiésemos la visual 4, como resultado del ajuste mediante las otras tres visuales obtenemos un valor ajustado para la visual de 4 a P, es decir, una estimación del valor real de esa visual, pero con una precisión muy mala.

Podemos decir también que la redundancia del conjunto de observaciones es  $r = 2$  (recordamos que  $r = m - n$ ), pero que esas dos

redundancias no están repartidas por igual, y se acumulan en las visuales 1, 2 y 3, mientras que 4 es casi imprescindible. Esta interpretación da a los  $r_i$  el nombre de redundancias parciales. Esta designación es consistente con el concepto de redundancia total, ya que se cumple  $\sum r_i = r$ , como vamos a demostrar.

Supongamos que tomamos  $\sigma_0 = 1$ , para no tener que arrastrar esta constante en todos los desarrollos. Entonces  $r_i$  es el elemento  $i$  de la diagonal principal de  $\Sigma_{vv}$  (no.) y por lo tanto

$$\sum r_i = \text{Tr}(\Sigma_{vv}).$$

Puede demostrarse fácilmente la igualdad  $\text{Tr}(AB) = \text{Tr}(BA)$ , y además es trivial que si A y B son matrices cuadradas entonces se cumple  $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$ . Entonces,

$$\begin{aligned} \text{Tr}(\Sigma_{vv}) &= \text{Tr}(\mathbf{I}) - \text{Tr}(\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T) = m - \text{Tr}(\mathbf{A}\mathbf{A}^T\mathbf{N}^{-1}) = m - \text{Tr}(\mathbf{I}) \\ &= m - n = r. \end{aligned}$$

Puesto que el estimador mínimo cuadrático es centrado, se cumple  $E[v_i] = 0$  para todo  $i$ , y entonces, por definición,  $E[v_i^2] = \sigma_{v_i}^2$ , de modo que también podemos definir la redundancia parcial como

$$r_i = \frac{E[v_i^2]}{\sigma_{\ell_i}^2}. \quad (10.20)$$

No obstante, la primera definición es más restrictiva, ya que atiende al signo, y además proporciona un resultado en función de  $\varepsilon_i$ , aunque para el estimador mínimo cuadrático es constante.

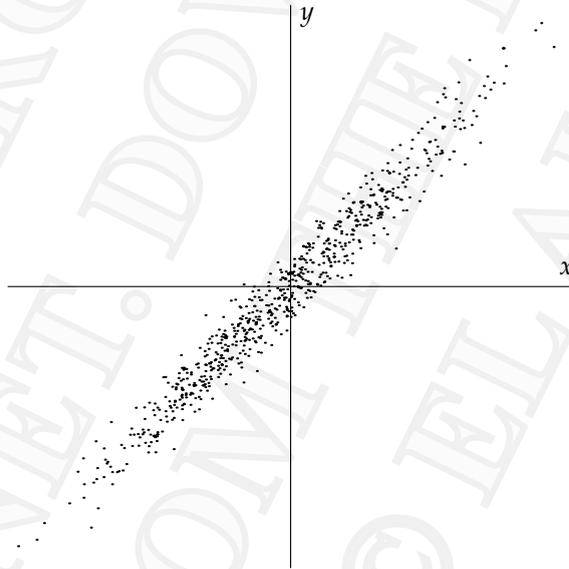
Además del caso visto, también puede tener muy poca redundancia una observación que es mucho más precisa que las demás. Supongamos que el ajuste lo realizamos eliminando esa observación; entonces el valor ajustado obtenido seguramente esté separado del valor observado varias veces la desviación típica. Si incluimos la observación, la anterior solución dará un valor muy elevado para  $v^2$  y  $v'^2$  en esa observación, y en consecuencia para  $\sum v'^2$ . Si variamos los parámetros de manera que el valor ajustado se aproxime al observado en esa observación,  $v'^2$  de la observación en cuestión disminuirá mucho, mientras que el aumento en las demás observaciones será menor, al tener una desviación típica mucho mayor, y en consecuencia

## 10 Regiones de confianza

la solución se ajusta al valor de la observación precisa, como es lógico. Dicho de otra manera, una variación de los parámetros produce en la observación precisa una variación de  $v'$  mucho mayor que en las demás. Este caso y el anterior tienen en común que si realizamos el ajuste sin la observación, el valor ajustado obtenido para ella tiene una precisión muy mala en relación a la de la propia observación. En caso contrario se obtendrán redundancias altas.

### Correlaciones

Supongamos que se tienen dos variables  $x, y$  que siguen la distribución que se muestra:



Centradas en un punto  $(x, y)$ , resultado de un cálculo.

Cada una de las dos variables se conoce con una cierta precisión, pero la precisión de la resta es mucho más alta que la de cada una de las dos individualmente:

$$\sigma_{y-x} \ll \sigma_x, \sigma_y.$$

## 10 Regiones de confianza

Esto es así porque los errores en  $x$  y en  $y$  están muy relacionados, de manera que si el valor  $x$  es mayor que  $x$  es muy probable que  $y$  también sea mayor que  $y$ , de modo que la resta  $y - x$  varía muy poco aunque  $x$  e  $y$  varíen.

Si obtenemos las elipses de probabilidad constante del par de variables  $x, y$  éstas serán cortas en el sentido de variación de  $y - x$  y largas en el sentido perpendicular. Si obtenemos  $\sigma_{y-x}$  según lo visto en este capítulo para la transmisión de varianzas obtenemos

$$\sigma_{y-x}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}.$$

El último término de esta relación es la covarianza entre las variables  $x$  e  $y$ , definida como  $E[(x - \mu_x)(y - \mu_y)]$  según ya vimos. Si las dos variables son independientes; es decir, si el conocimiento del valor de una de ellas no aporta nada al conocimiento, o a «lo que podemos esperar» de la otra, se tiene simplemente

$$\sigma_{xy} = E[(x - \mu_x)(y - \mu_y)] = E[x - \mu_x]E[y - \mu_y] = 0 \cdot 0 = 0.$$

Pero si el conocimiento de una de ellas afecta a lo que podemos esperar de la otra, como en este caso en el que las cantidades  $y$  y  $x$  son siempre muy parecidas, la esperanza de ese producto puede ser muy distinta de cero, positiva si la correlación es positiva y negativa si es negativa. En este caso es positiva y bastante alta y por ello la suma  $\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$  es mucho menor que  $\sigma_x^2 + \sigma_y^2$ . Nunca puede ser mayor que  $\sigma_x\sigma_y$ . Cuanto más se acerque a este límite más se agrupan los pares  $(x, y)$  en torno a una recta; de pendiente positiva si la covarianza es positiva y negativa si es negativa. Por esta razón se define el *coeficiente de correlación* como

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}, \quad -1 \leq \rho \leq 1. \quad (10.21)$$

El caso de dos coordenadas, una de un punto y otra de otro, la misma en ambos (p.e., dos coordenadas  $x$  o dos coordenadas  $z$ ) que tras un ajuste quedan con un valor de  $\rho$  próximo a 1 es muy común. Esto es así porque la precisión de las coordenadas se refiere a la definición del sistema, que se habrá hecho como sea, y en una zona de la red que se ajusta es de esperar que los errores se den en el mismo sentido, de modo que si las coordenadas calculadas para un punto que

## 10 Regiones de confianza

ha resultado con precisión de 5 cm tienen un error (que no podemos saber) de 4 cm es de esperar que un punto próximo tenga un error en torno a  $4 \text{ cm} \pm 5 \text{ cm}$ .

Un ejemplo clarísimo de esta situación es una gran red de nivelación con el datum fijado en un punto. Si comenzamos nivelando en Alicante con error 0 y la precisión de cada nivelada es menor de una décima de milímetro, al llegar a Galicia la precisión de los puntos será de, por ejemplo, 5 mm. De modo que un punto A de la red puede tener una precisión en su cota de 5 mm y otro punto próximo B de 5,005 mm. Esto no significa que la precisión del desnivel A-B sea de

$$\sqrt{5^2 + 5,005^2} \approx 7 \text{ mm}$$

como la mayoría de los topógrafos calcularía, sino que puede perfectamente ser de 0,1 mm. Para conocer el valor exacto necesitamos saber  $\sigma_{AB}$ . Supongamos  $\sigma_{AB} = 25$ . Entonces

$$\sigma_{B-A}^2 = 5^2 + 5,005^2 - 2 \cdot 25 = 0,05,$$

$$\sigma_{B-A} = \sqrt{0,05} \approx 0,22 \text{ mm}.$$

En este ejemplo el coeficiente de correlación es

$$\rho_{AB} = \frac{25}{5 \cdot 5,005} = 0,9990.$$

Un valor tan alto significa que el error en una de las dos cotas queda prácticamente determinado por el error en la otra. También se puede interpretar como que la causa de error en una de ellas es esencialmente la misma que en la otra. Teniendo en cuenta que los errores se componen cuadráticamente, podemos decir que los puntos A y B tienen una proporción de error común de 0,999 frente a una componente independiente del otro punto de solamente 0,032. Por ello la precisión con la que se conoce B - A es mucho mejor que la de B y A individualmente.

Otro caso muy habitual es el de los parámetros de orientación de fotogramas como resultado de una aerotriangulación. Podemos entender el conjunto de puntos de apoyo, puntos de aerotriangulación no de apoyo y centros de proyección de las fotografías como una red peculiar, que queda fijada por las coordenadas de unos puntos (los

## 10 Regiones de confianza

puntos de apoyo) y en la que solamente se han hecho observaciones desde un subconjunto de sus puntos (los centros de proyección) a los demás puntos. Es como si los centros de proyección fuesen sólo puntos de estación y los demás puntos solamente puntos visados, determinándose las coordenadas de los puntos de estación por intersecciones inversas entrelazadas. Siendo así la red y teniendo en cuenta la geometría de las aerotriangulaciones, en donde a diferencia de una buena intersección inversa topográfica los puntos visados se encuentran todos en torno a una dirección respecto a la estación, la precisión de los puntos de estación, e.d., de los centros de proyección, es mucho peor que la de los puntos de terreno, y también la precisión de sus giros. Sin embargo la precisión de la coordenada  $X$  y el giro  $\Phi$  por una parte, y de la coordenada  $Y$  y el giro  $\Omega$  por otra, se combinan de manera que la precisión de la intersección de un rayo proyectivo con el suelo, estando definida esta precisión por la precisión de los parámetros de orientación del centro de proyección, y suponiendo que no existe error en las fotocoordenadas, es mucho mejor que la precisión de la posición  $(X, Y)$  del centro de proyección (lo mismo se sigue cumpliendo para fotografía no vertical de objetos cercanos).

El siguiente ejemplo está tomado de una aerotriangulación cualquiera de varias pasadas con los puntos de apoyo bien distribuidos:

$$\sigma_X^2 = 0,5806 \text{ (m}^2\text{)}, \quad \sigma_\Phi^2 = 0,00003835 \text{ (deg}^2\text{)},$$

$$\sigma_{X\Phi} = -0,004435 \text{ (m} \cdot \text{deg)}.$$

Estos valores dan como resultado un coeficiente de correlación de  $\rho_{X\Phi} = -0,9400$ . Por ser negativo quiere decir que a mayor valor de  $X$  podemos esperar un menor valor de  $\Phi$ . Haciendo un cálculo aproximado, teniendo en cuenta que la fotografía es casi horizontal y la altura de vuelo sobre el terreno de unos 6 000 m, la coordenada  $X$  de la intersección del eje óptico de la foto con el terreno es

$$X_T = X_C + 6\,000 \Phi = X_C + 105 \Phi^\circ.$$

La precisión de  $X_T$  según los valores anteriores de precisiones es

$$\begin{aligned} \sigma_{X_T}^2 &= 0,5806 + 105^2 \cdot 0,00003835 + 2 \cdot 105 \cdot (-0,004435) = 0,072, \\ \sigma_{X_T} &= 0,27 \text{ m.} \end{aligned}$$

## 10 Regiones de confianza

Compárese esto con la precisión de la coordenada X del centro de proyección:  $\sigma_X = \sqrt{0,58} = 0,76$  m.

Normalmente las correlaciones juegan a favor del topógrafo, ya que si dos magnitudes tienen cada una una cierta precisión pero la resta u otra combinación de ellas queda con una precisión mucho mejor, suele ser esta última cantidad la que interesa. En una red, si se va a emplear un pequeño número de puntos para un trabajo local, la precisión de este último viene dada por la precisión interna de los puntos de la red empleados, es decir, la precisión de unos respecto a otros. Normalmente las correlaciones se ignoran, y si estos puntos tienen una precisión de 5 cm, por ejemplo, se harán los cálculos de la precisión del trabajo tomando cada uno de esos 5 cm como independientes, dando lugar por ejemplo a precisiones calculadas de 7 cm, cuando la realidad es que pueden ser de 3 cm. Incluso la precisión absoluta respecto a los puntos que definen el sistema de coordenadas será menor de 7 cm; serán los 5 cm de los puntos de la red empleados, ya que siendo la precisión de uno respecto a otro mucho mejor, el error añadido al dar los puntos del trabajo local a partir de ellos será pequeño respecto a los 5 cm. La cuestión de las precisiones relativas entre puntos se tratará con detalle en dos capítulos dedicados a ello.

La ignorancia de las correlaciones, o en términos menos técnicos, de que del valor de precisión de dos o más coordenadas o parámetros una parte importante es un error común a ambos, da lugar como acabamos de ver a que se especifiquen para coordenadas que se derivan de ellos valores de precisión mayores, es decir peores, que los que realmente son. Por ello en topografía la precisión de los puntos suele ser mejor que la que el topógrafo dice que tienen. Hoy en día con el empleo cada vez mayor del GPS esto está cambiando, pudiendo ser que la precisión sea peor que el valor que el topógrafo da, debido a causas de error que no están incluidas en el valor de precisión que el aparato está mostrando y que el topógrafo no tiene en cuenta. Aun así la precisión de las coordenadas de puntos próximos entre sí suele ser mejor que el valor dado debido a las causas de error comunes.

# 11 Altimetría y otros

## Ajuste altimétrico

En topografía se presentan básicamente dos tipos distintos de ajuste altimétrico: uno en el que las observaciones son directamente los desniveles (nivelación geométrica), y otro en el que se miden distancia y ángulo cenital (nivelación trigonométrica). El primero de ellos da lugar a ecuaciones lineales muy simples del tipo

$$v_H = H - H_0 - (\Delta Z_2 - \Delta Z_1),$$

cuya solución no entraña ninguna particularidad. Empleamos H para el desnivel entre dos puntos.

En el segundo caso habrá que plantear ecuaciones de residuo para la distancia D y para el ángulo cenital V. Sin embargo se acostumbra a proceder de otro modo. Se calcula el desnivel a partir de los valores observados y el resultado se toma como la observación original:

$$H = D \cos V + \frac{S^2}{2R}.$$

El último término es la corrección por esfericidad terrestre, siendo R el radio de la Tierra y S la distancia horizontal entre los dos puntos, al igual que en el capítulo dedicado a ángulos y distancias. En teoría debe emplearse la distancia S calculada mediante los parámetros (las coordenadas (X, Y) de ambos puntos), si se dispone de ellos, pero la corrección por esfericidad no varía significativamente con pequeños

## 11 Altimetría y otros

cambios en la distancia, así que puede tomarse  $S = D \sin V$ . Puede añadirse también una corrección por refracción, aunque la influencia de este fenómeno puede ser muy variable. Las observaciones de cara al ajuste son entonces los desniveles, y se resuelve igual que una nivelación geométrica.

Respecto a los parámetros a tomar como fijos, el estudio es idéntico al efectuado en planimetría. Habrá que fijar una coordenada  $Z$ , salvo que la situación de la figura venga obligada por varias coordenadas ya fijas, típicamente de una red superior. Pueden existir también coordenadas observadas directamente, y se planteará su ecuación de residuo. Si de esta especie sólo hay una (y no hay coordenadas ya fijas), entonces no entra en el ajuste y simplemente sirve para tener que tomar como fija obligatoriamente esa coordenada, que de no existir la observación escogeríamos a discreción.

### Ecuaciones de distancia geométrica y ángulo cenital

El ajuste de ángulos y distancias visto unos capítulos atrás toma como observaciones las lecturas y las distancias horizontales. Estas últimas normalmente no son observaciones originales, sino que se obtienen como  $D \sin V$  y sus valores se toman como observaciones originales, procediendo de modo análogo que con los desniveles. Ambas operaciones pueden interpretarse conjuntamente como un cambio de variables  $(L, D, V) \rightarrow (L, S, H)$ . Para ambos conjuntos resultan adecuados los parámetros  $(\Sigma, X, Y, Z)$ , pero el segundo tiene la propiedad de que para expresar las observaciones  $(L, S)$  sólo son necesarios los parámetros  $(\Sigma, X, Y)$ , mientras que para las observaciones  $H$  sólo se necesitan los parámetros  $Z$ . Entonces, si las observaciones  $(L, S, H)$  son estadísticamente independientes se pueden efectuar por separado el ajuste planimétrico y el altimétrico. De hecho el cálculo conjunto lo sería sólo aparentemente, ya que las matrices  $A$  y  $N$  tendrían un bloque planimétrico, otro altimétrico y otro(s) de ceros, y en el cálculo matricial nunca se operarían entre sí dos números provenientes del bloque planimétrico y del altimétrico.

Suponemos que las observaciones  $(L, D, V)$  son independientes, y entonces la independencia de  $(L, S, H)$  depende de la de  $S$  y  $H$ . Éstas últimas podemos suponer sin cometer gran error que son independientes, gracias a que las visuales tienen normalmente poca inclinación. Se puede ver gráficamente que en esa situación el error en  $H$

depende casi en exclusiva del error en  $V$ , y el error en  $S$  del error en  $D$ , y entonces podemos admitir que son independientes.

En situaciones de grandes inclinaciones, como por ejemplo en presas, se saca más partido a las observaciones si se realiza un ajuste conjunto de planimetría y altimetría, con las observaciones originales. La ecuación de residuo de la lectura no varía, y la de la distancia  $D$  es una extensión obvia de la de  $S$ :

$$v_D \approx D - D_0 - \left\{ \left( \frac{X_2 - X_1}{D} \right)_0 \Delta X_1 - \left( \frac{Y_2 - Y_1}{D} \right)_0 \Delta Y_1 - \left( \frac{Z_2 - Z_1}{D} \right)_0 \Delta Z_1 + \left( \frac{X_2 - X_1}{D} \right)_0 \Delta X_2 + \left( \frac{Y_2 - Y_1}{D} \right)_0 \Delta Y_2 + \left( \frac{Z_2 - Z_1}{D} \right)_0 \Delta Z_2 \right\}. \quad (11.1)$$

La ecuación del ángulo cenital sin linealizar e ignorando la esfericidad terrestre es

$$V = \arctan \frac{S}{Z_2 - Z_1}.$$

Para el cálculo de  $V_0$  así como para el de  $D_0$  debe tenerse en cuenta la esfericidad terrestre, en el grado que el problema requiera, pero las derivadas de ambas magnitudes en función de los parámetros apenas varían si se emplean las fórmulas exactas en lugar de las que corresponderían a un sistema  $X, Y, Z$  cartesiano, y podemos emplear estas últimas. Ya lo hicimos para  $D$  y en su momento para  $S$ , aunque no lo mencionamos. Para  $V$  resulta la ecuación linealizada

$$v_V \approx V - V_0 - \left\{ \left( \frac{(X_2 - X_1)(Z_2 - Z_1)}{SD^2} \right)_0 \Delta X_1 - \left( \frac{(Y_2 - Y_1)(Z_2 - Z_1)}{SD^2} \right)_0 \Delta Y_1 + \left( \frac{S}{D^2} \right)_0 \Delta Z_1 + \dots \right\}. \quad (11.2)$$

Las derivadas respecto a las coordenadas del segundo punto son las mismas que respecto al primero cambiadas de signo.

### Error de centrado

Existen muchos errores que si bien son accidentales, una vez producidos son constantes para un conjunto de observaciones, como por ejemplo el error de centrado al estacionar, que permanece constante

## 11 Altimetría y otros

en toda la estación. Estos errores se pueden incluir como parámetros en el ajuste. Representaremos el error de estación en un punto  $i$  por  $x_i, y_i$ . Al considerar como parámetros (modelizar) lo que antes era una componente del error de cada lectura, ahora se incluye en la lectura ajustada, que pasa a ser

$$L_1^2 = \arctan \frac{Y_2 - (Y_1 + y_1)}{X_2 - (X_1 + x_1)} - \Sigma_1.$$

Ahora los errores y sus estimaciones los residuos serán lo que reste, eliminada una causa de error. Desde un punto de vista conceptual para el ajuste, error (o residuo) es todo lo que no está modelizado mediante los parámetros.

No podemos tomar el (des)centrado de una estación como incógnita sin más, ya que equivale a no haber centrado en absoluto, o a considerar el punto de estación y el mismo punto cuando es visado como dos puntos distintos. Si hemos centrado al estacionar, y mediante algún dispositivo de observación (masa suspendida, visual vertical, ya sea a través de un retículo o con láser...) hemos situado en la medida de lo posible el eje vertical del aparato sobre el punto; es decir, hemos observado  $x$  e  $y$  y obtenido los valores  $x = 0, y = 0$ . Si los valores ajustados son otros  $(x, y)$ , hay que considerarlo como un error, o más exactamente como una estimación del error de estación, y hay que plantear las ecuaciones de residuo

$$v_x = 0 - x_0 - \Delta x, \quad v_y = 0 - y_0 - \Delta y.$$

Normalmente tomaremos  $x_0 = y_0 = 0$ , lo que simplifica aún más las ecuaciones. Estas ecuaciones habrá que multiplicarlas como todas por su correspondiente  $\sigma_0/\sigma_i$ .

### Estaciones excéntricas

El procurar centrar al estacionar, que para el ajuste se interpreta como observar los valores  $x = y = 0$ , es un caso particular de otro más general: estaciones excéntricas. En este caso general tenemos unos valores observados  $(x, y) \neq (0, 0)$ . Al igual que en un ajuste con estaciones «normales», podemos incluir los parámetros de estación  $(x, y)$

## 11 Altimetría y otros

en el ajuste o no hacerlo. Si los incluimos, las ecuaciones de residuo para la estación serán

$$v_x = x - x_0 - \Delta x, \quad v_y = y - y_0 - \Delta y.$$

En caso de no incluirlos estas ecuaciones no existirán, y en el cálculo de lecturas ajustadas los valores  $(x, y)$  se tomarán como constantes. Por ejemplo, si son  $x = 0,08$ ,  $y = 0,42$ , entonces la fórmula para la lectura ajustada es

$$L_1^2 = \arctan \frac{Y_2 - (Y_1 + 0,42)}{X_2 - (X_1 + 0,08)} - \Sigma_1.$$

Existen muchas posibilidades a la hora de incluir parámetros que modelizan errores, y muchas más combinaciones entre ellas. Debe tenerse cuidado siempre de no incluir conjuntos de parámetros que den lugar a infinitas soluciones. Por ejemplo, no podemos tomar para un punto de estación las coordenadas  $(X + x', Y + y')$  y para el mismo punto cuando es visado otras  $(X + x'', Y + y'')$ , pretendiendo estimar un descentrado también en la marca a la que se visa, porque entonces, ¿qué es el punto  $(X, Y)$ ? Los parámetros  $(x, y)$  son el desplazamiento entre el punto de estación y el punto visado medio, y  $(X, Y)$  las coordenadas de este último. Se ve bien en el caso de las estaciones excéntricas: podemos tomar como referencia la estación o el punto visado indistintamente, y los parámetros  $(x, y)$  son el desplazamiento de uno respecto al otro.

### Alturas de instrumento y de mira

El equivalente en altimetría a los parámetros de centrado son las alturas de instrumento y de mira. Podemos incluir una de las dos como parámetro, pero no las dos. Si las hemos medido, que será lo normal, entonces habrá que plantear la correspondiente ecuación de residuo.

Otra posibilidad puede ser que la mira tenga (o pensemos que puede tener) un error constante, situación muy habitual. Entonces las ecuaciones de desnivel tienen todas un parámetro de cero:

$$v_H = H - H_0 - (\Delta Z_2 - \Delta Z_1 + \Delta z), \quad H_0 = Z_{20} + z_0 - Z_{10}.$$

## 11 Altimetría y otros

Si realizamos un ajuste conjunto tomando las observaciones originales,  $(L, D, V)$ , tendremos en las fórmulas  $(Z_2 + z)$  en lugar de  $Z_2$ , y las derivadas de  $z$  son iguales a las de  $Z_2$  (pero no son el mismo parámetro porque  $z$  aparece en un conjunto de ecuaciones distinto a  $Z_2$ ).

Si no tenemos ninguna garantía sobre la ausencia de error constante en la mira entonces el valor  $z$  no se considerará un residuo a incluir en la minimización. Sin embargo si se trata de una señal precisa, cuya altura está calibrada con una precisión conocida, sí que habrá ecuación de residuo.

### Error de refracción

Otro posible parámetro a considerar es el error de refracción, siempre que las condiciones atmosféricas no varíen mucho a lo largo de las observaciones. Suele ser un efecto proporcional al cuadrado de la distancia.

### Error de verticalidad de la estación

Otro más es el error de verticalidad de una estación. Si es pequeño podemos modelarlo mediante dos parámetros  $(\omega, \varphi)$ . Llamaremos  $L^*$  y  $V^*$  a los valores que corresponderían con  $\omega = \varphi = 0$ . Si  $(\omega, \varphi)$  no son muy grandes (y no lo serán), podemos emplear las siguientes relaciones:

$$\begin{aligned}\cos V &= \cos V^* + (\omega \cos L^* - \varphi \sin L^*) \sin V^*, \\ \tan L &= \frac{\sin L^* + \varphi \cotg V^*}{\cos L^* - \omega \cotg V^*}.\end{aligned}$$

Que a su vez se pueden simplificar a las siguientes:

$$\begin{aligned}V &= V^* - \omega \cos L^* + \varphi \sin L^*, \\ L &= L^* + (\omega \sin L^* + \varphi \cos L^*) \cotg V^*.\end{aligned}$$

El error cometido al emplear estas expresiones es del orden de  $\omega^2 + \varphi^2$ . Habrá que incluir también las ecuaciones de residuo de  $\omega$  y  $\varphi$ , cuyo  $\sigma$  es la precisión con la que se nivela el aparato.

**Parámetros adicionales: precisiones y redundancias**

La inclusión de parámetros extra permite obtener el máximo provecho de todas las observaciones, incluyendo entre ellas el centrado y la nivelación del aparato, que realmente lo son. En estos dos casos, además, no disminuye el número de redundancias al considerarlas como incógnita, ya que cada nuevo parámetro va acompañado de una observación. No obstante, si su precisión es mucho mayor que la que se obtendría para esos mismos valores a partir de las demás observaciones no merece la pena incluirlas. Por ejemplo, si la precisión con la que se obtendría un punto por intersección de las visuales que concurren en él es mucho menor que la precisión de centrado, entonces los parámetros  $(x, y)$  no se incluirán. Sí que disminuirán las redundancias si se modelizan errores que no se han observado, como por ejemplo el de refracción.



# 12 Transformación de semejanza plana

Otro problema muy habitual en topografía o en cualquier disciplina que trabaje con coordenadas de puntos en diversos sistemas es el cálculo de los parámetros de transformación entre dos sistemas de coordenadas distintos. Muchas veces se trata de transformaciones de semejanza, aunque pueden ser también transformaciones lineales cualesquiera u otras más complicadas. El estudio de la semejanza en el plano abarca la mayoría de los aspectos que se pueden presentar en transformaciones más complejas y merece por ello un análisis pormenorizado.

## Notación y convenciones generales

Habrá siempre un conjunto de puntos  $P_1, \dots, P_k$ , cuyas coordenadas conocemos en dos sistemas, 1 y 2. Las coordenadas de un punto en un sistema determinado las representaremos mediante un subíndice. Así, las coordenadas de los puntos en el sistema 1 son  $P_{11}, P_{21}, \dots, P_{k1}$ , y en el sistema 2 son  $P_{12}, P_{22}, \dots, P_{k2}$ . Las coordenadas de ambos sistemas son medidas con ciertas precisiones, pero algunas veces es necesario considerar las coordenadas del sistema 1 como si fuesen perfectas. Estudiaremos ambos casos. De los dos sistemas normalmente hay uno que se considera el de referencia, que para nosotros será siempre el sistema 1, y cuando se da el valor numérico de los parámetros de la transformación se entiende que son los que transforman del sistema de referencia al secundario, en nuestro caso

## 12 Transformación de semejanza plana

esto es del sistema 1 al sistema 2. Cuando queramos hacer referencia a todo el conjunto de coordenadas de los puntos en algún sistema  $i$  diremos «los puntos  $P_i$ ». Así,  $P_i$  puede significar el punto  $P_i$  o el conjunto de coordenadas de los puntos  $P$  en el sistema  $i$ . En cada momento el contexto determinará claramente cuál de los dos significados se está empleando. Análogamente habrá magnitudes asociadas a cada punto y sistema, que representaremos mediante alguna letra, por ejemplo  $r$ , teniendo  $r_i$  los mismos dos significados. Cuando queramos referirnos a un punto cualquiera del espacio, generalmente cuando escribimos la fórmula de la transformación, emplearemos la letra  $X$ .

Aunque la realidad sea un único conjunto de puntos con coordenadas en dos sistemas, para comprender los parámetros suele ser más claro representarlo como dos conjuntos de puntos bajo un mismo sistema, y así se suele dibujar. Alguna vez, pensando en esa representación, haremos referencia a los puntos de tal o cuál conjunto, entendiendo que en realidad se trata de las coordenadas del conjunto de puntos en el sistema tal o cuál.

### Convenciones para los parámetros de semejanza

Los parámetros de una transformación hay que definirlos muy claramente, porque pueden tener distintos valores según en qué orden se apliquen, y según sean los que pasan del sistema de referencia al secundario o viceversa. Respecto a esto último ya hemos aclarado que serán siempre los que transformen el sistema de referencia, 1, en el secundario, 2. Respecto a lo primero, en las transformaciones de semejanza y en cualquiera que incluya una traslación más algún otro parámetro, el valor de la traslación es distinto según sea su posición relativa respecto a los demás parámetros, que en las semejanzas son el giro y el escalado. Lo normal es situarla en un extremo, es decir, lo primero o lo último que se aplica. Tenemos entonces dos posibilidades:

$$X_2 = \lambda R(X_1 - T),$$

$$X_2 = \lambda R X_1 + T;$$

en donde  $T = (T_x, T_y)$ ,  $\lambda$  es el factor de escala y  $R$  es la matriz de rotación.

## 12 Transformación de semejanza plana

La matriz de rotación a su vez podemos definirla de dos maneras según el ángulo de giro  $\alpha$ :

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}, \quad \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \sin \alpha \end{pmatrix}.$$

La matriz de rotación es única, lo que cambia de una definición a otra es el signo del ángulo  $\alpha$ ; es decir, que realmente hay dos definiciones posibles para el ángulo  $\alpha$ , no para la matriz de rotación.

Los valores de  $R$  y  $\lambda$  son los mismos independientemente del orden de aplicación. Sin embargo podemos definir  $R$  y  $\lambda$  como la rotación y el factor de escala que pasan del sistema 1 al sistema 2 o viceversa, y en un caso los valores son los inversos de los otros. Nosotros emplearemos siempre la siguiente definición:

$$X_2 = \lambda R(X_1 - T), \quad R = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}. \quad (12.1)$$

La transformación de  $X_2$  a  $X_1$  según esta definición es

$$X_1 = \frac{1}{\lambda} R^T X_2 + T, \quad (12.2)$$

ya que la inversa de una matriz de rotación es su traspuesta.

### Planteamiento mínimo cuadrático

Empezaremos suponiendo que las coordenadas en el sistema 1 son fijas. En ese caso el problema queda determinado mediante los parámetros de transformación:  $(T_x, T_y, \alpha, \lambda)$ , y entonces  $n = 4$ . Las coordenadas de cada punto  $P$  en el sistema 2 son dos observaciones  $(x_2, y_2)$ . Las derivadas de cada una de ellas respecto de cada parámetro una vez simplificadas son

$$\begin{aligned} \frac{\partial x_2}{\partial T_x} &= -\lambda \cos \alpha, & \frac{\partial x_2}{\partial T_y} &= \lambda \sin \alpha, & \frac{\partial x_2}{\partial \alpha} &= -y_2, & \frac{\partial x_2}{\partial \lambda} &= \frac{x_2}{\lambda}, \\ \frac{\partial y_2}{\partial T_x} &= -\lambda \sin \alpha, & \frac{\partial y_2}{\partial T_y} &= -\lambda \cos \alpha, & \frac{\partial y_2}{\partial \alpha} &= x_2, & \frac{\partial y_2}{\partial \lambda} &= \frac{y_2}{\lambda}. \end{aligned}$$

Y se plantean las ecuaciones y se resuelven. Sin embargo se pueden calcular los parámetros de una manera más sencilla, que desarrollamos en las siguientes secciones. Lo mismo sucede en el caso en que no suponemos que las coordenadas en el sistema de referencia son fijas.

## 12 Transformación de semejanza plana

### Traslaciones

Continuaremos con la suposición de que las coordenadas del sistema 1 son fijas. Observemos en primer lugar que la transformación se puede escribir de la forma

$$(X_2 - T_2) = \lambda R(X_1 - T_1).$$

Hay infinitas maneras de escribirla así; todas las que cumplan  $T_1 - \frac{1}{\lambda}R^T T_2 = T$ . Supongamos que  $T_1$  es el centro de gravedad del conjunto de puntos del sistema 1, que llamaremos  $G_1 = (x_{g1}, y_{g1})$ . Tras restar  $T_1$ , el centro de gravedad queda con coordenadas  $(0,0)$ . Entonces  $R$  y  $\lambda$  son un giro y un factor de escala respecto al centro de gravedad, y tras aplicarlos las coordenadas de ese punto siguen siendo  $(0,0)$ . Ya sólo queda aplicar  $T_2$ , que es una traslación en  $x$  de valor  $T_{x2}$  y una en  $y$  de valor  $T_{y2}$ . A las coordenadas del conjunto antes de aplicar  $T_2$  las llamaremos  $P_1''$ . Tras aplicar  $T_2$  serán coordenadas en el sistema 2, y su diferencia con  $P_2$  un residuo. Si vamos variando  $T_{x2}$ , los residuos  $x$  de los puntos varían todos una cantidad constante, y análogamente para  $T_{y2}$ . Supondremos que las observaciones  $x$  e  $y$  son independientes, y entonces este es otra vez el problema de la media aritmética, que tiene por solución  $\sum pv = 0$ , en donde  $p$  son los pesos, inversamente proporcionales a  $\sigma^2$ . En este caso y suponiendo que las precisiones de las coordenadas son todas iguales,  $\sum v_x = 0$  y  $\sum v_y = 0$ . Ahora bien,  $\sum v_x = \sum x_2 - \sum(x_1'' + T_{x2})$ . Pero  $\sum x_1'' = 0$ , ya que  $x_g$  en ese sistema es 0, y entonces

$$T_{x2} = \frac{\sum x_2}{k} = x_{g2},$$

y análogamente  $T_{y2} = y_{g2}$ , por lo que  $T_2 = G_2$ , y el punto  $G_1$  se transforma en  $G_2$ :

$$(X_2 - G_2) = \lambda R(X_1 - G_1). \quad (12.3)$$

Si hubiésemos fijado  $T_2 = G_2$  entonces la conclusión sería  $T_1 = G_1$ .

Si las precisiones de las observaciones no son todas las mismas, la condición que cumple la traslación es

$$0 = \sum v'_x = \sum p(x_1'' + T_{x2}) - \sum px_2. \quad (12.4)$$

## 12 Transformación de semejanza plana

Entonces para que se cumpla  $\sum px_{1''} = 0$  tenemos que emplear el centro de gravedad ponderado según los pesos  $p$ :

$$T_{x1} = x_{g1} = \frac{\sum px_1}{\sum p}, \text{ y entonces } T_{x2} = \frac{\sum px_2}{\sum p} = x_{g2},$$

y análogamente para la coordenada  $y$ .

### Escala y giro: planteamiento

Llamaremos  $X_{1'}$  a las coordenadas  $X_1 - T_1$  y  $X_{2'}$  a  $X_2 - T_2$ . Si  $T_1 = G_1$  y  $T_2 = G_2$  entonces la transformación es

$$X_{2'} = \lambda R X_{1'},$$

en donde tenemos que calcular  $\alpha$  y  $\lambda$ . Sea  $P_{1''} = \lambda R P_{1'}$  y  $\delta_2 = (\delta_{x_2}, \delta_{y_2}) = P_{2'} - P_{1''}$  (podríamos llamarlo  $\delta_{2'}$ , pero la diferencia entre los sistemas 2 y 2' es una traslación, así que  $\delta_{2'} = \delta_2$ ). Si  $\lambda = \lambda$  y  $R = R$  entonces  $P_{1''}$  coincide con  $P_{2'}$  y  $\delta_2$  con  $(v_x, v_y)$ .

Supondremos que  $\sigma_x = \sigma_y$  para todos los puntos. Entonces hay que minimizar

$$\sum p((x_2 - x_{1''})^2 + (y_2 - y_{1''})^2) = \sum p(\delta_{x_2}^2 + \delta_{y_2}^2) = \sum p \overline{P_2 P_{1''}}^2.$$

Cada segmento  $\overline{P_2 P_{1''}}$  podemos separarlo en una componente radial y otra tangencial, curva, a las que llamaremos  $\delta_{r_2}$  y  $\delta_{t_2}$  (fig. 12.1).

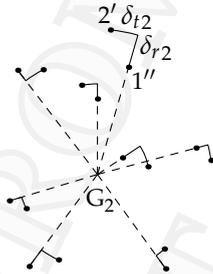


Fig. 12.1: Componentes radial y tangencial



## 12 Transformación de semejanza plana

Hay que minimizar  $\sum \left( \frac{v_2^2}{\sigma_2^2} + \frac{v_1^2}{\sigma_1^2} \right)$ ,  
que equivale a

$$\sum \left( \frac{v_{2'}^2}{\sigma_2^2} + \frac{v_{1''}^2}{(\lambda\sigma_1)^2} \right).$$

Entonces el punto  $P_{2'}$  es una media entre los puntos  $P_{2'}$  y  $P_{1''}$  con pesos  $1/\sigma_2^2$  y  $1/(\lambda\sigma_1)^2$ . La posición de  $P_{2'}$  así obtenida está a unas distancias de  $P_{2'}$  y  $P_{1''}$  proporcionales a  $\sigma_2^2$  y  $(\lambda\sigma_1)^2$ , es decir,

$$\overline{P_{2'}P_{2'}} = \frac{\sigma_2^2}{\sigma_2^2 + \lambda^2\sigma_1^2} \overline{P_{2'}P_{1''}}, \quad \overline{P_{1''}P_{2'}} = \frac{\lambda^2\sigma_1^2}{\sigma_2^2 + \lambda^2\sigma_1^2} \overline{P_{2'}P_{1''}}. \quad (12.5)$$

De donde

$$\begin{aligned} \frac{v_{2'}^2}{\sigma_2^2} + \frac{v_{1''}^2}{(\lambda\sigma_1)^2} &= \frac{\sigma_2^2}{(\sigma_2^2 + \lambda^2\sigma_1^2)^2} \overline{P_{2'}P_{1''}}^2 + \frac{\lambda^2\sigma_1^2}{(\sigma_2^2 + \lambda^2\sigma_1^2)^2} \overline{P_{2'}P_{1''}}^2 \\ &= \frac{1}{\sigma_2^2 + \lambda^2\sigma_1^2} \overline{P_{2'}P_{1''}}^2. \end{aligned}$$

Recordamos que la longitud  $\overline{P_{2'}P_{1''}}$  la habíamos denominado  $\delta_2$ . Entonces hay que hacer mínimo

$$\sum p\delta_2^2, \quad 1/p = \sigma_2^2 + \lambda^2\sigma_1^2. \quad (12.6)$$

Una vez calculados los parámetros ajustados conocemos  $\lambda$  y podemos calcular  $P_{2'}$ . Cuando minimizamos (12.6) los valores  $p$  son función de  $\lambda$ , y habría que tenerlo en cuenta a la hora de derivar para encontrar el mínimo. Sin embargo, si en la expresión de  $p$  empleamos en su lugar un valor  $\lambda_0$  fijo puede verse que la dependencia de la solución de ese  $\lambda_0$  es mínima. Si  $\lambda_0$  tiende a infinito los valores de  $p$  tienden a ser proporcionales a  $\sigma_1$ , y el resultado que se obtendría es el equivalente a considerar fijos los puntos del sistema 2. Si al contrario tiende a cero, entonces el resultado es el de considerar fijos los puntos  $P_1$ . La diferencia en los parámetros de la transformación entre un caso y otro es mucho menor que la precisión con la que se obtienen. Sí habrá diferencia apreciable en la posición de los puntos

## 12 Transformación de semejanza plana

ajustados, más próximos a  $P_{2'}$  o a  $P_{1''}$ . Con más razón la diferencia en los parámetros es despreciable si empleamos un valor  $\lambda_0$  cercano al valor real. Una vez obtenidos los parámetros, para calcular  $P_{2'}$  habría que emplear  $\lambda$ , pero normalmente incluso para esto la diferencia con  $\lambda_0$  será despreciable. Hay que tener en cuenta que  $\lambda$  aparece en el cálculo de  $P_{2'}$  porque  $\lambda\sigma_1$  es la desviación típica de los puntos  $P_{1''}$ , y normalmente el error en  $\lambda_0$  es mucho menor que la precisión con la que conocemos  $\sigma_1$ .

Visto esto, el problema es el mismo que el ya estudiado de considerar fijos los puntos del sistema 1, porque en ambos casos hay que minimizar  $\sum p\delta_2^2$ . Ahora los pesos no son  $1/\sigma_2^2$  sino que vienen dados por (12.6) (en ambos casos es  $1/\sigma_\delta^2$ ). Si además el cociente  $\sigma_1/\sigma_2$  se mantiene constante para todos los puntos, entonces también es constante  $\sigma_{\delta_2}/\sigma_1$ , y podemos obtener  $p$  como  $1/\sigma_1^2$  ó  $1/\sigma_2^2$  indistintamente (pero siempre la misma expresión para todos los puntos, claro está).

### Primeras fórmulas para escala y giro

Sea  $r_1$  la distancia al centro de gravedad de un punto del conjunto 1,  $\theta_1$  el ángulo que la semirrecta con origen en  $G_1$  y que pasa por el punto forma con una dirección origen de ángulos fija (por ejemplo, el sentido positivo del eje  $x$ ), medida en sentido antihorario desde el origen de ángulos, y  $t_1 = r_1\theta_1$ . Análogamente se definen  $r_2, \theta_2$  y  $t_2$ . El origen de ángulos ha de ser el mismo (en cada sistema ha de formar el mismo ángulo con los ejes).

Se tiene  $r_{1''} = \lambda r_1$ , de donde  $\delta_{r_2}^2 = (r_2 - \lambda r_1)^2$  y

$$\frac{d\sum p\delta_{r_2}^2}{d\lambda} = -2\sum p(r_2 - \lambda r_1)r_1,$$

y el valor de  $\lambda$  es

$$\lambda = \frac{\sum pr_1 r_2}{\sum pr_1^2}. \quad (12.7)$$

Por otra parte  $\theta_{1''} = \theta_1 + \alpha$ , de donde

$$\sqrt{p}\delta_{t_2} = \sqrt{p}r_2((\theta_1 - \theta_2) - \alpha),$$

## 12 Transformación de semejanza plana

y el valor  $\alpha$  es la media aritmética de  $\theta_2 - \theta_1$  con pesos  $pr_2^2$ :

$$\alpha = \frac{\sum pr_2^2(\theta_2 - \theta_1)^2}{\sum pr_2^2}. \quad (12.8)$$

En todas las expresiones anteriores los pesos vienen dados por (12.6).

### Fórmulas simétricas para escala y giro

Si intercambiamos los sistemas 1 y 2 obtenemos

$$\lambda' = \frac{\sum pr_1 r_2}{\sum pr_2^2}, \quad \alpha' = \frac{\sum pr_1^2(\theta_1 - \theta_2)}{\sum pr_2^2}.$$

Los pesos no es necesario cambiarlos ya que serían

$$1/p = \sigma_1^2 + \frac{1}{\lambda_0^2} \sigma_2^2 = \frac{1}{\lambda_0^2} (\sigma_2^2 + \lambda_0^2 \sigma_1^2);$$

es decir, proporcionales a los ya existentes.

El resultado en un caso y en otro no es el mismo; es decir, no se cumple  $\lambda' = 1/\lambda$  (incluso se puede ir más allá y demostrar que se cumple siempre  $\lambda\lambda' > 1$ ) ni  $\alpha' = -\alpha$ . La razón es en ambos casos bastante sutil. En el primero radica en que no es lo mismo minimizar  $\sum p(r_2 - \lambda r_1)^2$  que  $\sum p(\frac{r_2}{\lambda} - r_1)^2$ ; en el segundo en que para calcular  $\alpha$  estamos tomando los arcos  $\delta_i$  desde los puntos  $2'$ , y por lo tanto calculándolos con las distancias  $r_2$ , y en el segundo desde  $1'$  y los radios son  $r_1$ . La diferencia entre ambas soluciones es ínfima. De hecho, cuando descompusimos  $\delta_2$  en  $\delta_{r_2}$  y  $\delta_{t_2}$ , que pudimos hacerlo porque  $\delta_{r_2}^2 + \delta_{t_2}^2 \approx \delta_{x_2}^2 + \delta_{y_2}^2$ , fue una decisión arbitraria tomar los arcos desde  $P_{2'}$  y no desde  $P_{1''}$ .

También se puede ver que la diferencia es totalmente despreciable advirtiendo que la expresión de  $\lambda$  es una media aritmética de los valores  $r_2/r_1$  ponderada con los pesos  $pr_1 r_2$ , mientras que  $\lambda'$  es una media de  $r_1/r_2$  con los mismos pesos. A su vez,  $\alpha$  es una media de  $\theta_2 - \theta_1$  con los pesos  $pr_2^2$ , y  $-\alpha'$  la media de los mismos valores con los pesos  $pr_1^2$ . El cociente  $r_2/r_1$  es prácticamente constante para todos los puntos, y en consecuencia la diferencia entre ambas soluciones es

## 12 Transformación de semejanza plana

despreciable. Puede darse un valor de  $r_2/r_1$  más distinto si un punto coincide casi exactamente con el centro de gravedad, pero en ese caso la contribución de ese punto a cualquiera de los sumatorios es prácticamente cero.

Sin embargo conviene tener una definición de los parámetros que sea independiente de cuál consideremos el conjunto 1 y cuál el conjunto 2. Para ello lo más fácil es tomar para  $\lambda$  la media geométrica entre las dos soluciones y para  $\alpha$  el resultado de calcularlo con los pesos  $pr_1r_2$ :

$$\lambda = \sqrt{\frac{\sum pr_1r_2}{\sum pr_1^2} \frac{\sum pr_2^2}{\sum pr_1r_2}} = \sqrt{\frac{\sum pr_2^2}{\sum pr_1^2}}, \quad \alpha = \frac{\sum pr_1r_2(\theta_2 - \theta_1)}{\sum pr_1r_2}. \quad (12.9)$$

### Análisis de las fórmulas obtenidas

La fórmula para el factor de escala es tan buena desde el punto de vista práctico como cabría desear. Aunque sobre el papel parezca que  $r^2$  es una expresión más elaborada que  $r$ , en la práctica lo primero que se obtiene es  $r^2$  como  $x^2 + y^2$ , y  $r$  es la raíz cuadrada de esa cantidad. Por lo tanto una fórmula que opere con los cuadrados de las distancias es más rápida de calcular y más sencilla de escribir en términos de las coordenadas de los puntos que una que operase con las propias distancias. Por otra parte, puesto que se trata de obtener un factor de escala en algún momento habrá que efectuar una división, y la fórmula obtenida solamente incluye una. Compárese la expresión de la fórmula propuesta en función de las coordenadas  $x, y$  de los puntos y se verá que es más sencilla que cualquiera otra.

Además de esta ventaja práctica incluye otra mucho más importante. Supongamos que alguien decide calcular el factor de escala promediando los factores de escala individuales que se deducirían para cada punto y análogamente para el giro (para evitar complicaciones supondremos que no se aplican pesos):

$$\lambda = \frac{1}{n} \sum \frac{r_2}{r_1}, \quad \alpha = \frac{1}{n} \sum (\alpha_2 - \alpha_1).$$

A parte de que en el caso de  $\lambda$  la fórmula no es simétrica, ya que de hecho se cumple  $\lambda\lambda'_1 > 1$ , estas fórmulas tienen el inconveniente de

## 12 Transformación de semejanza plana

que si uno de los puntos está muy próximo al centro de gravedad sus valores de  $r_2/r_1$  y  $\alpha_2 - \alpha_1$  pueden ser muy distintos de los demás y estropean la media.

Para solventar este problema habría que ponderar cada valor individual de manera que si el punto está muy próximo al centro de gravedad en cualquiera de los dos sistemas su peso se haga cero. Así, un peso igual a  $r_1 + r_2$  no sirve porque si una de las dos distancias es muy pequeña pero la otra no el peso no es pequeño, o no es lo suficientemente pequeño. Piénsese que esto es perfectamente posible sin necesidad de que el punto sea erróneo. Supongamos para simplificar que la mayoría de los puntos se encuentran a una distancia del centro de gravedad en torno a 1 y que el factor de escala entre los dos sistemas es también próximo a 1. El punto en cuestión puede estar a una distancia de  $1/40$  del centro de gravedad en uno de los dos sistemas y de  $1/400$  en el otro, lo que da un término  $r_2/r_1$  de 10, que a pesar de un peso de 0,025 (de acuerdo a  $r_1 + r_2$ ) sigue alterando el valor medio como pone de manifiesto un simple cálculo, por ejemplo con cinco puntos en total:

$$\lambda = \frac{2 \cdot 1 + 2 \cdot 1 + 2 \cdot 1 + 2 \cdot 1 + 0,025 \cdot 10}{8,025} \approx 1 + \frac{0,25}{8} \approx 1,03.$$

Téngase en cuenta además que el cálculo del factor de escala no ha de efectuarse necesariamente entre dos sistemas que se relacionan mediante una semejanza, sino que puede ser un valor inicial para el cálculo de una transformación más complicada, como una afinidad, una entre dos proyecciones cartográficas, etc., o simplemente un valor al que se le puede dar un cierto significado, como por ejemplo el factor de escala de la zona común de dos fotos sacadas desde puntos distintos. En todos estos casos no resulta seguro aplicar un peso a cada factor de escala individual de  $r_1 + r_2$ , pues puede estar un punto cerca del centro de gravedad en uno de los dos sistemas y alejado en el otro sin que haya ningún error, siquiera pequeño, en los puntos.

Lo dicho anteriormente se aplica por igual al giro, ya que los ángulos  $\alpha_1$  y  $\alpha_2$  se obtienen en última instancia a partir de cocientes entre las cantidades  $x$ ,  $y$  y  $r$ .

El peso puede ser  $r_1 r_2$ , pues si uno de los valores de  $r$  se hace cero el peso se hace cero en la misma proporción. Si aplicamos estos pesos

## 12 Transformación de semejanza plana

a los promedios obtenemos

$$\lambda = \frac{\sum r_1 r_2 \frac{r_2}{r_1}}{\sum r_1 r_2} = \frac{\sum r_2^2}{\sum r_1 r_2}, \quad \alpha = \frac{\sum r_1 r_2 (\theta_2 - \theta_1)}{\sum r_1 r_2};$$

que son las fórmulas ya obtenidas. En el caso del giro la definitiva y en el del factor de escala la primera de ellas, a falta de *simetrizarla*.

### Cómo promediar valores angulares

El promediar ángulos suele plantear una pequeña dificultad cuando los ángulos se sitúan en torno al origen de ángulos, de manera que unos tienen unos valores muy pequeños y otros valores próximos a  $400^\circ$ . Obviamente esto se puede evitar tomando los ángulos entre  $-200^\circ$  y  $200^\circ$ , pero entonces el problema aparece cuando los ángulos están en torno a  $200^\circ$ .

Para una persona no ofrece ninguna dificultad ver qué valores toman los ángulos y partir el círculo en uno u otro punto. El problema surge si se quiere programar una pequeña aplicación que lo calcule siempre bien. Se puede tomar uno de los ángulos a promediar como referencia y llevar todos los valores angulares a ese valor  $\pm 200^\circ$ . Esto fallará si el valor de referencia es erróneo por la razón que sea. A continuación indicamos un método que obtiene la solución correcta en todos los casos posibles:

Se divide el círculo en dos mitades, de manera que una de las divisiones coincida con la ruptura en la escala. Por ejemplo, de  $-200^\circ$  a  $0^\circ$  y de  $0^\circ$  a  $200^\circ$ , o bien de  $0^\circ$  a  $200^\circ$  y de  $200^\circ$  a  $400^\circ$ . Se promedian los ángulos de cada una de las dos mitades de acuerdo a la fórmula (12.9) o la que en su caso se emplee, recordando la suma de los pesos en ambos casos; esto es,  $p = (\sum r_1 r_2)_1$  y  $q = (\sum r_1 r_2)_2$ . Evidentemente si alguna de estas dos cantidades es cero, es decir, si no hay valores angulares en la mitad correspondiente, el resultado es el valor promedio de la otra mitad y no hay que hacer nada más.

En caso contrario miramos si para pasar de uno de los dos valores promedio, digamos  $\alpha_1$ , al otro, digamos  $\alpha_2$ , el camino es más corto a través de uno de los dos valores que limitan las dos mitades de la circunferencia o del otro y se promedia acorde con ello. Por ejemplo, si la circunferencia la hemos dividido por los valores  $0^\circ$  y  $\pm 200^\circ$  se mira

## 12 Transformación de semejanza plana

si  $\alpha_2 - \alpha_1$  es  $< 200$  ó  $> 200$ . En el primer caso se promedian  $\alpha_1$  y  $\alpha_2$  directamente:

$$\alpha = \frac{p\alpha_1 + q\alpha_2}{p + q}.$$

En el segundo caso primero se han de sumar  $400^\circ$  al valor de  $\alpha_1$ .

Si la diferencia fuese exactamente  $200^\circ$  es indistinto efectuar uno u otro promedio. Piénsese lo que significa que haya ángulos con diferencias de  $200^\circ$  para promediar.

### Giro: fórmula definitiva

El valor del ángulo de giro para una transformación de semejanza no se quiere para emplearlo él mismo sino para emplear sus razones trigonométricas  $\sin \alpha$  y  $\cos \alpha$ . Cuando se calcula el valor de un ángulo  $\alpha_1$  o  $\alpha_2$  se calcula primero una de esas dos razones o bien la tangente y a ella se aplica la función trigonométrica inversa correspondiente, esto es, arcsin, arc cos o arctan. Tras promediar todas las restas  $\alpha_2 - \alpha_1$  según la fórmula (12.9), a partir del valor obtenido  $\alpha$  calculamos  $\cos \alpha$  y  $\sin \alpha$ . Tal vez sería posible promediar directamente valores de  $\cos(\alpha_2 - \alpha_1)$  y  $\sin(\alpha_2 - \alpha_1)$  sin tener que pasar por las funciones trigonométricas inversas para después volver a pasar a sin y cos.

Esto no sólo es posible sino que elimina de raíz el problema abordado en la sección anterior.

El primer problema que surge ante este planteamiento es que, apliquemos las fórmulas que apliquemos, si promediamos por una parte valores de  $\cos(\alpha_2 - \alpha_1)$  y por otra valores de  $\sin(\alpha_2 - \alpha_1)$ , y puesto que los datos no son perfectos, los valores obtenidos no serán compatibles, es decir, que si a partir del coseno obtenido calculamos el valor del seno no nos va a dar lo mismo que el valor de  $\sin \alpha$  obtenido en el promedio y viceversa. Sean C y S los valores obtenidos. Para resolver esto no hay más que multiplicar ambas cantidades por una misma constante de manera que la suma de sus cuadrados pase a ser 1; es decir, *normalizarlos*:

$$\cos \alpha = \frac{C}{\sqrt{C^2 + S^2}}, \quad \sin \alpha = \frac{S}{\sqrt{C^2 + S^2}}. \quad (12.10)$$

## 12 Transformación de semejanza plana

Los valores  $\cos(\alpha_2 - \alpha_1)$  y  $\sin(\alpha_2 - \alpha_1)$  son

$$\cos(\alpha_2 - \alpha_1) = \frac{x_1x_2 + y_1y_2}{r_1r_2}, \quad \sin(\alpha_2 - \alpha_1) = \frac{x_1y_2 - y_1x_2}{r_1r_2},$$

valores que por las razones ya expuestas hay que promediar según los pesos  $r_1r_2$ , lo que da

$$C = \frac{\sum(x_1x_2 + y_1y_2)}{\sum r_1r_2}, \quad S = \frac{\sum(x_1y_2 - y_1x_2)}{\sum r_1r_2}.$$

Ahora bien, puesto que estos dos valores aún se han de normalizar según (12.10) podemos multiplicarlos por una constante cualquiera, y en particular podemos eliminar el  $\sum r_1r_2$  del denominador:

$$C = \sum(x_1x_2 + y_1y_2), \quad S = \sum(x_1y_2 - y_1x_2).$$

Incluyendo pesos para cada punto:

$$C = \sum p(x_1x_2 + y_1y_2), \quad S = \sum p(x_1y_2 - y_1x_2), \quad (12.11)$$

que junto con (12.10) dan la solución para el giro.

### Solución completa

$$(X_2 - G_2) = \lambda R(X_1 - G_1), \quad R = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (12.12)$$

Sin pesos:

$$G_2 = \frac{\sum P_2}{n}, \quad G_1 = \frac{\sum P_1}{n},$$
$$\lambda = \sqrt{\frac{\sum(x_2x_2 + y_2y_2)}{\sum(x_1x_1 + y_1y_1)}}, \quad (12.13)$$

$$C = \sum(x_1x_2 + y_1y_2), \quad S = \sum(x_1y_2 - y_1x_2),$$
$$\cos \alpha = \frac{C}{\sqrt{C^2 + S^2}}, \quad \sin \alpha = \frac{S}{\sqrt{C^2 + S^2}}.$$

## 12 Transformación de semejanza plana

Con pesos:

$$\begin{aligned}(\sigma_{\delta_2}^2)_i &= (\sigma_2^2 + \lambda_0^2 \sigma_1^2)_i, & p_i &= 1/(\sigma_{\delta_2}^2)_i \\ G_2 &= \frac{\sum p P_2}{\sum p}, & G_1 &= \frac{\sum p P_1}{\sum p}, \\ \lambda &= \sqrt{\frac{\sum p(x_2 x_2 + y_2 y_2)}{\sum p(x_1 x_1 + y_1 y_1)}}, & (12.14) \\ C &= \sum p(x_1 x_2 + y_1 y_2), & S &= \sum p(x_1 y_2 - y_1 x_2), \\ \cos \alpha &= \frac{C}{\sqrt{C^2 + S^2}}, & \sin \alpha &= \frac{S}{\sqrt{C^2 + S^2}}.\end{aligned}$$

La solución con pesos se aplica cuando las precisiones varían de un punto a otro. Por tanto no es necesario aplicar pesos si  $\sigma_1 \neq \sigma_2$  pero para todos los puntos los valores son iguales. La fórmula para el peso sigue siendo válida y se puede emplear también cuando  $\sigma_2 = 0$  ó  $\sigma_1 = 0$ .

La solución mínimo cuadrática rigurosa es bastante más complicada y la diferencia es insignificante (estará comprendida entre los valores anteriores de  $\lambda$  y  $\lambda'$  y de  $\alpha$  y  $\alpha'$ ), así que no hay ninguna razón para emplearla. Recordamos además que el estimador mínimo cuadrático se aplica muchas veces porque es el más fácil, no porque las observaciones se ajusten exactamente a las hipótesis. Además, en este caso, según como se planteen las hipótesis acerca de las precisiones a priori en uno y otro conjuntos de puntos la solución mínimo cuadrática rigurosa puede ser la expuesta en las fórmulas anteriores.

### Interpretación del parámetro $T$ y su precisión

Vamos ahora y en las siguientes secciones a obtener las precisiones de los parámetros de la transformación y del transformado de un punto según esos parámetros

En la fórmula de la transformación de (12.12) aparecen dos traslaciones, pero no son independientes ya que hay infinitas parejas equivalentes. Por lo tanto no se pueden tomar ambas como parámetros.

## 12 Transformación de semejanza plana

Podemos interpretar sus valores de dos maneras: Una de ellas consiste en tomar el punto  $G_1$  como fijo, lo que quiere decir que si se ha obtenido, por ejemplo,  $G_1 = (20, 10)$ , entonces en la fórmula de la transformación  $G_1$  lo definimos como el punto  $(20, 10)$ , no como el centro de gravedad ponderado de los puntos del conjunto 1. Entonces en el miembro de la izquierda tenemos  $X_2 - T_2$ , y  $T_2$  se obtiene mediante (12.4) y coincide con  $G_2$ , ya que  $\sum p(P_1 - (10, 20)) = (0, 0)$ . Si repitiésemos las observaciones entonces ya no se cumpliría  $\sum p(P_1 - (10, 20)) = (0, 0)$  y  $T_2$  sería distinto de  $G_2$ . La otra interpretación es la contraria, que supongamos como un punto fijo en la fórmula los valores obtenidos para  $G_2$ , y entonces el parámetro es  $T_1$ .

La segunda interpretación se ajusta mejor a la definición que adoptamos de los parámetros de la transformación, aunque no se cumple  $\sigma_{T_1} = \sigma_T$ . La razón está en el significado de  $T$ . Cuando empleamos este parámetro no existe  $T_2$ , y entonces la rotación y el escalado, que se aplican a los puntos  $P_{1'}$ , tienen por origen el origen de coordenadas del sistema 2, y puesto que son lo último que se aplica, los puntos resultantes  $P_{1''}$  quedan ya ajustados a los puntos  $P_2$ . Si los puntos del conjunto 2 están muy alejados del origen, entonces un giro o un escalado equivale en la zona de los puntos casi a una translación, por lo que antes de su aplicación los puntos  $P_{1'}$  están bastante separados de su posición final  $P_{1''}$ . Entonces, incluso aunque  $P_1$  y  $P_2$  casi coincidan, la translación  $T$  puede ser un valor muy grande simplemente para desplazar el conjunto  $P_1$  a la posición  $P_{1'}$  adecuada para que tras aplicar  $\lambda R$  los puntos vayan a su posición correcta (cf. el ejemplo correspondiente en el libro de ejercicios). Debido a este comportamiento, una variación pequeña en las observaciones que causase una variación también pequeña en los valores ajustados  $\lambda$  y  $\alpha$ , podría dar lugar a una variación grande de  $T$ . Esto significa que la precisión de  $T$  y de  $T$  es muy mala (valores de  $\sigma_{T_x}$  y  $\sigma_{T_y}$  grandes) y que la translación está muy correlada con el factor de escala y el ángulo de giro (valores de  $\sigma_{T_x, \lambda}$ , etc. también grandes). En consecuencia la precisión de  $T$  no interesa realmente, porque no tiene ningún significado útil. Todo esto no ocurre si tomamos  $T_2 = G_2$ , y por ello el concepto de precisión en la translación viene expresado mediante la precisión de  $T_1$ . Así, la transformación queda

$$(X_2 - G_2) = \lambda R(X_1 - T_1),$$

en donde  $G_2$  se debe interpretar como una constante dada. Sabemos

## 12 Transformación de semejanza plana

que para el conjunto de observaciones de que disponemos  $T_1 = G_1$ , pero no por ello deja de ser un parámetro; simplemente sabemos de antemano cuál será su resultado, al igual que ocurre con el valor ajustado para un conjunto de medidas de una misma magnitud, que sabemos que será la media aritmética. Podríamos tomar además  $G_1$  como un punto dado, y en lugar de  $(X_1 - T_1)$  tener  $(X_1 - G_1 - T_{\Delta_1})$ , y será  $T_{\Delta_1} = (0, 0)$ . Simplemente se trata de un desplazamiento en el parámetro  $T_1$ , así que ambas formas son equivalentes.

### Precisión de los parámetros de la transformación

La precisión de los parámetros se puede obtener a partir de las ecuaciones (12.13), pero es algo laborioso porque  $\lambda$  y  $\alpha$  dependen de los valores  $r$  y  $\theta$ , que a su vez dependen ambos de  $G_1$  y  $G_2$ , y por otra parte es fácil equivocarse en el cálculo de  $\sigma_{T_1}$ . Calcularemos entonces la precisión mediante la matriz N del método de mínimos cuadrados. Aunque no hayamos empleado ese método para calcular los parámetros, puesto que la diferencia respecto a la solución obtenida es totalmente despreciable, podemos calcular las precisiones como si hubiésemos tomado la solución mínimo cuadrática.

La solución mínimo cuadrática sin ningún conjunto de puntos fijo vimos que es equivalente al caso en el que los puntos del sistema 1 son fijos, sin más que sustituir  $\sigma_2$  por  $\sigma_{\delta_2} = \sqrt{(\sigma_2^2 + \lambda_0^2 \sigma_1^2)}$ . Las derivadas parciales son las de la página 127, con  $x_{2'}$  e  $y_{2'}$  en lugar de  $x_2$  e  $y_2$ . Las derivadas se evalúan en los valores aproximados. Tomaremos como valores aproximados de  $P_1$  los propios valores observados, y  $T_{10} = G_1$ ,  $\lambda_0 = \lambda$  y  $\alpha_0 = \alpha$ . Entonces  $P_{2'0} = \lambda R(P_1 - G_1)$ , por lo que el centro de gravedad de los puntos  $P_{2'0}$  es también  $(0, 0)$ , y además se cumple

$$P_{1'} = \frac{1}{\lambda} R^T P_{2'0};$$

es decir,

$$x_{1'} = \frac{1}{\lambda} (\cos \alpha x_{2'0} + \sin \alpha y_{2'0}), \quad y_{1'} = \frac{1}{\lambda} (-\sin \alpha x_{2'0} + \cos \alpha y_{2'0}).$$

Estas relaciones permiten expresar de manera más simple algunos elementos de la matriz N:

## 12 Transformación de semejanza plana

$$\begin{pmatrix} (\Sigma p) \lambda^2 & 0 & \lambda^2 \Sigma p x_{1'} & -\lambda \Sigma p x_{1'} \\ (\Sigma p) \lambda^2 & -\lambda^2 \Sigma p y_{1'} & -\lambda \Sigma p y_{1'} & \\ & \Sigma p r_{20}^2 & 0 & \\ & & \frac{1}{\lambda^2} \Sigma p r_{20}^2 & \end{pmatrix}.$$

Se cumple  $\Sigma p x_{1'} = 0$  y  $\Sigma p y_{1'} = 0$ . Entonces la matriz N es diagonal y su inversa es inmediata. Podemos sustituir sin error apreciable  $r_{20}$  por  $r_2$ . Si además tomamos  $1/p = \sigma_{\delta_2}^2$  entonces  $\sigma_0 = 1$ , y se obtiene

$$\sigma_{T_{1x}}^2 = \sigma_{T_{1y}}^2 = \frac{1}{\lambda^2 \Sigma p} = \frac{1}{\Sigma \frac{1}{\frac{1}{\lambda^2} \sigma_2^2 + \sigma_1^2}} = \frac{1}{\Sigma \frac{1}{\sigma_{\delta_1}^2}},$$

$$\sigma_{\alpha}^2 = \frac{\sigma_{\lambda}^2}{\lambda^2} = \frac{1}{\Sigma p r_2^2} = \frac{1}{\Sigma \frac{r_2^2}{\sigma_{\delta_2}^2}} = \frac{1}{\Sigma \frac{r_2^2}{\sigma_{\delta_1}^2}}.$$

Estas son las varianzas de los valores ajustados. Las que interesan son las de los valores reales, pero sabemos que en el estimador mínimo cuadrático ambos conjuntos coinciden. La precisión de  $\lambda$  interesa en proporción a sí misma, es decir,  $\sigma_{\lambda}/\lambda$ . Es lo que multiplicado por  $10^6$  se llama precisión en partes por millón. Finalmente, la traslación  $T_1$  se efectúa lo primero al pasar del sistema 1 al 2, por lo que su desviación típica está en unidades del sistema 1. En unidades del sistema 2 es la misma multiplicada por  $\lambda$ , y vemos que es igual a la desviación típica que tendría  $T_2$  si  $G_1$  fuese un punto fijo y  $T_2$  el parámetro. Entonces ambos valores son iguales, y en definitiva podemos decir que la precisión de cualquiera de ellos es la precisión de la traslación.

$$\sigma_{T_1} = \frac{1}{\sqrt{\Sigma \frac{1}{\sigma_{\delta_1}^2}}}, \quad \sigma_{T_2} = \frac{1}{\sqrt{\Sigma \frac{1}{\sigma_{\delta_2}^2}}},$$

$$\sigma_{\alpha} = \frac{\sigma_{\lambda}}{\lambda} = \frac{1}{\sqrt{\Sigma \frac{r_1^2}{\sigma_{\delta_1}^2}}} = \frac{1}{\sqrt{\Sigma \frac{r_2^2}{\sigma_{\delta_2}^2}}}. \quad (12.15)$$

Si además todos los puntos tienen el mismo valor de  $\sigma_1$  así como de  $\sigma_2$

## 12 Transformación de semejanza plana

entonces también son constantes  $\sigma_{\delta_1}$  y  $\sigma_{\delta_2}$ , y en ese caso

$$\begin{aligned}\sigma_{T_1} &= \frac{\sigma_{\delta_1}}{\sqrt{k}}, & \sigma_{T_2} &= \frac{\sigma_{\delta_2}}{\sqrt{k}}, \\ \sigma_{\alpha} &= \frac{\sigma_{\lambda}}{\lambda} = \frac{\sigma_{\delta_1}}{\sqrt{\sum r_1^2}} = \frac{\sigma_{\delta_2}}{\sqrt{\sum r_2^2}}.\end{aligned}\quad (12.16)$$

### Precisión del transformado de un punto

Las variables  $(T_{1x}, T_{1y}, \alpha, \lambda)$  son independientes, por ser la matriz  $N^{-1}$  diagonal. Entonces la precisión del transformado de un punto  $X_1$  de coordenadas libres de error es muy fácil de obtener:

$$\begin{aligned}\sigma_{x_2}^2 &= \left( \frac{\partial x_2}{\partial T_{1x}} \right)^2 \sigma_{T_{1x}}^2 + \dots + \left( \frac{\partial x_2}{\partial \lambda} \right)^2 \sigma_{\lambda}^2 = \\ &= \lambda^2 \sigma_{T_1}^2 + y_2'^2 \sigma_{\alpha}^2 + \frac{x_2'^2}{\lambda^2} \sigma_{\lambda}^2 = \sigma_{T_2}^2 + r_2^2 \sigma_{\alpha}^2,\end{aligned}$$

y entonces

$$\sigma_{x_2}^2 = \sigma_{y_2}^2 = \sigma_{T_2}^2 + r_2^2 \sigma_{\alpha}^2. \quad (12.17)$$

Si transformamos un punto  $X_2$  a  $X_1$  simplemente hay que cambiar 2 por 1 en la expresión.

Si el punto  $X_1$  no es perfecto entonces será igual al punto exacto más un cierto error, y el error en  $X_2$  será la suma del error al transformar el punto exacto más el error en  $X_1$  transformado. La varianza de la suma de dos variables independientes es la suma de las varianzas, así que

$$\Sigma_{X_2 X_2} = \begin{pmatrix} \sigma_{T_2}^2 + r_2^2 \sigma_{\alpha}^2 & 0 \\ 0 & \sigma_{T_2}^2 + r_2^2 \sigma_{\alpha}^2 \end{pmatrix} + \lambda^2 R \Sigma_{X_1 X_1} R^T. \quad (12.18)$$

Esto no es de aplicación a los puntos del conjunto  $P$ , ya que los parámetros de transformación se han obtenido a partir de estos puntos y entonces el error en aquellos no es independiente del error en los puntos. No obstante, si el número de puntos empleados para calcular la transformación es muy grande podemos admitir que sí son

## 12 Transformación de semejanza plana

independientes. Una vez obtenida la precisión de un punto  $P_1'$ , la de  $P_2$  se obtiene promediando las matrices de varianzas de ambos. De todos modos la precisión de los  $P$  no es algo que suele interesar.

Si  $\sigma_{x1} = \sigma_{y1} = \sigma_1$  y  $\sigma_{x1y1} = 0$  entonces

$$\sigma_{x2}^2 = \sigma_{y2}^2 = \sigma_2^2 + r_2^2 \sigma_\alpha^2 + \lambda^2 \sigma_1^2. \quad (12.19)$$

# 13 Otras transformaciones de coordenadas

## Semejanza espacial. Rotación

La transformación de semejanza en el espacio es en muchos aspectos una extensión de la semejanza en el plano. La fórmula escrita en forma matricial es igual, y las expresiones para  $G_1$ ,  $G_2$  y  $\lambda$  las mismas, así como  $\sigma_T$  y  $\sigma_\lambda$ . La diferencia radica en la rotación, que ahora queda definida mediante tres parámetros. Una rotación en el espacio que deje fijo el origen de coordenadas es un giro de ángulo  $\alpha$  respecto a una recta que pasa por el origen, que necesita dos parámetros para ser definida. Sin embargo una parametrización de la rotación que incluya los parámetros de la recta no es conveniente, porque cuando el giro es pequeño la posición de la recta queda determinada con una precisión muy mala. Existen muchas parametrizaciones posibles para una matriz de rotación  $3 \times 3$ . Todas ellas presentan problemas para la definición de alguno de los parámetros si la rotación esta próxima a una en particular (singularidad). Para solucionar este problema no tomaremos directamente los parámetros de la matriz  $R$  como parámetros a ajustar, sino que haremos  $R = R_\Delta R_0$ , en donde  $R_0$  es una rotación aproximada. La obtención de  $R_0$  puede hacerse de diversas maneras, aunque no vamos a extendernos aquí en el estudio de las matrices de rotación. Una manera muy sencilla consiste en obtener los nueve elementos de la matriz como si fuesen independientes, mediante un sistema de ecuaciones lineales, y tomar una matriz  $R_0$  que se aproxime a esos nueve valores, lo que es muy fácil. Ahora denotaremos por  $P_1'$  los puntos  $R_0(P_1 - T_1)$ . Los parámetros a calcular son los que

### 13 Otras transformaciones de coordenadas

definen la matriz  $R_\Delta$ . Basta tomar cualquier parametrización que no de problemas para rotaciones pequeñas.

En particular, si  $R_\Delta$  es muy pequeña, es decir, muy próxima a I (y si no es así podemos iterar), muchas parametrizaciones son el primer orden

$$\begin{pmatrix} 1 & -\kappa & \varphi \\ \kappa & 1 & -\omega \\ -\varphi & \omega & 1 \end{pmatrix}. \quad (13.1)$$

Esto corresponde, en primer orden, a pequeños giros de valor  $\omega$ ,  $\varphi$ ,  $\kappa$  respecto a los ejes  $x$ ,  $y$ ,  $z$  respectivamente (tratándose de una aproximación de primer orden y valores de los giros pequeños el orden de los giros es indiferente). La elección de los signos es consistente con la definición que tomamos para la matriz de rotación de dos dimensiones. Con esta parametrización tenemos

$$\begin{aligned} x_{1''} &= \lambda(x_{1'} - \kappa y_{1'} + \varphi z_{1'}), \\ y_{1''} &= \lambda(y_{1'} - \omega z_{1'} + \kappa x_{1'}), \\ z_{1''} &= \lambda(z_{1'} - \varphi x_{1'} + \omega y_{1'}). \end{aligned} \quad (13.2)$$

Sería deseable una división del segmento  $\delta_2$  en componentes tales que cada parámetro modificase una de ellas pero no las demás, como hicimos en dos dimensiones. Esto es imposible ya que en tres dimensiones las componentes deberían ser tres, y sin embargo tenemos cuatro parámetros. La variación en un punto  $P_{1''}$  debida a uno cualquiera de los parámetros de la rotación es tangencial, y en consecuencia se mantiene la independencia de  $\lambda$  respecto a la rotación, y su valor ajustado viene dado mediante la misma fórmula que en el caso bidimensional. El ajuste de los parámetros  $(\omega, \varphi, \kappa)$  debe hacerse en conjunto.

#### Cálculo de los parámetros de la rotación

Los valores ajustados de  $(\omega, \varphi, \kappa)$  serán los que minimicen  $\sum p \delta_{t_2}^2$ . En primer lugar tenemos que encontrar una expresión para  $\delta_{t_2}$ , que no es tan sencillo como en dos dimensiones. Calcularemos la longitud del segmento mediante las coordenadas de sus extremos. De esta manera estaremos empleando como  $\delta_{t_2}$  la cuerda en lugar del arco, lo

### 13 Otras transformaciones de coordenadas

que podemos hacer ya que en la aproximación  $\delta_2^2 \approx \delta_{r_2}^2 + \delta_{t_2}^2$  podemos entender  $\delta_{t_2}$  como el arco o la cuerda indistintamente.

Un extremo del segmento es  $P_{2'}$  y el otro  $\frac{r_2}{r_1}P_{1'} = P_{1''}$ . Podríamos tomar  $\delta_{t_2}$  desde  $P_{1''}$  en lugar de  $P_{2'}$ . Empleando este último, es decir, los radios  $r_2$ , llegaremos igual que en dos dimensiones a una fórmula en la que aparece  $r_2^2$ , que sustituiremos por  $r_2 r_1$ . Las coordenadas de  $P_{1''}$  son

$$\begin{aligned}x_{1''} &= \frac{r_2}{r_1}(x_{1'} - \kappa y_{1'} + \varphi z_{1'}), \\y_{1''} &= \frac{r_2}{r_1}(y_{1'} - \omega z_{1'} + \kappa x_{1'}), \\z_{1''} &= \frac{r_2}{r_1}(z_{1'} - \varphi x_{1'} + \omega y_{1'}).\end{aligned}$$

Hay que minimizar  $\delta_{t_2}^2$ , que es la suma de los cuadrados de sus componentes. La expresión de estas componentes en forma matricial es la siguiente:

$$\begin{pmatrix} x_{2'} - x_{1''} \\ y_{2'} - y_{1''} \\ z_{2'} - z_{1''} \end{pmatrix} = \begin{pmatrix} x_{2'} - \frac{r_2}{r_1}x_{1'} \\ y_{2'} - \frac{r_2}{r_1}y_{1'} \\ z_{2'} - \frac{r_2}{r_1}z_{1'} \end{pmatrix} - \frac{r_2}{r_1} \begin{pmatrix} 0 & z_{1'} & -y_{1'} \\ -z_{1'} & 0 & x_{1'} \\ y_{1'} & -x_{1'} & 0 \end{pmatrix} \begin{pmatrix} \omega \\ \varphi \\ \kappa \end{pmatrix}.$$

Y puesto que hay que minimizar la suma de los cuadrados (con los pesos  $p$ ), se trata de una estimación mínimo cuadrática. Tras operar obtenemos las matrices

$$\mathbf{N} = \begin{pmatrix} \sum p \frac{r_2^2}{r_1} (y_{1'}^2 + z_{1'}^2) & -\sum p \frac{r_2^2}{r_1} x_{1'} y_{1'} & -\sum p \frac{r_2^2}{r_1} z_{1'} x_{1'} \\ -\sum p \frac{r_2^2}{r_1} x_{1'} y_{1'} & \sum p \frac{r_2^2}{r_1} (z_{1'}^2 + x_{1'}^2) & -\sum p \frac{r_2^2}{r_1} y_{1'} z_{1'} \\ -\sum p \frac{r_2^2}{r_1} z_{1'} x_{1'} & -\sum p \frac{r_2^2}{r_1} y_{1'} z_{1'} & \sum p \frac{r_2^2}{r_1} (x_{1'}^2 + y_{1'}^2) \end{pmatrix},$$

$$\mathbf{A}^T \mathbf{L} = \begin{pmatrix} \sum p \frac{r_2}{r_1} (y_{1'} z_{2'} - z_{1'} y_{2'}) \\ \sum p \frac{r_2}{r_1} (z_{1'} x_{2'} - x_{1'} z_{2'}) \\ \sum p \frac{r_2}{r_1} (x_{1'} y_{2'} - y_{1'} x_{2'}) \end{pmatrix}.$$

En lo sucesivo, para mostrar matrices como estas de manera más clara escribiremos el signo de sumatorio y el peso fuera de la matriz, entendiendo que afectan ambos a todas las componentes de la matriz.

### 13 Otras transformaciones de coordenadas

Para que la expresión del resultado sea simétrica respecto a los conjuntos 1' y 2' sustituimos  $r_2^2$  por  $r_1 r_2$ , lo que en la matriz  $A^T L$  equivale a eliminar  $r_2/r_1$ , y en la matriz  $N$  a escribir  $r_2/r_1$  en lugar de  $r_2^2/r_1^2$ . En esta última es necesario además realizar una pequeña modificación:

$$\frac{r_2}{r_1}(x_1'^2 + y_1'^2) \approx (x_1'x_2' + y_1'y_2'), \quad \text{etc.},$$

$$\frac{r_2}{r_1}x_1'y_1' \approx x_1'y_2' \approx x_2'y_1' \approx \frac{1}{2}(x_1'y_2' + y_1'x_2'), \quad \text{etc.}$$

Estas aproximaciones son posibles porque la rotación  $R_\Delta$  es pequeña y las componentes  $x_1'$ ,  $y_1'$  y  $z_1'$  son casi proporcionales a  $x_2'$ ,  $y_2'$  y  $z_2'$ . Para estas últimas podríamos escribir  $x_2'$ , etc. pero mantenemos el superíndice por simetría y porque como veremos más adelante los puntos  $P_2'$  también pueden incluir una rotación. Con estas modificaciones las matrices quedan

$$N = \sum p \begin{pmatrix} (y_1'y_2' + z_1'z_2') & -\frac{1}{2}(x_1'y_2' + y_1'x_2') & \dots \\ -\frac{1}{2}(x_1'y_2' + y_1'x_2') & (z_1'z_2' + x_1'x_2') & \dots \\ -\frac{1}{2}(z_1'x_2' + x_1'z_2') & -\frac{1}{2}(y_1'z_2' + z_1'y_2') & \dots \end{pmatrix}, \quad (13.3 \text{ a})$$

$$A^T L = \sum p \begin{pmatrix} y_1'z_2' - z_1'y_2' \\ z_1'x_2' - x_1'z_2' \\ x_1'y_2' - y_1'x_2' \end{pmatrix}. \quad (13.3 \text{ b})$$

y la solución es  $N^{-1}A^T L$ .

Si los elementos fuera de la diagonal fuesen cero la estimación de cada parámetro  $\omega$ ,  $\varphi$ ,  $\kappa$ , que recordemos que equivalen en primer orden a pequeños giros, sería independiente e igual a una estimación bidimensional, empleando en cada caso el par de coordenadas correspondiente. Para estimar el ángulo  $\kappa$ , por ejemplo, dado que este parámetro no modifica las coordenadas  $z$  (cf. las expresiones (13.1) y (13.2)  $z$ ), para la estimación de ese ángulo aisladamente se emplearían únicamente coordenadas  $x, y$  como si la coordenada  $z$  no existiese, y sería:

$$S = \sum p(x_1'y_2' - y_1'x_2'), \quad C = \sum p(x_1'x_2' + y_1'y_2');$$

### 13 Otras transformaciones de coordenadas

$$\sin \kappa = \frac{S}{\sqrt{C^2 + S^2}}.$$

Puesto que el ángulo  $\kappa$  será muy pequeño  $\sqrt{C^2 + S^2} \approx C$ , y de ahí  $\sin \kappa \approx S/C = \tan \kappa$ , que es el valor que se obtiene si los elementos de fuera de la diagonal de  $N$  fuesen cero.

Se puede buscar una matriz  $R_1$  tal que  $P_{1'} = R_1(P_1 - G_1)$  cumpla  $\sum p x_{1'} y_{1'} = 0$ , etc. Entonces podemos hacer  $R_0 = R_2^{-1} R_1$  (de donde calculamos  $R_2$ ), y entonces

$$R_2(P_2 - G_2) = R_\Delta R_1(P_1 - G_1), \quad P_{2'} = R_\Delta P_{1'}.$$

La matriz  $R_\Delta$  no es la misma que antes. Seguirá siendo una rotación pequeña y podemos suponer de cara a la matriz  $N$  que se cumple también  $\sum p x_{2'} y_{2'} = 0$  y  $\sum p x_{1'} x_{2'} y_{1'} y_{2'} = 0$  (se cumpliría exactamente para  $R_\Delta = 0$  si las coordenadas no tuviesen error).

De todos modos esta descomposición tiene pocas ventajas desde un punto de vista práctico. Hasta cierto punto las matrices  $R_1$  y  $R_2$  son a las rotaciones lo que los puntos  $G_1$  y  $G_2$  a las traslaciones, pero sólo hasta cierto punto. Dependiendo de la forma del conjunto de puntos  $P$  no es raro que estén muy indefinidas; es decir, que matrices muy alejadas de  $R_1$  pueden seguir dando lugar a unos valores muy pequeños para los sumatorios  $\sum p x_{1'} y_{1'} = 0$ , etc., y en consecuencia una variación muy pequeña en las coordenadas de los puntos puede dar lugar a variaciones muy grandes de la matriz  $R_1$ . Por esa razón no se puede definir  $R_2$  de manera análoga a  $R_1$ , ya que la rotación  $R_\Delta$  podría no ser pequeña. Hay que obtener  $R_2$  a partir de  $R_1$  y una  $R_0$ , como hicimos. Por otra parte, la independencia de los valores  $\omega$ ,  $\phi$  y  $\kappa$  no simplifica el cálculo de las precisiones de los puntos transformados, porque posteriormente hay que aplicar  $R_2^T$  (ó  $R_1^T$  si transformamos en el otro sentido), y la matriz  $\sum x_2 x_2$  (ó  $\sum x_1 x_1$ ) deja de ser diagonal.

#### Construcción de una matriz $R_\Delta$ exacta

Una vez calculados los parámetros  $\omega$ ,  $\phi$  y  $\kappa$  conviene construir la matriz  $R_\Delta$  mediante una parametrización exacta de una matriz de rotación, para que la matriz  $R = R_\Delta R_0$  resultante sea exactamente una matriz de rotación. Esto es imprescindible si todavía se va a hacer alguna iteración más. La parametrización escogida ha de ser en

### 13 Otras transformaciones de coordenadas

primer orden igual a (13.1) para que la introducción de los parámetros  $(\omega, \varphi, \kappa)$  calculados tenga sentido. Mostramos aquí una que no incluye razones trigonométricas ni raíces cuadradas.

$$p = \frac{1}{1 + \frac{1}{4}(\omega^2 + \phi^2 + \kappa^2)},$$

$$R = \begin{pmatrix} 1 - \frac{1}{2}p(\phi^2 + \kappa^2) & -p\kappa + \frac{1}{2}p\omega\phi & p\phi + \frac{1}{2}p\kappa\omega \\ p\kappa + \frac{1}{2}p\omega\phi & 1 - \frac{1}{2}p(\kappa^2 + \omega^2) & -p\omega + \frac{1}{2}p\phi\kappa \\ -p\phi + \frac{1}{2}p\kappa\omega & p\omega + \frac{1}{2}p\phi\kappa & 1 - \frac{1}{2}p(\omega^2 + \phi^2) \end{pmatrix} =$$

$$\begin{pmatrix} 1 & -p\kappa & p\phi \\ p\kappa & 1 & -p\omega \\ -p\phi & p\omega & 1 \end{pmatrix} + \frac{p}{2} \begin{pmatrix} -(\phi^2 + \kappa^2) & \omega\phi & \kappa\omega \\ \omega\phi & -(\kappa^2 + \phi^2) & \phi\kappa \\ \kappa\omega & \phi\kappa & -(\omega^2 + \phi^2) \end{pmatrix}. \quad (13.4)$$

#### Descomposición planimétrica / altimétrica

La situación más común en el cálculo de una transformación de semejanza tridimensional es que el conjunto P sean puntos sobre la superficie terrestre, y los valores  $(z - z_g)$ , ya sea en el sistema 1 o en el sistema 2, sean muy pequeños en comparación con  $(x - x_g)$  e  $(y - y_g)$ . Los giros  $\omega$  y  $\varphi$  suelen ser por ello también muy pequeños, y entonces para obtener  $R_0$  sólo hay que calcular un giro en el plano mediante las coordenadas  $(x, y)$ . Si ese giro lo llamamos  $\kappa_0$ , entonces

$$R_0 = \begin{pmatrix} \cos \kappa_0 & -\operatorname{sen} \kappa_0 & 0 \\ \operatorname{sen} \kappa_0 & \cos \kappa_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

En estas circunstancias es habitual separar el cálculo de la transformación en una parte planimétrica y otra altimétrica. En primer lugar se calcula una transformación bidimensional mediante las coordenadas  $(x, y)$ , ignorando la coordenada  $z$ . El factor de escala que se obtiene se toma como el de la transformación tridimensional, lo que es posible ya que los valores de las distancias al centro de gravedad sin considerar la coordenada  $z$  difieren poco de las distancias correctas. Se obtienen los valores  $\sin \kappa_0$  y  $\cos \kappa_0$  como se vio y en función de

### 13 Otras transformaciones de coordenadas

ellos la matriz  $R_0$ . Posteriormente se calculan  $z_{g1}$  y  $z_{g2}$  y se aplica el giro a las coordenadas  $(P_1 - G_1)$ , obteniendo  $\hat{P}_{1'}$ . Dado lo pequeño de los valores  $(z - z_g)$ , los elementos  $n_{13} = n_{31}$  y  $n_{23} = n_{32}$  serán próximos a cero, lo que se aprovecha para calcular  $\omega$  y  $\varphi$  de manera independiente de  $\kappa$ . Entonces se obtiene el ángulo  $\kappa$  exclusivamente mediante las coordenadas  $(x, y)$ , de modo que su valor es ya  $\kappa_0$  (es decir, en  $R_\Delta$  se tiene  $\kappa = 0$ ). Finalmente se calculan  $\omega$  y  $\varphi$ .

Esta que hemos descrito es la aplicación correcta del método de separación de planimetría y altimetría, pero la realidad es que por desconocimiento del método se suelen emplear soluciones que no son más que la obtención de unos valores aproximados, en especial para  $\omega$  y  $\varphi$ , lo que no es en absoluto aceptable. Por otra parte, esta división del ajuste en planimétrico y altimétrico era útil cuando había que efectuar las operaciones a mano, pero hoy en día es parco el beneficio, ya que los únicos ahorros en cálculos son: I) Se obtienen los valores de  $r$  sin necesidad de calcular  $z^2$ , y II) Hay que invertir una matriz  $2 \times 2$  en lugar de una matriz  $3 \times 3$ .

#### Precisiones

La matriz de varianzas de los parámetros  $(\omega, \varphi, \kappa)$  de  $R_\Delta R_0$  (que son los de  $R_\Delta$ ) es  $\frac{1}{\lambda} N^{-1}$  y la precisión de un punto transformado es la siguiente:

$$\Sigma_{X_2 X_2} = \sigma_2^2 I + \frac{\sigma_\lambda^2}{\lambda^2} \begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix}_{2'} + A_2 N^{-1} A_1^T + \lambda^2 R \Sigma_{X_1 X_1} R^T;$$

$$A_2 = \begin{pmatrix} 0 & z_{2'} & -y_{2'} \\ -z_{2'} & 0 & x_{2'} \\ y_{2'} & -x_{2'} & 0 \end{pmatrix}, \quad A_1 = \frac{1}{\lambda} A_2 = \begin{pmatrix} 0 & z_{1'} & -y_{1'} \\ -z_{1'} & 0 & x_{1'} \\ y_{1'} & -x_{1'} & 0 \end{pmatrix}.$$

#### Afinidad

El siguiente tipo de transformaciones que vamos a estudiar son las transformaciones de afinidad. A la hora de calcular son con diferencia las más fáciles. La expresión es:

$$X_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} (X_1 - T). \quad (13.5)$$

### 13 Otras transformaciones de coordenadas

Por las mismas razones que en la semejanza es mejor calcularla como

$$(X_2 - G_2) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} (X_1 - G_1).$$

Es un problema lineal, y más aún, las ecuaciones de  $\delta_{x_2}$  son totalmente independientes de las de  $\delta_{y_2}$ , ya que

$$\delta_{x_2} = x_2 - (ax_1 + by_1),$$

$$\delta_{y_2} = y_2 - (cx_1 + dy_1).$$

Y se pueden realizar dos ajustes independientes. Al igual que en la semejanza no trabajamos con los residuos originales  $v_1$  y  $v_2$ , sino con  $\delta_2$ , la minimización de cuyos cuadrados es equivalente si tomamos los  $\sigma_{\delta_2}$  adecuados:  $\sigma_{\delta_2}^2 = \sigma_2^2 + \lambda^2 \sigma_1^2$ . El valor de  $\lambda$  lo podemos obtener mediante la fórmula vista para la semejanza.

Aunque el problema sea lineal conviene ajustar por el método general tomando valores aproximados, para que el resultado no sea muy sensible a pequeñas variaciones en los elementos de la matriz N. Se pueden calcular los valores aproximados, bien mediante tres puntos, lo que da una solución única para los parámetros, o bien calculando el giro y el factor de escala como si fuese una semejanza, y tomar como matriz aproximada  $\lambda R$ , siempre que la desviación de la afinidad respecto a una semejanza sea pequeña. Con parámetros aproximados,

$$\delta_{x_2} = x_2 - (x_1 a_0 + y_1 b_0) - (x_1 \Delta a + y_1 \Delta b),$$

$$\delta_{y_2} = y_2 - (x_1 c_0 + y_1 d_0) - (x_1 \Delta c + y_1 \Delta d).$$

La solución es

$$\begin{pmatrix} a \\ b \end{pmatrix} = N^{-1} \begin{pmatrix} \sum p x_1' L_x \\ \sum p y_1' L_x \end{pmatrix}, \quad \begin{pmatrix} c \\ d \end{pmatrix} = N^{-1} \begin{pmatrix} \sum p x_1' L_y \\ \sum p y_1' L_y \end{pmatrix}, \quad (13.6)$$

$$N = \sum p \begin{pmatrix} x_1'^2 & x_1' x_1' \\ x_1' x_1' & y_1'^2 \end{pmatrix}.$$

La extensión a tres dimensiones es inmediata.

### 13 Otras transformaciones de coordenadas

#### Descomposición de una afinidad

Una matriz de una afinidad (es decir, una matriz invertible cualquiera) se puede descomponer en giro, factor de escala y una componente afín, entendiendo por esta última una transformación lineal que no produce ningún «giro total» ni ningún «escalado total». En el plano es:

$$\lambda, R = \begin{pmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{pmatrix}, B = \frac{1}{\Delta} \begin{pmatrix} 1 + \alpha & \beta \\ \beta & 1 - \alpha \end{pmatrix};$$
$$\Delta = \sqrt{1 - \alpha^2 - \beta^2}. \quad (13.7)$$

Según el orden de aplicación de la afinidad respecto al giro hay dos posibilidades:

$$A = \lambda RB: \quad \lambda = |A|^{1/2}, \quad \psi = \arctan \frac{c-b}{a+d},$$
$$\alpha = \frac{a^2 - d^2 + c^2 - b^2}{(a+d)^2 + (c-b)^2}, \quad \beta = \frac{2(ab+dc)}{(a+d)^2 + (c-b)^2}; \quad (13.8)$$

y la otra,

$$A = \lambda BR: \quad \lambda = |A|^{1/2}, \quad \psi = \arctan \frac{c-b}{a+d},$$
$$\alpha = \frac{a^2 - d^2 + b^2 - c^2}{(a+d)^2 + (c-b)^2}, \quad \beta = \frac{2(ac+db)}{(a+d)^2 + (c-b)^2}. \quad (13.9)$$

#### Solución rigurosa en casos extremos

Si una afinidad tiene una componente afín considerable, entonces la minimización de  $\sum p\delta_2^2$  ya no es equivalente a la minimización de los cuadrados de los residuos. No obstante la diferencia no es significativa, salvo tal vez cuando el número de puntos es muy pequeño (4-6). Si queremos una solución rigurosa, entonces

$$\Sigma_{\delta_2\delta_2} = \Sigma_{X_2X_2} + A\Sigma_{X_1X_1}A^T. \quad (13.10)$$

### 13 Otras transformaciones de coordenadas

Las variables  $\delta_{x_2}$  y  $\delta_{y_2}$  ya no son independientes, y se trata entonces del caso general cuya solución se vio en el capítulo 8. La nueva expresión para el centro de gravedad es

$$\left(\sum \Sigma_{\delta_2 \delta_2}^{-1}\right)^{-1} \sum \Sigma_{\delta_2 \delta_2}^{-1} P_2, \quad (13.11)$$

y análogamente para el conjunto 1. Es fundamental que ambos centros de gravedad se calculen con pesos correspondientes; es decir, el mismo peso pero transformado al otro sistema. Por ello se emplea  $\Sigma_{\delta_2 \delta_2}$  para promediar los  $P_2$  y obtener  $G_2$  y  $\Sigma_{\delta_1 \delta_1}$  para promediar los  $P_1$  y obtener  $G_1$ . No se puede emplear  $\Sigma_{22}$  para el conjunto 2 y  $\Sigma_{11}$  para el conjunto 1.

Si por alguna razón queremos la solución rigurosa pero no deseamos trabajar con variables que no son independientes, se pueden emplear los residuos originales  $v_{x_1}$ ,  $v_{y_1}$ ,  $v_{x_2}$  y  $v_{y_2}$ , tomando entonces los parámetros  $(T_x, T_y, a, b, c, d, x_{11}, y_{11}, \dots, x_{n1}, y_{n1})$ . Sin embargo para conjuntos de puntos grandes es mucho menor la complicación de tener que emplear la matriz  $\Sigma_{\delta_2 \delta_2}$  que la de considerar tantos parámetros, que aumentaría enormemente los cálculos necesarios.\*

#### Transformaciones en general

Cualquier otra transformación que se pueda presentar quedará definida mediante un conjunto de parámetros  $(a_1, \dots, a_n)$ , en función de los cuales se escribe  $X_2$  a partir de  $X_1$ . Se puede siempre minimizar los vectores  $\delta_2$ , con la consiguiente reducción en el número de parámetros. Si la transformación conserva los ángulos (aunque sea sólo aproximadamente), tendrá en cada punto un factor de escala local,  $\lambda_X$ . Si estamos trabajando en dimensión  $d$ ,

$$\lambda_X = \sqrt[d]{|A_X|}, \quad A_X = \left(\frac{\partial x_{2i}}{\partial x_{1j}}\right)_X. \quad (13.12)$$

En donde  $x_i$  es la componente  $i$  del punto  $X$ ; es decir,  $x_1 = x$ ,  $x_2 = y$  y  $x_3 = z$ . Para cada punto  $X$  calculamos  $\sigma_{\delta_2}^2$  con el valor  $\lambda_X$ . Si la

---

\*Se pueden emplear métodos de inversión para la matriz  $N$  que tengan en cuenta su estructura particular, con muchos bloques de ceros, y efectuar sólo las operaciones necesarias, pero equivale a eliminar los parámetros coordenada obteniendo las matrices  $\Sigma_{\delta_2 \delta_2}$ ; es decir, lo mismo que emplearlas directamente.

### 13 Otras transformaciones de coordenadas

transformación no conserva los ángulos entonces hay que emplear (si queremos la solución rigurosa) la expresión más general

$$\Sigma_{\delta_2\delta_2} = \Sigma_{X_2X_2} + A\Sigma_{X_1X_1}A^T, \quad (13.13)$$

siendo A la misma matriz que antes.

El caso de conservación de ángulos (a efectos de varianzas, es decir, puede ser aproximado) es el de transformaciones que son iguales a una semejanza o un giro más pequeñas componentes de distorsión o error. El de no conservación de ángulos suele presentarse cuando hay que calcular los parámetros entre una imagen «poco o nada deformada» y otra «muy deformada», como por ejemplo entre un mapa y una imagen de satélite, o cuando interviene una imagen muy oblicua, por la razón que sea.



# 14 Cambio de los parámetros fijos, I: Puntual

Cada vez que como resultado de un ajuste se den precisiones para la posición de unos puntos, dichas precisiones dependen mucho de cómo se haya definido el sistema de coordenadas, pues las coordenadas de los puntos están referidas a ese sistema y por ello los valores de precisión que resultan son los de las coordenadas de cada punto respecto al sistema definido de tal forma.

Se puede cambiar la definición del sistema de coordenadas y con ello los valores de precisión que resultan. La manera más fácil de volver a calcular las precisiones es formar de nuevo la matriz  $N$ , eliminando de la matriz  $A$  las columnas de los nuevos parámetros fijos y añadiendo las que se habían eliminado, es decir, las de los parámetros que se habían tomado como fijos y que ya no lo son. Pero es posible que ya no dispongamos de la matriz  $A$ . Es posible incluso que la matriz  $A$  nunca se haya llegado a almacenar; nótese que para el cálculo del ajuste no es necesario guardar los valores de esta matriz, pues cada vez que se obtiene una fila pueden multiplicarse sus elementos entre sí y sumar los resultados a las posiciones respectivas de las matrices  $N$  y  $A^T L$ . En este capítulo se verá cómo es posible cambiar la definición del sistema y obtener las nuevas precisiones a partir únicamente de la matriz  $N^{-1}$ .

## **Precisión respecto a un punto**

Empezaremos por el caso más fácil: la obtención de la precisión de los puntos respecto a uno cualquiera en concreto. La precisión de

## 14 Cambio de los parámetros fijos, I: Puntual

un punto respecto a otro es la precisión en la diferencia de sus coordenadas. Así por ejemplo, la expresión que da las diferencias en las coordenadas de cinco puntos respecto al cuarto de ellos, escrita en forma matricial es:

$$\begin{pmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ X_3 \\ Y_3 \\ X_5 \\ Y_5 \end{pmatrix}_{\{4\}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ X_3 \\ Y_3 \\ X_4 \\ Y_4 \\ X_5 \\ Y_5 \end{pmatrix}.$$

El subíndice  $\{4\}$  significa que todas las coordenadas del vector son respecto al punto 4.

Escribimos esta relación en forma matricial:

$$X_{\{4\}} = AX.$$

Queremos obtener las regiones de confianza para las variables  $X_{\{4\}}$ . Se trata entonces de un problema normal de transmisión de varianzas y las regiones de confianza buscadas vendrán dadas por las varianzas

$$\Sigma_{X_{\{4\}}X_{\{4\}}} = A\Sigma_{XX}A^T.$$

Pero hay que tener cuidado en la formación de la matriz  $A$ . Sus columnas se deben corresponder exactamente con las filas de  $\Sigma_{XX}$ . Si alguno de los puntos se tomó como fijo no existirá en la matriz  $\Sigma_{XX}$  y habrá que eliminar sus correspondientes columnas de la matriz  $A$ . Si por ejemplo los cinco puntos proceden de un ajuste en el que se tomaron como fijos el punto 1 y la coordenada  $X$  del punto 2, esas tres columnas se eliminarán de la matriz  $A$ . En segundo lugar, las columnas de  $A$  deben estar en el mismo orden que las filas de  $\Sigma_{XX}$  y esta última no puede incluir parámetros que no aparezcan como columnas de  $A$ . Si existen tales parámetros deberá extraerse la submatriz de  $\Sigma_{XX}$  formada por las filas y columnas de los parámetros que

## 14 Cambio de los parámetros fijos, I: Puntual

interesan, o bien incluir columnas de ceros en la matriz A. Continuando con la suposición de que eran fijos el punto 1 y la coordenada  $X_2$  la matriz A será

$$A = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}.$$

Si además la matriz  $\Sigma_{XX}$  incluye las desorientaciones de los cinco puntos, deberá eliminarse sus filas y columnas de la matriz y emplear solamente la parte de coordenadas. Se muestra un ejemplo en el libro de ejercicios.

### Dos puntos fijos

Vamos ya a abordar el problema de un auténtico cambio en la definición del sistema de coordenadas. Puede suceder por ejemplo que en un ajuste se hayan tomado como fijas las coordenadas de un punto 1 y la coordenada  $X$  de otro punto 2, se haya obtenido la matriz  $\Sigma_{XX}$  (es decir,  $\sigma_0^2 N^{-1}$ ) y estemos interesados en la precisión de los puntos en un sistema que quede definido por las coordenadas y desorientación del punto 11, o por las coordenadas de 11 y la coordenada  $Y$  de 12, etc.

El problema es más complejo que la simple obtención de precisiones respecto a un punto porque al intervenir varios puntos ya no es suficiente con efectuar una resta. Lo resolveremos primero para la situación en la que el sistema queda definido mediante las coordenadas de dos puntos. Esto es lo que sucede cuando solamente se han medido ángulos. Para plantearlo vamos primero a analizar en detalle el caso precedente. Restar a todos los puntos las coordenadas de uno de ellos, de 4 por ejemplo, supone llevar toda la figura a un sistema en el que las coordenadas de 4 son fijas, en concreto  $(0,0)$ . Podrían haber sido otras cualesquiera, por ejemplo  $(1000, 500)$ . En ese caso la

## 14 Cambio de los parámetros fijos, I: Puntual

transformación a aplicar a todos los puntos sería

$$(X, Y)_{\{4\}} = (X, Y) - (X_4, Y_4) + (1000, 500).$$

Sumar el punto  $(1000, 500)$  es una traslación constante que no afecta a las precisiones.

Ahora lo que se pide es que dos puntos sean fijos, digamos A y B. Hay que plantear entonces la transformación que lleva los puntos A y B a dos puntos de coordenadas fijas, por ejemplo  $(0, 0)$  y  $(1000, 0)$ . Dicha transformación es una semejanza: traslación, giro y escala. Como queremos conservar la escala de la figura deberemos escoger como segundo punto uno que esté a la misma distancia de  $(0, 0)$  que B de A. Si ésta es, por ejemplo, 1037,64 m, se tomará para B el punto  $(0, 1037,64)$ , que podemos redondear a  $(0, 1040)$ .

No es ninguna barbaridad modificar 2 m una distancia para redondearla si atendemos al problema que nos ocupa. No queremos obtener las coordenadas de todos los puntos en un sistema definido por unas coordenadas fijas para A y B, sino simplemente obtener cuáles serían las precisiones en ese sistema. Desde luego que las precisiones no van a cambiar mucho por pasar de 1038 a 1040. Además hay una buena razón para tomar un valor distinto del exacto como en seguida se verá.

De esta manera la dirección A-B se lleva al eje X. Está bien si es eso lo que queremos, pero ahora lo que nos interesa es mantener las coordenadas de los puntos y solamente cambiar la definición de cuáles son fijos. Hay pues que restituir la dirección de esa línea, de lo contrario obtendremos las elipses giradas. Entonces llevaremos B a unas coordenadas B - A, que podemos redondear. Supongamos por ejemplo

$$B - A = (1013,25, 223,64) \approx (1010, 220).$$

La transformación que lleva A y B a los puntos  $(0, 0)$  y  $(1010, 220)$  respectivamente es

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} X - X_A \\ Y - Y_A \end{pmatrix} \quad (14.1)$$

para unos valores  $a$  y  $b$  que hay que calcular. No hay una traslación final porque queremos que el punto A se transforme en  $(0, 0)$ .

Por supuesto que por haber elegido el punto  $(1010, 220)$  como B - A ya sabemos que  $a$  y  $b$  van a ser 1 y 0, es decir, que no hay escalado

## 14 Cambio de los parámetros fijos, I: Puntual

ni giro. Pero la cuestión no es obtener  $a$  y  $b$  como números sino su expresión en función de las coordenadas de A y B. Hay que calcular la transformación que transforma A y B en  $(0,0)$  y  $(1010,220)$  fueren cuales fueren las coordenadas de esos dos puntos. Por eso hay que obtener los parámetros de la transformación en función de las «letras»  $X_A, Y_A, X_B$  e  $Y_B$ . Una vez obtenidos hay que continuar con las letras para calcular las derivadas parciales y obtener la matriz A. Sólo en ese momento se sustituyen por sus valores.

Por esta misma razón la transformación incluye en general un factor de escala, aunque sepamos siempre de antemano que va para nuestra elección de coordenadas va ser 1. Esto es algo en todo análogo a lo que ocurría en el cap. 12, cuando ya sabíamos que por elegir los centros de gravedad la traslación iba a ser cero. Ello no es óbice para que siga siendo un parámetro.

Ocurre lo mismo pero de forma más sencilla cuando queremos las precisiones respecto a un punto como en el caso anterior. En ese momento restamos a todos los puntos las coordenadas del punto 4. Si quisiésemos mantener las coordenadas en su sitio y solamente cambiar la definición del punto fijo habría que desplazar todo para volver a llevar el punto 4 a su sitio, por lo que al final no hay que restar nada. Pero sería un error interpretar que la ecuación consiste en restar  $(0,0)$ . Si las coordenadas de 4 son por ejemplo  $(118.21,2884.92)$ , la ecuación que cambia la definición del punto fijo y lo lleva a 4 y además mantiene los puntos en su sitio es

$$(X', Y') = (X, Y) - (X_4, Y_4) + (118,21, 2884,92).$$

Hay que reconocer que es ciertamente difícil al principio distinguir qué es lo que ha de tomarse como un parámetro y qué no es más que un simple número. Es por esto que se apuntó antes que es mejor redondear las coordenadas a las que se llevan los puntos, en este caso A y B, incluso aunque las tomemos iguales a sus propias coordenadas, para así ver claramente que se trata de un número y no de los parámetros «coordenada de A» o «coordenada de B». Ayuda también llevar uno de los puntos fijos siempre a  $(0,0)$ .

Como regla general recuerdese que *las coordenadas exactas de los puntos nunca aparecen como números en la fórmula*. Si hemos redondeado, claro.

El objetivo del desarrollo que estamos llevando a cabo es llegar a obtener los elementos de la matriz A. Éstos serán siempre núme-

## 14 Cambio de los parámetros fijos, I: Puntual

ros en cada problema concreto. Como ahora queremos obtener una expresión general para esos elementos que se pueda aplicar a cada caso concreto sin más que sustituir, al igual que hicimos en su momento cuando hallamos las derivadas de las ecuaciones de lectura y distancia respecto de las coordenadas de los puntos, cuando hayamos llegado a esas expresiones volveremos a escribir  $X_B - X_A$  e  $Y_B - Y_A$  en lugar de (1010, 220), pero no antes.

Hecha esta aclaración obtenemos  $a$  y  $b$  en función de  $X_A$ ,  $Y_A$ ,  $X_B$  e  $Y_B$ . Para ello sustituimos  $(X, Y)$  en (14.1) por  $(X_B, Y_B)$ . Según lo que acabamos de explicar éstos se mantienen como parámetros, letras, mientras que a la izquierda del igual están las coordenadas (1010, 220) que es el punto fijo a donde queremos llevar B:

$$\begin{pmatrix} 1010 \\ 220 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} X_B - X_A \\ Y_B - Y_A \end{pmatrix}.$$

O sea,

$$1010 = (X_B - X_A)a - (Y_B - Y_A)b,$$

$$220 = (Y_B - Y_A)a + (X_B - X_A)b.$$

Resolviendo para  $a$  y  $b$  se obtiene

$$1010(X_B - X_A) + 220(Y_B - Y_A) = ((X_B - X_A)^2 + (Y_B - Y_A)^2)a,$$

$$220(X_B - X_A) - 1010(Y_B - Y_A) = ((X_B - X_A)^2 + (Y_B - Y_A)^2)b;$$

$$a = \frac{1010(X_B - X_A) + 220(Y_B - Y_A)}{(X_B - X_A)^2 + (Y_B - Y_A)^2}, \quad (14.2)$$

$$b = \frac{220(X_B - X_A) - 1010(Y_B - Y_A)}{(X_B - X_A)^2 + (Y_B - Y_A)^2}.$$

Sea  $S^2 = (X_B - X_A)^2 + (Y_B - Y_A)^2$ . Sustituyendo estas expresiones en (14.1) obtenemos las expresiones para  $X'$  e  $Y'$  en función de los parámetros  $X_A$ ,  $Y_A$ ,  $X_B$  e  $Y_B$ :

$$X' = \frac{1010(X_B - X_A) + 220(Y_B - Y_A)}{S^2} X - \frac{220(X_B - X_A) - 1010(Y_B - Y_A)}{S^2} Y \quad (14.3)$$

#### 14 Cambio de los parámetros fijos, I: Puntual

$$Y' = \frac{220(X_B - X_A) - 1010(Y_B - Y_A)}{S^2} X + \frac{1010(X_B - X_A) + 220(Y_B - Y_A)}{S^2} Y \quad (14.4)$$

Las coordenadas  $X', Y'$  dependen por tanto de un total de 6 parámetros:  $X, Y, X_A, Y_A, X_B$  e  $Y_B$ . Hay que obtener las derivadas parciales respecto de cada uno de ellos. Puede ser un poco largo pero no tiene ninguna dificultad. En vez de derivar directamente estas expresiones lo haremos en dos pasos. Primero obtendremos las derivadas de  $a$  y  $b$  respecto de los cuatro parámetros de que dependen y después obtendremos las de  $X'$  e  $Y'$  a partir de esas y de (14.1). Las derivadas de  $a$  y  $b$  son

$$\frac{\partial a}{\partial X_A} = -\frac{1010}{S^2} + \frac{2(X_B - X_A)a}{S^2} = -\frac{\partial a}{\partial X_B},$$

$$\frac{\partial a}{\partial Y_A} = -\frac{220}{S^2} + \frac{2(Y_B - Y_A)a}{S^2} = -\frac{\partial a}{\partial Y_B},$$

$$\frac{\partial b}{\partial X_A} = -\frac{220}{S^2} + \frac{2(X_B - X_A)b}{S^2} = -\frac{\partial b}{\partial X_B},$$

$$\frac{\partial b}{\partial Y_A} = \frac{1010}{S^2} + \frac{2(Y_B - Y_A)b}{S^2} = -\frac{\partial b}{\partial Y_B}.$$

En las derivadas parciales, que en cada problema concreto serán números, ya podemos escribir  $(X_B - X_A)$  en lugar de 1010 e  $(Y_B - Y_A)$  en lugar de 220 para obtener unas expresiones generales aplicables a cualquier problema. También escribimos 1 por  $a$  y 0 por  $b$ . Con ello las expresiones para las derivadas se simplifican considerablemente:

$$\begin{aligned} \frac{\partial a}{\partial X_A} &= \frac{X_B - X_A}{S^2} = -\frac{\partial a}{\partial X_B}, & \frac{\partial a}{\partial Y_A} &= \frac{Y_B - Y_A}{S^2} = -\frac{\partial a}{\partial Y_B}, \\ \frac{\partial b}{\partial X_A} &= -\frac{Y_B - Y_A}{S^2} = -\frac{\partial b}{\partial X_B}, & \frac{\partial b}{\partial Y_A} &= \frac{X_B - X_A}{S^2} = -\frac{\partial b}{\partial Y_B}. \end{aligned} \quad (14.5)$$

Derivando ahora en (14.1) obtenemos para  $X'$

$$\frac{\partial X'}{\partial X} = a = 1, \quad \frac{\partial X'}{\partial Y} = -b = 0,$$

## 14 Cambio de los parámetros fijos, I: Puntual

$$\begin{aligned}\frac{\partial X'}{\partial X_A} &= \frac{\partial a}{\partial X_A}(X - X_A) - a - \frac{\partial b}{\partial X_A}(Y - Y_A) \\ &= \frac{X_B - X_A}{S^2}(X - X_A) + \frac{Y_B - Y_A}{S^2}(Y - Y_A) - 1,\end{aligned}$$

$$\begin{aligned}\frac{\partial X'}{\partial Y_A} &= \frac{\partial a}{\partial Y_A}(X - X_A) - \frac{\partial b}{\partial Y_A}(Y - Y_A) + b \\ &= \frac{Y_B - Y_A}{S^2}(X - X_A) - \frac{X_B - X_A}{S^2}(Y - Y_A),\end{aligned}$$

$$\begin{aligned}\frac{\partial X'}{\partial X_B} &= \frac{\partial a}{\partial X_B}(X - X_A) - \frac{\partial b}{\partial X_B}(Y - Y_A) \\ &= -\frac{X_B - X_A}{S^2}(X - X_A) - \frac{Y_B - Y_A}{S^2}(Y - Y_A),\end{aligned}$$

$$\begin{aligned}\frac{\partial X'}{\partial Y_B} &= \frac{\partial a}{\partial Y_B}(X - X_A) - \frac{\partial b}{\partial Y_B}(Y - Y_A) \\ &= -\frac{Y_B - Y_A}{S^2}(X - X_A) + \frac{X_B - X_A}{S^2}(Y - Y_A).\end{aligned}$$

Si  $(X, Y)$  no es el punto A ni el punto B el problema es totalmente simétrico respecto a A y B, por lo que las derivadas de  $X'$  respecto de  $X_B$  e  $Y_B$  han de ser necesariamente las mismas que respecto de  $X_A$  e  $Y_A$  cambiando A por B. Esto lo podemos comprobar operando en las expresiones precedentes. Así, se llega a las siguientes expresiones, en las que para simplificar su escritura

$$[1] \text{ representa } \frac{X_B - X_A}{S^2} \quad y \quad [2] \text{ representa } \frac{Y_B - Y_A}{S^2}.$$

$$\begin{aligned}\frac{\partial X'}{\partial X_A} &= [1](X - X_A) + [2](Y - Y_A) - 1, \\ \frac{\partial X'}{\partial Y_A} &= [2](X - X_A) - [1](Y - Y_A), \\ \frac{\partial X'}{\partial X_B} &= -[1](X - X_B) - [2](Y - Y_B) - 1, \\ \frac{\partial X'}{\partial Y_B} &= -[2](X - X_B) + [1](Y - Y_B); \end{aligned} \tag{14.6}$$

## 14 Cambio de los parámetros fijos, I: Puntual

y las correspondientes a  $Y'$ :

$$\begin{aligned}\frac{\partial Y'}{\partial X_A} &= -[2](X - X_A) + [1](Y - Y_A), \\ \frac{\partial Y'}{\partial Y_A} &= [1](X - X_A) + [2](Y - Y_A) - 1, \\ \frac{\partial Y'}{\partial X_B} &= [2](X - X_B) - [1](Y - Y_B), \\ \frac{\partial Y'}{\partial Y_B} &= -[1](X - X_B) - [2](Y - Y_B) - 1.\end{aligned}\tag{14.7}$$

Las fórmulas precedentes se simplifican mucho en cada problema concreto porque en ellas lo único que es variable es  $X$  e  $Y$ . Así por ejemplo, para los valores que estamos empleando de  $B - A \approx (1013, 224)$  quedan reducidas a

$$\begin{aligned}\frac{\partial X'}{\partial X_A} &= 0,00094(X - X_A) + 0,00021(Y - Y_A) - 1, \\ \frac{\partial X'}{\partial Y_A} &= 0,00021(X - X_A) - 0,00094(Y - Y_A), \\ \frac{\partial X'}{\partial X_B} &= -0,00094(X - X_B) - 0,00021(Y - Y_B) - 1, \\ \frac{\partial X'}{\partial Y_B} &= -0,00021(X - X_B) + 0,00094(Y - Y_B); \\ \frac{\partial Y'}{\partial X_A} &= -0,00021(X - X_A) + 0,00094(Y - Y_A), \\ \frac{\partial Y'}{\partial Y_A} &= 0,00094(X - X_A) + 0,00021(Y - Y_A) - 1, \\ \frac{\partial Y'}{\partial X_B} &= 0,00021(X - X_B) - 0,00094(Y - Y_B), \\ \frac{\partial Y'}{\partial Y_B} &= -0,00094(X - X_B) - 0,00021(Y - Y_B) - 1,\end{aligned}$$

en donde las coordenadas de  $A$  y  $B$  también serán constantes, a las que hay que añadir

$$\frac{\partial X'}{\partial X} = 1, \quad \frac{\partial X'}{\partial Y} = 0, \quad \frac{\partial Y'}{\partial X} = 0, \quad \frac{\partial Y'}{\partial Y} = 1.$$

Todas estas derivadas se sitúan en sus posiciones respectivas de la matriz  $A$ . Sea por ejemplo

$$(X_A, Y_A) = (30\ 897,52, 16\ 021,82),$$

## 14 Cambio de los parámetros fijos, I: Puntual

$$(X_B, Y_B) = (31\,910,77, 16\,245,46), \\ (X_P, Y_P) = (31\,755,48, 15\,305,74).$$

Se obtienen las siguientes diferencias:

$$P - A \approx (858, -716), \quad P - B \approx (-155, -940).$$

Los ocho valores de las derivadas son, en orden,

$$-0,34, \quad 0,85, \quad -0,66, \quad -0,85; \quad -0,85, \quad -0,34, \quad 0,85, \quad -0,66.$$

Las filas correspondientes a  $X_P$  e  $Y_P$  serían entonces

$$\begin{pmatrix} \dots & -0,34 & 0,85 & \dots & -0,66 & -0,85 & \dots & 1 & 0 & \dots \\ \dots & -0,85 & -0,34 & \dots & 0,85 & -0,66 & \dots & 0 & 1 & \dots \end{pmatrix}.$$

Los ... son las posiciones correspondientes a las columnas de los demás puntos, que valen todas 0. El orden de las columnas puede ser distinto; tiene que ser el mismo que el de filas de la matriz  $N^{-1}$ . Si alguno de los 6 parámetros no apareciese en la matriz  $N^{-1}$  su columna tampoco se incluiría en A.

### Un punto y orientación

Veamos ahora cómo es la matriz A si el sistema queda definido por tres parámetros, esto es, cuando se permite trasladar y girar pero no escalar. La transformación es

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} X - X_A \\ Y - Y_A \end{pmatrix}, \quad (14.8)$$

siendo A el punto que se fija en coordenadas.

Consideraremos dos casos según los parámetros que se tomen como fijos para definir el sistema: I) las dos coordenadas de un punto y una desorientación o II) las dos coordenadas de un punto y una dirección. El caso de las dos coordenadas de un punto y una coordenada de otro punto veremos que conduce de manera lógica al caso II.

En el primer caso el giro  $\alpha$  queda definido por la desorientación del punto, que puede ser el punto A u otro distinto. Lo llamaremos B

## 14 Cambio de los parámetros fijos, I: Puntual

en cualquier caso. Al igual que antes escribiremos unos números concretos en lugar de letras para los parámetros que se fijan para así facilitar los desarrollos. Sea entonces  $100^\circ$  el valor que escribiremos para representar  $\Sigma'_B$ . El giro es

$$\alpha = \Sigma_B - 100.$$

Como vamos a derivar habría que escribir esta ecuación en radianes, pero no importa porque en ambos casos la derivada de  $\alpha$  respecto de  $\Sigma_B$  es 1. Entonces

$$\begin{aligned}\frac{\partial \cos \alpha}{\partial \Sigma_B} &= \frac{\partial \cos \alpha}{\partial \alpha} \frac{\partial \alpha}{\partial \Sigma_B} = -\sin \alpha, \\ \frac{\partial \sin \alpha}{\partial \Sigma_B} &= \frac{\partial \sin \alpha}{\partial \alpha} \frac{\partial \alpha}{\partial \Sigma_B} = \cos \alpha.\end{aligned}$$

Y como  $\alpha$  va a ser 0 estos valores serán 0 y 1 respectivamente.

Los parámetros  $X_A$  e  $Y_A$  no intervienen en la matriz de giro, y como esta va a ser la matriz identidad se obtienen las siguientes derivadas

$$\frac{\partial X'}{\partial X_A} = -1, \quad \frac{\partial X'}{\partial Y_A} = 0, \quad \frac{\partial Y'}{\partial X_A} = 0, \quad \frac{\partial Y'}{\partial Y_A} = -1.$$

Sea cual sea la transformación, como escogeremos los parámetros de manera que los puntos se queden en su sitio, las siguientes derivadas siempre valen lo mismo:

$$\frac{\partial X'}{\partial X} = 1, \quad \frac{\partial X'}{\partial Y} = 0, \quad \frac{\partial Y'}{\partial X} = 0, \quad \frac{\partial Y'}{\partial Y} = 1.$$

Sólo falta por obtener las derivadas respecto de  $\Sigma_B$ :

$$\begin{aligned}\frac{\partial X'}{\partial \Sigma_B} &= \frac{\partial \cos \alpha}{\partial \Sigma_B} (X - X_A) - \frac{\partial \sin \alpha}{\partial \Sigma_B} (Y - Y_A) = \\ &= -\sin \alpha (X - X_A) - \cos \alpha (Y - Y_A) = -(Y - Y_A), \\ \frac{\partial Y'}{\partial \Sigma_B} &= \frac{\partial \sin \alpha}{\partial \Sigma_B} (X - X_A) + \frac{\partial \cos \alpha}{\partial \Sigma_B} (Y - Y_A) = \\ &= \cos \alpha (X - X_A) - \sin \alpha (Y - Y_A) = (X - X_A),\end{aligned}\tag{14.9}$$

## 14 Cambio de los parámetros fijos, I: Puntual

y ya tenemos todas las derivadas que hacen falta.

Si empleamos los mismos puntos P y A que en el ejemplo anterior las filas de la matriz A para  $X_P, Y_P$  son

$$\begin{pmatrix} \dots & -1 & 0 & \dots & 1 & 0 & \dots & 716 \\ \dots & 0 & -1 & \dots & 0 & 1 & \dots & 858 \end{pmatrix}.$$

La última columna es la correspondiente a  $\Sigma_B$ . Como aparece en la matriz A ahora es necesario incluir en la matriz  $N^{-1}$  la fila y columna correspondientes a  $\Sigma_B$ , al contrario que en los casos anteriores en los que no intervenía ninguna desorientación.

Si el sistema se define tomando como fijas las coordenadas de A y una coordenada de B el valor de  $\alpha$  en (14.8) queda definido por las coordenadas de esos dos puntos. Si la dirección A-B es aproximadamente la del eje X sabemos que para que el giro quede bien definido es necesario tomar como fija la Y de B y no la X, y al contrario si la dirección A-B es aproximadamente la del eje Y. En general la solución óptima para dar orientación al conjunto no consiste en tomar una u otra coordenada fijas, sino en mantener la relación X/Y en un valor fijo, que como queremos que los puntos no se muevan tomamos igual al que tienen. Sea por ejemplo A-B =  $60^\circ$  respecto al eje X, la orientación queda fijada entonces obligando a A-B =  $60^\circ$ . De esta manera no hay que distinguir dos casos según la coordenada de B que se fije sea la X o la Y, además de proporcionar elipses más pequeñas.

Si A-B =  $60^\circ$  se tiene que cumplir

$$\frac{Y'_B - Y'_A}{X'_B - X'_A} = \tan 60^\circ \approx 1,38.$$

$(X'_B - X'_A, Y'_B - Y'_A)$ , o lo que es lo mismo  $(X'_B, Y'_B)$ , se obtienen aplicando (14.8) al punto B:

$$\begin{aligned} X'_B &= \cos \alpha (X_B - X_A) - \sin \alpha (Y_B - Y_A), \\ Y'_B &= \sin \alpha (X_B - X_A) + \cos \alpha (Y_B - Y_A). \end{aligned}$$

sustituyendo en la ecuación anterior,

$$\frac{(X_B - X_A) \cos \alpha - (Y_B - Y_A) \sin \alpha}{(X_B - X_A) \sin \alpha + (Y_B - Y_A) \cos \alpha} = 1,38.$$

## 14 Cambio de los parámetros fijos, I: Puntual

En esta expresión se puede despejar  $\alpha$  pasando el denominador al miembro derecho, dividiendo luego todo por  $\cos \alpha$  y despejando  $\tan \alpha$ . De ahí se puede derivar  $\alpha$  respecto de cada una de las variables de que depende. Una vez obtenidas las derivadas se introduce  $\frac{Y_B - Y_A}{X_B - X_A}$  en lugar de 1,38 y se simplifica. En lugar de hacerlo así obtendremos gráficamente las derivadas, que es más fácil.

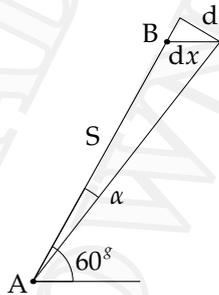


Fig. 14.1

Esta figura muestra el esquema. La dirección de  $60^\circ$  respecto al eje X es a la que vamos a obligar la línea A'-B'. Como hemos tomado el ángulo de manera que coincida con el de la línea A-B, ambas líneas coinciden. Si ahora B se desplaza una pequeña cantidad dx, para que la línea A'-B' se sitúe según el valor fijado de  $60^\circ$  hay que aplicar un pequeño ángulo  $\alpha$ :

$$\alpha = d\alpha = \frac{dt}{S} = dX_B \frac{Y_B - Y_A}{S^2},$$

de donde

$$\frac{\partial \alpha}{\partial X_B} = \frac{Y_B - Y_A}{S^2}.$$

Los cuatro valores que necesitamos son

$$\begin{aligned} \frac{\partial \alpha}{\partial X_B} &= \frac{Y_B - Y_A}{S^2} = -\frac{\partial \alpha}{\partial X_A}, \\ \frac{\partial \alpha}{\partial Y_B} &= -\frac{X_B - X_A}{S^2} = -\frac{\partial \alpha}{\partial Y_A}. \end{aligned}$$

## 14 Cambio de los parámetros fijos, I: Puntual

Si se aplica un pequeño giro  $d\alpha$  a un punto respecto al origen sabemos que se cumple

$$dX = -Yd\alpha, \quad dY = Xd\alpha.$$

Como en este caso el origen es el punto A,

$$\frac{\partial X'}{\partial \alpha} = -(Y - Y_A), \quad \frac{\partial Y'}{\partial \alpha} = X - X_A,$$

y finalmente obtenemos las derivadas de  $X'$ :

$$[1] \text{ representa } \frac{X_B - X_A}{S^2} \quad y \quad [2] \text{ representa } \frac{Y_B - Y_A}{S^2};$$

$$\begin{aligned} \frac{\partial X'}{\partial X_A} &= -1 + [2](Y - Y_A), & \frac{\partial X'}{\partial Y_A} &= -[1](Y - Y_A), \\ \frac{\partial X'}{\partial X_B} &= -[2](Y - Y_A), & \frac{\partial X'}{\partial Y_B} &= [1](Y - Y_A); \end{aligned} \quad (14.10)$$

y las de  $Y'$ :

$$\begin{aligned} \frac{\partial Y'}{\partial X_A} &= -[2](X - X_A), & \frac{\partial Y'}{\partial Y_A} &= -1 + [1](X - X_A), \\ \frac{\partial Y'}{\partial X_B} &= [2](X - X_A), & \frac{\partial Y'}{\partial Y_B} &= -[1](X - X_A). \end{aligned} \quad (14.11)$$

Siguiendo con los mismos puntos A y B de ejemplo se tiene

$$\begin{aligned} \frac{\partial X'}{\partial X_A} &= 0,00021(Y - Y_A) - 1, & \frac{\partial X'}{\partial Y_A} &= -0,00094(Y - Y_A), \\ \frac{\partial X'}{\partial X_B} &= -0,00021(Y - Y_A), & \frac{\partial X'}{\partial Y_B} &= 0,00094(Y - Y_A); \\ \frac{\partial Y'}{\partial X_A} &= -0,00021(X - X_A), & \frac{\partial Y'}{\partial Y_A} &= 0,00094(X - X_A) - 1, \\ \frac{\partial Y'}{\partial X_B} &= 0,00021(X - X_A), & \frac{\partial Y'}{\partial Y_B} &= -0,00094(X - X_A). \end{aligned}$$

Para el punto P de ejemplo las filas de la matriz A son:

$$\begin{pmatrix} \dots & -1,15 & 0,67 & \dots & 0,15 & -0,67 & \dots & 1 & 0 & \dots \\ \dots & -0,18 & -1,81 & \dots & 0,18 & -0,81 & \dots & 0 & 1 & \dots \end{pmatrix}.$$

## 14 Cambio de los parámetros fijos, I: Puntual

En las filas de la matriz A relativas al punto B, en las posiciones correspondientes a  $X_B$  e  $Y_B$ , tenemos por una parte las derivadas que proporcionan las fórmulas (14.10) y (14.11) y por otra los valores +1 por ser las columnas del propio punto para el que estamos derivando. Ambos valores deben sumarse. Así, para el punto B las derivadas respecto de sus propias coordenadas son las siguientes. Obsérvese que son los valores opuestos a las derivadas respecto de A:

$$\begin{aligned}
 \frac{\partial X'_B}{\partial X_B} &= 1 - [2](Y_B - Y_A) = \frac{(X_B - X_A)^2}{S^2}, \\
 \frac{\partial X'_B}{\partial Y_B} &= [1](X_B - X_A) = \frac{(X_B - X_A)(Y_B - Y_A)}{S^2}, \\
 \frac{\partial Y'_B}{\partial X_B} &= [2](X_B - X_A) = \frac{(X_B - X_A)(Y_B - Y_A)}{S^2}, \\
 \frac{\partial Y'_B}{\partial Y_B} &= 1 - [1](X_B - X_A) = \frac{(Y_B - Y_A)^2}{S^2}.
 \end{aligned} \tag{14.12}$$

Para los puntos de ejemplo los valores que resultan son

$$\begin{pmatrix} \dots & -0,95 & -0,21 & \dots & 0,95 & 0,21 & \dots \\ \dots & -0,21 & -0,05 & \dots & 0,21 & 0,05 & \dots \end{pmatrix}.$$

Las dos primeras columnas mostradas son las derivadas respecto de  $X_A$  e  $Y_A$ , las dos siguientes respecto de  $X_B$  e  $Y_B$ .

Para definir orientación mediante una dirección fija hemos tomado como uno de los extremos de la línea definitoria el propio punto fijo. Puede tomarse de otra manera, de modo que si la orientación queda definida obligando la dirección de la línea A-B, la posición lo sea fijando las coordenadas de otro punto C. Pueden obtenerse las derivadas para esta caso fácilmente a partir de las vistas: el -1 que aparece en dos de las derivadas respecto de las coordenadas de A se quita de ahí y pasa a la derivada análoga respecto de C, que es simplemente -1, mientras que las derivadas para el punto A respecto de sus propias coordenadas son iguales que las de (14.12) intercambiando las letras A y B. Pero si lo que se busca con una elección de este tipo es un punto fijo óptimo es mejor tomar el punto A como se hizo aquí y después aplicar el desplazamiento óptimo para definir la posición, que se explica más abajo.

## 14 Cambio de los parámetros fijos, I: Puntual

### Posición libre

Hemos estudiado los casos en los que el sistema de coordenadas queda determinado mediante las coordenadas de dos puntos, es decir, en el que la posición, orientación y escala son libres, y el caso en el que posición y orientación son libres pero no la escala. El último caso que falta por analizar es aquél en el que solamente la posición es libre. Entonces la transformación se reduce a una traslación, que queda definida fijando las coordenadas de un punto:

$$X' = X - X_A,$$

$$Y' = Y - Y_A,$$

más la traslación constante que devolviese A a su sitio, que estamos ignorando siempre porque no afecta a las precisiones. Se trata entonces del caso ya visto al principio del capítulo de obtener las precisiones respecto a un punto.

### Elección de los grados de libertad a aplicar

De las tres posibilidades vistas hay que decidir en cada situación cuál es correcto aplicar. Si no se piensa mucho parecería que el caso a aplicar viene determinado por los grados de libertad de la red. Así, si la red tiene orientación y escala absolutas, se aplicará la solución que solamente incluye una traslación: obtener la precisión respecto a un punto. Pero este razonamiento nos llevaría por ejemplo a concluir que en una red en coordenadas absolutas no se puede aplicar ningún cambio de los parámetros fijos, ya que no los hay. Es decir, ¿que no podemos ni siquiera calcular la precisión de todos los puntos en relación a un punto cualquiera? Obviamente esto se puede hacer siempre.

Para saber qué transformación conviene aplicar hay que ver qué se va a hacer con las elipses calculadas; es decir, para qué se quieren las precisiones respecto a tal o cual punto. Por ejemplo, si de una red grande en coordenadas absolutas vamos a emplear tres puntos próximos para desde ellos dar coordenadas a un conjunto de puntos de la zona, y esas coordenadas pueden estar en un sistema cualquiera pero en el que los metros sean metros (es decir, que la escala no es libre)

## 14 *Cambio de los parámetros fijos, I: Puntual*

entonces nos interesa definir el sistema de coordenadas de manera que las elipses de precisión de esos tres puntos sean lo más pequeñas posible. Para ello podemos tomar uno de ellos como fijo y definir una orientación, de acuerdo a las posibles libertades—posición y orientación— del sistema local. No podemos sin embargo tomar dos de los tres puntos como fijos porque eso daría lugar a que las elipses de precisión de ambos se reducen a cero, es decir, que estaríamos diciendo que las coordenadas de los dos puntos son perfectas, lo que es imposible si la escala del trabajo no es arbitraria. Si la orientación del sistema local tampoco fuese libre y estuviese determinada por la de la red, entonces tampoco se puede tomar como perfecta la desorientación de uno de los puntos o la dirección de la línea entre dos de ellos, y lo más que se puede hacer es tomar las coordenadas de uno de ellos como perfectas.

Así pues, para ver qué redefinición de parámetros fijos es posible aplicar no hay que mirar de dónde vienen las coordenadas, sino a dónde van. Interesa siempre aplicar la transformación con más grados de libertad posibles porque cuantos más parámetros se puedan tomar como perfectos más pequeñas serán las elipses de error.



# 15 Cambio de los parámetros fijos, II: Global

## Introducción

Los cambios en la definición del sistema de coordenadas vistos hasta ahora se reducen a cambiar el punto fijo, la dirección fija y la distancia fija. Según escojamos uno u otro punto las elipses de error se harán más pequeñas para los puntos próximos al punto escogido y más grandes cuanto más nos alejemos de él. Tal vez sería posible aplicar una redefinición que se ajustase a los grados de libertad pero de alguna manera tuviese en cuenta todos los puntos por igual y diese lugar a las elipses más pequeñas posibles en general. Esto es posible, pero antes de ponerse a aplicarlo hay que analizar si interesa o no y cómo se ha de aplicar.

Cuando se lleva la definición del sistema de coordenadas a unos u otros puntos es porque emplearemos esos puntos para algo y queremos que las elipses de precisión en la zona sean lo más pequeñas posible, así que son mejores las soluciones puntuales vistas en el capítulo anterior que las globales que se verán en este, que tienen en cuenta todos los puntos por igual sin que ninguna zona resulte favorecida. Sin embargo se pueden aplicar las soluciones globales que se desarrollan en este capítulo sin más que tomar como "global" el subconjunto de puntos de la red que interesa, con lo que se obtiene una definición local del sistema de coordenadas pero en la que ninguno de los puntos de interés resulta favorecido sobre otro.

En cuanto a la red en su conjunto, a veces es preferible tener unas elipses un poco más grandes que las mejores posibles pero saber cla-

## 15 Cambio de los parámetros fijos, II: Global

ramente cómo está definido el sistema de coordenadas. Normalmente sin embargo se prefiere homogeneidad de precisiones por toda la red y que estas sean lo más pequeñas posible.

Por último, para no tener que eliminar de las matrices  $A$  que se obtengan las columnas correspondientes a los parámetros que eran fijos y por tanto no aparecen en la matriz  $\Sigma$ , lo que evitará excepciones en las expresiones generales de las matrices  $A$ , supondremos que la matriz  $\Sigma$  está completa, lo que se consigue sin más que intercalar filas y columnas de ceros correspondientes a los parámetros fijos.

### Traslación promedio

Hasta ahora lo que hemos hecho ha sido definir un punto de la red como fijo, es decir, llevar el punto  $A$ , por ejemplo, a las coordenadas  $(0, 0)$ . Parece de sentido común que la solución que proporcionará el óptimo consiste en llevar, no un punto u otro particular, sino el centro de gravedad a unas coordenadas fijas. Se puede demostrar que esto así, como se verá unas secciones más adelante.

Si el número de puntos es  $n$ , para llevar el centro de gravedad a las coordenadas  $(0, 0)$  la traslación que hay que aplicar es

$$-\frac{1}{n}(X_1 + X_2 + X_3 + \dots)$$

y análogamente para la  $Y$ . Un punto  $P$  pasa entonces a

$$X'_P = X_P - \frac{1}{n}(X_1 + X_2 + X_3 + \dots) = \frac{n-1}{n}X_P - \frac{1}{n}X_1 - \frac{1}{n}X_2 - \dots, \quad (15.1)$$

en donde los puntos suspensivos incluyen todos los puntos que faltan excepto el propio  $P$ . Análogamente para la  $Y$ .

Si hay por ejemplo 5 puntos la matriz  $A$  quedaría como se muestra en la página siguiente. Las columnas se corresponden con  $X_1, Y_1, X_2, Y_2, X_3, Y_3, X_4, Y_4, X_5$  y  $Y_5$ , en ese orden. Habría que eliminar de  $A$  las columnas relativas a parámetros que no existiesen en la matriz  $N^{-1}$  por ser los que se tomaron como fijos cuando se calculó la red.

La matriz  $A$  se escribe mejor como  $I - M$ , en donde  $I$  es la matriz identidad:

$$A = I - M = I - \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (15.2)$$

## 15 Cambio de los parámetros fijos, II: Global

$$A = \begin{pmatrix} \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 \\ 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} \\ -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 \\ 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} \\ -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 \\ 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} \\ -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} & 0 \\ 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 & -\frac{1}{5} \\ -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} & 0 \\ 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & -\frac{1}{5} & 0 & \frac{4}{5} \end{pmatrix}$$

Esto se repetirá para la escala promedio y el giro promedio. En el producto  $A\Sigma A^T$  la matriz  $I$  representa las varianzas que los puntos ya tienen mientras que la matriz  $M$  es el ajuste debido a la redefinición del sistema de coordenadas; en este caso el ajuste por la redefinición de la traslación.

### Escala promedio

En el capítulo anterior redefiníamos la escala del sistema de coordenadas tomando como fija la distancia entre dos puntos de la red. La solución global tendrá en cuenta todas las distancias entre todos los puntos de la red. En el capítulo dedicado a las transformaciones de coordenadas vimos que una medida de la escala de un conjunto de puntos en su conjunto, que tiene en cuenta todos por igual, es

$$\sum r^2,$$

en donde las distancias son respecto al centro de gravedad. Por lo tanto se puede definir la escala de la red como que esta cantidad sea igual a un cierto valor, y puesto que las coordenadas de los puntos han de mantenerse, ese valor ha de ser el que ya tienen. Por ejemplo, si esa cantidad vale 14 400 para el conjunto de puntos que se tiene, la escala del sistema de coordenadas la definimos como  $\sum r^2 = 14\,400$ . Lo que es fijo no es la distancia entre tal y cual punto sino esta cantidad.

Si un punto  $i$  se moviese de manera que su distancia pasase a ser  $r_i + \Delta r_i$ , para que la cantidad que define la escala del sistema se mantenga fija las coordenadas de todos los puntos han de dividirse

## 15 Cambio de los parámetros fijos, II: Global

por el aumento (o disminución) de  $\sum r^2$ . En concreto, si este valor era, por ejemplo, 14 400, el nuevo valor de  $r^2$  para cada punto será

$$(r_P^2)' = \frac{14\,400}{\sum r^2} r_P^2,$$

o lo que es lo mismo

$$r_P' = \frac{120}{\sqrt{\sum r^2}} r_P,$$

El número 120 es el valor concreto de  $\sqrt{\sum r^2}$  para el conjunto de puntos que se tiene, al que llamaremos  $S_0$ , y el denominador es el nuevo valor, una vez que  $r_i$  ha pasado a ser  $r_i + \Delta r_i$ , al que llamaremos  $S$ :

$$r_P' = \frac{S_0}{\sqrt{\sum r^2}} r_P, \quad (15.3)$$

El cambio de  $S_0$  a  $S$  es, en primer orden,

$$\sqrt{\sum r^2} \rightarrow \sqrt{\sum r^2} + \frac{2r_i}{2\sqrt{\sum r^2}} \Delta r_i = S_0 \left( 1 + \frac{r_i \Delta r_i}{S_0^2} \right).$$

Por lo tanto

$$r_P' = \frac{1}{1 + \frac{r_i \Delta r_i}{S_0^2}} r_P \approx \left( 1 - \frac{r_i \Delta r_i}{S_0^2} \right) r_P = r_P - \frac{r_i \Delta r_i}{S_0^2} r_P,$$

de donde

$$\frac{\partial r_P}{\partial r_i} = -\frac{r_i r_P}{\sum r^2}. \quad (15.4)$$

Al propio punto  $i$  hay que añadir el  $+1$  de  $r_i \rightarrow r_i + \Delta r_i$ .

Hay que pasar estas derivadas de  $r_P$  respecto de  $r_i$  a derivadas de las coordenadas de  $P$  respecto de las coordenadas de  $i$ . En primer lugar, si una variación en  $r$  es de  $kr$  para cierta cantidad  $k$ , las variaciones en  $x$  e  $y$  son respectivamente de  $kx$  y  $ky$ :

$$\frac{\partial x_P}{\partial r_i} = -\frac{x_P r_i}{\sum r^2}, \quad \frac{\partial y_P}{\partial r_i} = -\frac{y_P r_i}{\sum r^2}.$$

## 15 Cambio de los parámetros fijos, II: Global

En segundo lugar, una variación en  $x_i$  da lugar a una variación en  $r_i$  igual a la anterior multiplicada por  $x_i/r_i$ , y análogamente para  $y_i$ :

$$\begin{pmatrix} \frac{\partial x_j}{\partial x_i} & \frac{\partial x_j}{\partial y_i} \\ \frac{\partial y_j}{\partial x_i} & \frac{\partial y_j}{\partial y_i} \end{pmatrix} = -\frac{1}{\sum r^2} \begin{pmatrix} x_j x_i & x_j y_i \\ y_j x_i & y_j y_i \end{pmatrix}.$$

A lo que hay que añadir los +1 en la diagonal si el punto  $j$  es el propio punto  $i$ .

Por lo tanto la matriz A de cambio de varianzas queda

$$A = I - M = I - \frac{1}{\sum r^2} \begin{pmatrix} x_j x_i & x_j y_i \\ y_j x_i & y_j y_i \end{pmatrix}. \quad (15.5)$$

Hay que considerar dos efectos que podrían causar variaciones a mayores en los puntos  $j$ . El primero se debe a que si el centro de gravedad del conjunto de puntos tiene unos ciertos valores, (100, 200) por ejemplo, hemos llevado a cabo todos los desarrollos atendiendo a  $\sum r^2$  respecto al punto (100, 200), no respecto al centro de gravedad nuevo. Éste varía debido al desplazamiento del punto  $i$  y al desplazamiento inducido por este en cada punto  $j$ . La suma de estos últimos, incluyendo al propio punto  $i$ , es cero, por ser desplazamientos radiales hacia el centro de gravedad proporcionales a las distancias de cada punto a dicho centro. Alternativamente, estos desplazamientos son los recogidos en la matriz M, y la suma de una columna de dicha matriz, que corresponde a una variación de  $x_i$  o  $y_i$ , sumando por una parte las filas correspondientes a las coordenadas  $x$  y por otra las correspondientes a las coordenadas  $y$ , es

$$\frac{1}{\sum r^2} (x_i \sum x_j, \quad x_i \sum y_j, \quad y_i \sum x_j, \quad y_i \sum y_j),$$

en donde se muestran las cuatro combinaciones posibles. Estas cantidades son todas cero porque  $\sum x_j = 0$  y  $\sum y_j = 0$ . Pero sigue quedando el desplazamiento  $x_i$  o  $y_i$  original (lo representado por la matriz I en  $A = I - M$ ), por lo que el centro de gravedad sufre una variación.

El otro efecto es que los desplazamientos en la dirección tangencial los hemos ignorado, y dichos desplazamientos producen un cambio en la posición del centro de gravedad y por consiguiente un cambio en los valores de  $r^2$  para todos los puntos, cuya posición habría que ajustar para que  $\sum r^2$  no varíe.

## 15 Cambio de los parámetros fijos, II: Global

Estos efectos pueden ser ignorados porque la cantidad  $\sum r^2$  es mínima cuando el punto respecto al cual se consideran las distancias es el centro de gravedad, lo que significa que la variación de esta cantidad para pequeños desplazamientos del centro de gravedad es en primer orden nula.

### Giro promedio

Si bien una definición de un giro promedio análoga a la de escala promedio es más difícil que esta última, no lo es la obtención de las variaciones diferenciales de cada coordenada debidas a la variación de una de ellas, que se obtiene de modo inmediato por analogía con las variaciones de  $r$ , sustituyendo desplazamientos en la dirección radial,  $r$ , por desplazamientos en la dirección tangencial,  $t$ :

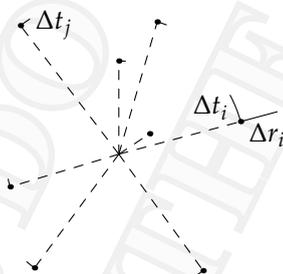


Fig. 15.1: Pequeños desplazamientos tangenciales

$$\frac{\partial t_j}{\partial t_i} = -\frac{r_j r_i}{\sum r^2}.$$

Los desplazamientos según las direcciones  $(t, x, y)$  están en proporción  $(r, -y, x)$ , de donde

$$\frac{\partial x_j}{\partial t_i} = -\frac{y_j r_i}{\sum r^2}, \quad \frac{\partial y_j}{\partial r_i} = -\frac{x_j r_i}{\sum r^2}.$$

y finalmente

$$\begin{pmatrix} \frac{\partial x_j}{\partial x_i} & \frac{\partial x_j}{\partial y_i} \\ \frac{\partial y_j}{\partial x_i} & \frac{\partial y_j}{\partial y_i} \end{pmatrix} = -\frac{1}{\sum r^2} \begin{pmatrix} y_j y_i & -y_j x_i \\ -x_j y_i & x_j x_i \end{pmatrix},$$

## 15 Cambio de los parámetros fijos, II: Global

$$A = I - M = I - \frac{1}{\sum r^2} \begin{pmatrix} y_j y_i & -y_j x_i \\ -x_j y_i & x_j x_i \end{pmatrix}. \quad (15.6)$$

### Combinación de las distintas libertades

Los desplazamientos diferenciales deducidos para variaciones radiales y tangenciales son ortogonales entre sí y además vimos que en el caso de la escala libre los desplazamientos tangenciales no inducen variaciones en los otros puntos, y por la misma razón que esto es así los desplazamientos radiales no inducen variación alguna en los otros puntos en el caso de giro libre, aunque no se mencionó. Por ello si tanto escala como giro son libres la matriz  $M$ , que representa las pequeñas variaciones en cada coordenada debidas a un desplazamiento en cada coordenada, será simplemente la suma de las dos.

Esto puede comprobarse efectuando el producto de las matrices  $A$  en

$$A_2 A_1 \Sigma A_1^T A_2^T,$$

siendo  $A_1$  la correspondiente a giro libre y  $A_2$  la correspondiente a escala libre:

$$A_2 A_1 = (I - M_2)(I - M_1) = I - (M_1 + M_2) + M_2 M_1,$$

y se comprueba que el producto  $M_2 M_1$  da cero. Este producto está compuesto de bloques de la forma

$$\begin{pmatrix} x_j x_i & x_j y_i \\ y_j x_i & y_j y_i \end{pmatrix} \begin{pmatrix} y_i y_k & -y_i x_k \\ -x_i y_k & x_i x_k \end{pmatrix},$$

ya que en un producto de dos matrices cada elemento de la columna  $i$  de la primera matriz se multiplica por cada elemento de la fila  $i$  de la segunda. Al desarrollar el producto queda

$$\begin{pmatrix} x_j(x_i y_i - y_i x_i) y_k & x_j(-x_i y_i + y_i x_i) x_k \\ y_j(x_i y_i - y_i x_i) y_k & y_j(-x_i y_i + y_i x_i) x_k \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

El que el resultado sea cero es una manifestación de la ortogonalidad de ambos fenómenos.

## 15 Cambio de los parámetros fijos, II: Global

La matriz  $M$  es entonces la suma de las dos matrices  $M$ , y finalmente

$$A = I - M = I - \frac{1}{\sum r^2} \begin{pmatrix} x_j x_i + y_j y_i & -(y_j x_i - x_j y_i) \\ y_j x_i - x_j y_i & x_j x_i + y_j y_i \end{pmatrix}. \quad (15.7)$$

También la traslación es independiente del giro y la escala. El producto de una fila de la matriz  $M$  de la traslación por una columna de la matriz  $M$  de escala o giro da lugar a una de las seis siguientes posibilidades:

$$\frac{x_i}{n \sum r^2} \sum x_j, \quad \pm \frac{y_i}{n \sum r^2} \sum x_j, \quad \pm \frac{x_i}{n \sum r^2} \sum y_j, \quad \frac{y_i}{n \sum r^2} \sum y_j.$$

Todas valen cero por ser cero  $\sum x_j$  y  $\sum y_j$ .

### Interpretaciones de las transformaciones anteriores

La interpretación de la traslación libre está clara: el sistema queda definido en posición fijando unas coordenadas para el centro de gravedad. Supongamos que estas coordenadas son  $(100, 200)$ . La obtención de las coordenadas ajustadas en un sistema definido de esta manera se hace en dos pasos. En primer lugar se obtienen las coordenadas ajustadas tomando unas coordenadas fijas para que el sistema quede definido y poder resolver el ajuste. A continuación se calcula el centro de gravedad  $(x_G, y_G)$  y se desplazan todos los puntos la cantidad  $(100 - x_G, 200 - y_G)$ . Este proceso en dos pasos para obtener las coordenadas  $(x, y)$  ajustadas es un estimador como otro cualquiera, y si repitiésemos el proceso muchas veces para los mismos puntos con distintas medidas obtendríamos la distribución de las coordenadas ajustadas, y son las varianzas de estas lo que se obtuvo más arriba en este capítulo.

Que no lleguemos a efectuar la suma  $(100 - x_G, 200 - y_G)$  para las coordenadas de los puntos no impide obtener las varianzas en ese sistema, quedando reducida la elección de coordenadas fijas del ajuste a un mero paso auxiliar para la resolución del problema. Pero si el conjunto de puntos sirve por ejemplo para definir un sistema de coordenadas en una sala, y cada cierto tiempo se vuelven a medir y ajustar, entonces conviene no omitir dicha suma. De esta manera estamos haciendo real esta manera de definir el sistema, y tras un cierto tiempo,

## 15 Cambio de los parámetros fijos, II: Global

cuando ya se hayan medido los puntos unas cuantas veces, se podrá comprobar que los conjuntos de coordenadas que resultan llevando siempre el centro de gravedad a un valor fijo tienen menos variación que los conjuntos que resultan si el sistema se define asignando unas coordenadas fijas a uno de los puntos, por ejemplo  $P_6 = (100, 1\ 000)$ .

La escala en caso de que sea libre la definimos en la solución global como que  $\sqrt{\sum r^2} = S_0$ . Al contrario que en el caso anterior, el valor elegido para  $S_0$  afecta a las varianzas que se obtienen. La razón es bien simple; si tomamos un valor de  $S_0$  de la décima parte las distancias y coordenadas respecto al centro de gravedad se dividen todas por diez y con ello sus incertidumbres. Por tanto en realidad las precisiones de los puntos no dependen de  $S_0$ , entendidas esas como su precisión en relación a la escala elegida para el conjunto. No tendría sentido que fuese de otra manera ya que el valor escogido para  $S_0$  es arbitrario, al igual que lo era el escogido para las coordenadas del centro de gravedad, y no puede ser que las precisiones intrínsecas de los puntos dependan del valor escogido.

Una vez más, si se tiene un conjunto de puntos cuyo objeto es definir un sistema de coordenadas, la escala de dicho sistema es arbitraria y los puntos se van a remedir al cabo de un tiempo, la mejor elección posible para la escala, en el sentido de que tiene en cuenta todos los puntos y, como veremos en la siguiente sección, da lugar a las elipses de error lo más pequeñas posibles en promedio, es la que hemos desarrollado en este capítulo.

La definición del giro es más problemática. Si en el cálculo del factor de escala entre dos sistemas aparecen productos  $xx + yy$ , lo que sirvió para motivar la elección de la definición global de escala en este capítulo, los productos  $xy - yx$  y  $xx + yy$  que aparecen en el cálculo de la rotación entre dos sistemas no se pueden emplear para definir el giro de un sistema. El primero de ellos porque obviamente vale cero y el segundo porque no varía al rotar el conjunto de puntos. El siguiente ejemplo muestra que no se puede emplear ninguna expresión simétrica en todos los puntos. Sea un conjunto de puntos situados sobre una circunferencia a intervalos regulares. Si hay un punto cada  $\varepsilon$  grados un giro de ángulo  $\varepsilon$  deja invariante cualquier otra expresión simétrica en las coordenadas de los puntos. El hecho de que lo se pueda emplear ninguna expresión simétrica en los puntos parece ir directamente en contra de la posibilidad de una definición global

## 15 Cambio de los parámetros fijos, II: Global

del giro al modo de las obtenidas para la traslación o la escala.

Sin embargo, la definición de la escala como que la cantidad  $\sum(xx + yy)$  no varíe, y el modo en todo similar en cómo se obtienen las matrices  $M$  en uno y otro caso, hace pensar que de algún modo la definición del giro está ligada a una invariabilidad de  $\sum(xy - yx)$  o  $\sum(xx + yy)$ . La interpretación correcta se tiene si se presta atención a que en la obtención del factor de escala entre dos conjuntos los productos que se suman son de coordenadas en un mismo sistema, mientras que en la obtención del giro están entremezclados uno y otro sistemas; v. g.,  $\sum(x_1x_1 + y_1y_1)$  en el primer caso y  $\sum(y_2x_1 - x_2y_1)$  así como  $\sum(x_2x_1 + y_2y_1)$  en el segundo.

La invariabilidad buscada es pues la de esas expresiones para las coordenadas  $(x, y)$  en relación a las coordenadas  $(x, y)$  que los puntos tienen, tomadas estas últimas como fijas, como números:

$$\sum(yx_0 - xy_0), \quad \sum(xx_0 + yy_0).$$

Estas dos cantidades son, como se vio en el capítulo 12, proporcionales a  $\sin \theta$  y  $\cos \theta$ , siendo  $\theta$  el ángulo de giro entre el sistema variable  $\{(x, y)\}$  y el sistema fijo  $\{(x_0, y_0)\}$ . Por lo tanto la definición global de la rotación en un sistema de giro libre es que dichas cantidades estén en una determinada proporción, es decir, en fijar un valor para

$$\frac{\sum(yx_0 - xy_0)}{\sum(xx_0 + yy_0)}. \quad (15.8)$$

Según el valor que escojamos sucede, al igual que con la escala, que las elipses de error varían en apariencia, pero la precisión intrínseca de los puntos no varía. Así, si giramos el sistema las elipses de error girarán, como no podía ser de otra manera pues de lo contrario las precisiones de los puntos sí que estarían realmente variando.

Dado que entre todos los valores que podemos escoger tomaremos  $\theta = 0$  los valores anteriores son

$$\sum(yx_0 - xy_0) = 0, \quad \sum(xx_0 + yy_0) = \sum r_0^2 = S_0^2. \quad (15.9)$$

La segunda de ellas no se puede emplear para definir el giro pues para un giro de  $0^\circ$  la derivada del coseno es 0 y su valor igual a 1, o lo que es lo mismo  $\sum(xx_0 + yy_0) = S_0^2$ , no permite distinguir en primer orden de aproximación los giros pequeños.

## 15 Cambio de los parámetros fijos, II: Global

Por tanto la definición del giro queda como

$$\sum (yx_0 - xy_0) = 0. \quad (15.10)$$

Si se mide el mismo conjunto de puntos en múltiples ocasiones y se ajusta su orientación mediante esta definición se verá que las elipses de error que resultan son en promedio más pequeñas que si se define tomando como fija la dirección entre un par de puntos cualesquiera de la red.

Si se quisiese tomar en (15.8) un valor distinto de cero es más fácil aplicar el giro correspondiente a las coordenadas  $(x_0, y_0)$  y después emplear la condición (15.10). Por otra parte, si queremos aplicar esta última condición y el giro entre los sistemas  $(x, y)$  y  $(x_0, y_0)$  es grande habrá que aplicar la corrección de manera exacta y no de modo lineal como incrementos  $\Delta t$ , puesto esto último es simplemente una aproximación de primer orden al giro. También se puede aplicar primero un giro aproximado y la corrección residual que reste aplicarla linealmente.

### Resumen de fórmulas

El ajuste varianzas óptimo para la posición, escala y/o orientación libres viene dado por una matriz  $A$ , siendo las varianzas resultantes en función de las originales  $\Sigma' = A\Sigma A^T$ .

$$A = I - M;$$

$M$  es la suma de las matrices  $M$  correspondientes a cada una de las libertades de la red. Estas matrices están compuestas por bloques  $2 \times 2$  correspondientes a las filas  $x_j$  e  $y_j$  y las columnas  $x_i$  e  $y_i$ , para cada  $i$  y para cada  $j$ . Estos bloques son las siguientes para cada libertad:

$$\begin{aligned} (M_{\text{tras.}})_{j,i} &= \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ (M_{\text{esc.}})_{j,i} &= \frac{1}{\sum r^2} \begin{pmatrix} x_j x_i & x_j y_i \\ y_j x_i & y_j y_i \end{pmatrix}, \\ (M_{\text{giro}})_{j,i} &= \frac{1}{\sum r^2} \begin{pmatrix} y_j y_i & -y_j x_i \\ -x_j y_i & x_j x_i \end{pmatrix}. \end{aligned} \quad (15.11)$$

## 15 Cambio de los parámetros fijos, II: Global

Si por ejemplo es libre tanto la traslación como la escala y el giro se tendrá

$$A = I - (M_{\text{tras.}} + M_{\text{esc.}} + M_{\text{giro}}).$$

Si se tiene un conjunto de coordenadas  $(x_0, y_0)$  y se obtiene un nuevo conjunto  $(x, y)$ , y se quiere mantener la misma definición del sistema de coordenadas, habrá que efectuar pequeños ajustes a las coordenadas  $(x, y)$  en traslación, escala y/o giro.

Sea

$$\frac{1}{n} (\sum x_0, \sum y_0) = (x_{G0}, y_{G0}), \quad \frac{1}{n} (\sum x, \sum y) = (x_G, y_G),$$

$$\sum (x_0^2 + y_0^2) = S_0^2, \quad \sum (x^2 + y^2) = S^2,$$

$$\sum (yx_0 - xy_0) = S_\theta, \quad \sum (xx_0 + yy_0) = C_\theta.$$

Los ajustes a aplicar son los que se indican a continuación. Para la escala y el giro las coordenadas  $(x, y)$  son respecto al centro de gravedad de los puntos.

Traslación:

$$(x, y) \rightarrow (x, y) - (x_G - x_{G0}, y_G - y_{G0}).$$

Escala:

$$(x, y) \rightarrow \frac{S_0}{S} (x, y) = (x, y) - \frac{S - S_0}{S} (x, y).$$

Giro:

$$\sin \theta = \frac{S_\theta}{\sqrt{S_\theta^2 + C_\theta^2}}, \quad \cos \theta = \frac{C_\theta}{\sqrt{S_\theta^2 + C_\theta^2}}.$$

$$(x, y) \rightarrow (x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta) \quad (\text{Girar } -\theta).$$

Si  $S_\theta$  es muy pequeño se puede tomar directamente  $\sin \theta \approx \theta \approx S_\theta / \sum r_0^2$ , y aplicar la corrección según la tangente  $t$  en vez de según el arco:

$$(x, y) \rightarrow (x, y) - \theta(-y, x).$$

En las expresiones (15.11) las coordenadas a emplear son las de subíndice 0. Previamente hay que multiplicar la matriz de varianzas por  $S_0^2/S^2$  y girarla  $-\theta$ .

**Demostración de óptimo para la traslación**

Vamos a demostrar aquí que las elecciones hechas para los casos de traslación, escala y giro libres son las mejores posibles. Que sean las mejores posibles significa que la suma de los elementos de la diagonal de la matriz  $\Sigma$  es la menor posible. La mayor dificultad consiste en identificar cuáles son todas elecciones posibles. Por ejemplo, en caso de que la traslación sea libre, ¿cuáles son todas las maneras posibles de definir el sistema de coordenadas de acuerdo a ese grado de libertad?

Para fijar la red en posición si la traslación es libre hemos empleado dos criterios hasta ahora: que las coordenadas de un punto tengan un determinado valor y que las coordenadas del centro de gravedad tengan un determinado valor. Centrándonos por ejemplo en la coordenada  $x$ , estos criterios son

$$x_1 = C, \quad \text{y} \quad \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n = C.$$

La letra C representa una constante cuyo valor exacto es irrelevante.

Parece que cualquier criterio lógico será que una combinación lineal de coeficientes positivos de los distintos puntos sea igual a un cierto valor, y siempre podremos multiplicar los coeficientes por la constante adecuada para que su suma sea 1. Pero nada impide tomar un criterio más excéntrico como

$$x_1^2 + x_2x_3 - x_7 = C, \quad \log x_1 - x_2 + \frac{1}{3}x_3 = C,$$

etcétera. Cualquiera que sea la elección, una pequeña variación  $\Delta x_i$  dará lugar a una variación en el valor de la expresión que se quiere hacer igual a C igual en primer orden a un cierto múltiplo de esa variación:  $k_i \Delta_i$ . Para que la expresión vuelva a ser igual a C habrá que restar (o sumar) a todas las coordenadas una cierta cantidad que será otro múltiplo de  $\Delta_i$ . Una variación de  $\varepsilon$  igual en todas las coordenadas  $x$  dará lugar a una variación en la expresión que es, en primer orden, un múltiplo de  $\varepsilon$ , y este múltiplo es  $\sum k_i$ . Por lo tanto habrá que restar a todas las coordenadas  $\varepsilon = \frac{k_i}{\sum k_i} \Delta_i$ .

Estos valores  $k_i / \sum k_i$  son los que hay que situar en la matriz M, en la columna correspondiente a  $x_i$ . Por tanto para cada variación de cada coordenada  $x_i$  las correcciones a aplicar a los puntos (los coeficientes de su columna en la matriz) son para todos iguales, y la matriz M en el caso más general posible está formada por columnas en

## 15 Cambio de los parámetros fijos, II: Global

las que todos los valores son iguales (para la coordenada correspondiente, obviamente, para la otra coordenada los coeficientes son cero) y la suma por filas es 1.

Antes de continuar con la demostración hacemos ver que esta conclusión significa que, localmente, la condición equivale a una combinación lineal de las coordenadas, cuyos coeficientes sumen 1, igual a una constante dada. La condición será de la forma

$$F(x_1, x_2, \dots, x_n) = C$$

para una función tan complicada como se quiera y una constante  $C$  cuyo valor exacto es irrelevante. Para los valores  $x_1, x_2, \dots, x_n$  que tienen las coordenadas de los puntos la función  $F$  será igual a un cierto valor  $C_1$ . Tomamos  $C$  igual a  $C_1$ . Para pequeñas variaciones de las coordenadas la condición es, en primer orden de aproximación,

$$F(x_1, x_2, \dots, x_n)_0 + k_1\Delta_1 + k_2\Delta_2 + \dots + k_n\Delta_n = C_1.$$

Es decir,

$$C_1 + k_1\Delta_1 + k_2\Delta_2 + \dots + k_n\Delta_n = C_1,$$

$$k_1\Delta_1 + k_2\Delta_2 + \dots + k_n\Delta_n = 0.$$

Y podemos multiplicar toda la ecuación por una constante de manera que  $\sum k_i = 1$ . Los valores  $k_i / \sum k$  no van a cambiar.

La forma exacta de la condición da igual, lo único que importa es la matriz  $M$  a la que da lugar, hasta el punto que de las distintas condiciones que dan lugar a la misma matriz  $M$  pueden considerarse en esencia la misma, al menos localmente.

Como la forma de la matriz  $M$  no varía si se multiplican todos los valores de  $k$  por una constante, supondremos que  $\sum k_i = 1$ . Por otro lado, ya que las coordenadas  $x$  e  $y$  se tratan de manera independiente separamos ambas en las matrices, y la matriz  $M$  para una de ellas queda

$$M = \begin{pmatrix} k_1 & k_2 & \dots & k_n \\ k_1 & k_2 & \dots & k_n \\ \dots & \dots & \dots & \dots \\ k_1 & k_2 & \dots & k_n \end{pmatrix}.$$

La traza (suma de los elementos de la diagonal) de  $A\Sigma A^T$  es

$$\text{Tr} \left[ (I - M)\Sigma(I - M)^T \right] = \text{Tr}(\Sigma) - \text{Tr}(M\Sigma) - \text{Tr}(\Sigma M^T) + \text{Tr}(M\Sigma M^T).$$

## 15 Cambio de los parámetros fijos, II: Global

La traza de una matriz es igual a la de su traspuesta, y  $(M\Sigma)^T = \Sigma^T M^T = \Sigma M^T$ , de modo que la expresión anterior se reduce a

$$\text{Tr}(\Sigma) - 2\text{Tr}(\Sigma M^T) + \text{Tr}(M\Sigma M^T). \quad (15.12)$$

Por otra parte  $\text{Tr}(\Sigma)$  es fijo, de modo que lo que hay que minimizar es  $-2\text{Tr}(\Sigma M^T) + \text{Tr}(M\Sigma M^T)$ .

En la matriz  $M^T$  todas las columnas son iguales, por lo que en el producto  $\Sigma M^T$  todas las columnas son iguales, y la traza de esta matriz es pues igual a la suma de los elementos de una columna. A su vez, el producto  $M(\Sigma M^T)$  es el producto de una matriz en la que todas las filas son iguales por otra en la que todas las columnas son iguales, por lo que todos los elementos de la matriz producto son iguales. Esto también se puede razonar de la siguiente manera: Si las variables representadas en la matriz  $\Sigma$  son  $X$ , sea  $Y = MX$ . Como todas las filas de  $M$  son iguales, todas las variables de  $Y$  son iguales, es decir, se trata de la misma variable  $y$ , por lo que su matriz de varianzas constará de un único elemento  $\sigma_{yy}$  repetido en todas las posiciones.

Si los elementos de una columna de  $\Sigma M^T$  son  $c_1, c_2, \dots, c_n$ , los de  $M\Sigma M^T$  son  $k_1 c_1 + k_2 c_2 + \dots + k_n c_n$ , y las trazas de una y otra matriz son

$$c_1 + c_2 + \dots + c_n \quad \text{y} \quad n(k_1 c_1 + k_2 c_2 + \dots + k_n c_n)$$

y la suma a minimizar

$$(nk_1 - 2)c_1 + (nk_2 - 2)c_2 + \dots + (nk_n - 2)c_n.$$

Sustituyendo los elementos  $c_i$  por lo que valen queda

$$\begin{aligned} &(nk_1 - 2)(k_1 \sigma_{11} + k_2 \sigma_{12} + \dots + k_n \sigma_{1n}) \\ &(nk_2 - 2)(k_1 \sigma_{21} + k_2 \sigma_{22} + \dots + k_n \sigma_{2n}) \\ &\dots \\ &(nk_n - 2)(k_1 \sigma_{n1} + k_2 \sigma_{n2} + \dots + k_n \sigma_{nn}). \end{aligned}$$

Si todos los  $k_i$  son iguales a  $1/n$  la suma de todo esto se reduce a

$$-\frac{1}{n} \sum \sigma_{ij}.$$

## 15 Cambio de los parámetros fijos, II: Global

Esta cantidad es negativa (es decir, el sumatorio es positivo) ya que el sumatorio es la varianza de la variable "suma de todas las variables de la matriz  $\Sigma$ ", puesto que

$$\sum \sigma_{ij} = (1 \ 1 \ \dots \ 1) \Sigma \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}.$$

Si no son todos  $1/n$  será  $k_1 = 1/n + \alpha_1, k_2 = 1/n + \alpha_2, \dots, k_n = 1/n + \alpha_n$ , con  $\sum \alpha_i = 0$ , pues se ha de cumplir  $\sum k_i = 1$ . El primero de los anteriores sumandos, por ejemplo, queda

$$\begin{aligned} & (n\alpha_1 - 1) \left( (\alpha_1 + \frac{1}{n})\sigma_{11} + (\alpha_2 + \frac{1}{n})\sigma_{12} + \dots + (\alpha_n + \frac{1}{n})\sigma_{1n} \right) \\ &= -\frac{1}{n} \sum \alpha_{1j} + n(\alpha_1^2\sigma_{11} + \alpha_1\alpha_2\sigma_{12} + \dots + \alpha_1\alpha_n\sigma_{1n}) \\ & \quad + (\alpha_1 - \alpha_2)\sigma_{12} + \dots + (\alpha_1 - \alpha_n)\sigma_{1n}. \end{aligned}$$

Todos los elementos  $(\alpha_1 - \alpha_j)\sigma_{1j}$  de la segunda línea se anulan con el correspondiente  $(\alpha_j - \alpha_1)\sigma_{j1}$  que aparece en el desarrollo del sumando  $j$ -ésimo. Una vez eliminados todos los términos que se anulan la suma total queda

$$-\frac{1}{n} \sum \alpha_{1j} + n \sum \alpha_i \alpha_j \sigma_{ij}.$$

El sumatorio lo podemos escribir como

$$(\alpha_1 \ \alpha_2 \ \dots \ \alpha_n) \Sigma \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{pmatrix};$$

el decir, se trata de la varianza de una variable y es por tanto positivo, salvo que tomemos todos los  $\alpha_i$  iguales a cero.

Hemos demostrado pues que el mínimo para la suma se da si  $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ ; es decir, si todos los  $k$  son iguales a  $1/n$ . Esto, como ya sabemos, significa que la condición que define el sistema en traslación consiste en fijar las coordenadas del centro de gravedad.

## 15 Cambio de los parámetros fijos, II: Global

Nótese que no ha sido necesario hacer uso de la condición  $\sum \alpha = 0$ , resp.  $\sum k = 1$ . El valor mínimo de la traza, (15.12), es

$$\text{Tr}(\Sigma) - \frac{1}{n} \sum \sigma_{ij}. \quad (15.13)$$

Si volvemos a redefinir el sistema en traslación, y volvemos a definirlo fijando las coordenadas del centro de gravedad, nada cambiará y las varianzas seguirán siendo las mismas. Sea  $\Sigma' = A\Sigma A^T$ . Esto implica que

$$\text{Tr}(\Sigma') - \frac{1}{n} \sum \sigma'_{ij} = \text{Tr}(\Sigma'),$$

es decir,  $\sum \sigma'_{ij} = 0$ . Ya vimos que esta suma es la varianza de una variable definida como la suma de todas las variables que están representadas en la matriz  $\Sigma'$ . En este caso esto es la suma de todas las coordenadas. Dado que el valor del centro de gravedad es fijo también lo es el de la suma de todas las coordenadas, de ahí que su varianza sea 0.

Dada la matriz  $\Sigma$ , la varianza del centro de gravedad es pues  $\frac{1}{n^2} \sum \sigma_{ij}$ . En (15.13) se está restando  $n$  veces esta cantidad, lo que podemos interpretar como que al redefinir la traslación del sistema fijando unas coordenadas para el centro de gravedad, de la varianza de cada punto restamos por término medio la varianza que tenía el centro de gravedad. Y por ser esta la solución óptima ya no es posible disminuir más las varianzas de los puntos.

### Independencia de las tres optimizaciones

La definición de la escala de la red será fijando un valor para una distancia, o para una combinación de distancias, o para cualquier función de las coordenadas, pero tal que para ajustar esa función al valor fijado se multiplican todas las coordenadas respecto a un punto. Este punto puede ser un punto de coordenadas fijas, por ejemplo el origen de coordenadas, o un punto definido a su vez como una función de las coordenadas de los puntos, como por ejemplo uno de los puntos o el centro de gravedad de todos los puntos.

La diferencia entre aplicar un factor de escala respecto a uno u otro punto es una traslación, de modo que la multiplicación de distancias

## 15 Cambio de los parámetros fijos, II: Global

respecto a un punto cualquiera se puede descomponer en una multiplicación respecto a otro punto cualquiera más una traslación. Si además de la escala también es libre la traslación es indistinto el punto respecto al cual escalemos, ya que la traslación se puede variar libremente. Escogemos entonces el centro de gravedad como origen del escalado, que tiene la ventaja de que la suma de las variaciones en las coordenadas de los puntos como resultado de un escalado respecto a ese punto es cero. Como consecuencia de ello la matriz  $M$  de un escalado respecto al centro de gravedad y la matriz  $M$  de una traslación son ortogonales, de lo que se sigue, como ya se vio:

$$A_{\text{tras. esc.}} \Sigma A_{\text{tras. esc.}}^T = A_{\text{tras.}} A_{\text{esc.}} \Sigma A_{\text{esc.}}^T A_{\text{tras.}}^T$$

de donde la optimización de la matriz  $A$  para una traslación más un escalado se obtiene optimizando de manera independiente cada una de estas dos libertades.

El caso de que la escala sea libre pero no la traslación resulta difícil de imaginar. Precisamente porque un escalado de todos los puntos produce una traslación neta, salvo que se efectúe respecto al centro de gravedad, no parece que tenga sentido decir que la escala es libre pero la traslación no. Si así fuese parece que lo lógico es entender que las coordenadas del centro de gravedad han de mantenerse.

De modo que el escalado se aplicará respecto al centro de gravedad, siendo de esta manera independientes (ortogonales) el escalado y la traslación, por lo que ambos óptimos obtenidos de manera independiente, una vez combinados, constituyen el óptimo total en el caso de que traslación y escala sean libres.

El razonamiento para el giro libre es en todo punto análogo al de la escala: La diferencia entre aplicar una cierta rotación respecto a un punto u otro es una traslación, y si la rotación se aplica respecto al centro de gravedad la traslación total es cero. Por lo tanto se sigue la misma relación de independencia respecto a la traslación y la posibilidad de optimizar ambas libertades por separado. Por otra parte un escalado y un giro aplicados desde un mismo punto, no necesariamente el centro de gravedad, siempre son ortogonales.

Por tanto, si el factor de escala y el giro se aplican respecto al centro de gravedad, como por otra parte parece lógico, la optimización de cualquier combinación de las tres libertades se reduce a una aplicación sucesiva de los óptimos para cada una de ellas.

## 15 Cambio de los parámetros fijos, II: Global

### Principio de demostración de óptimo para escala y giro

La condición de escala es una función cualquiera de las coordenadas igualada a un valor:

$$F(x_1, y_1, x_2, y_2, \dots, x_n, y_n) = C.$$

Para aplicar una escala o un giro respecto al centro de gravedad la forma más adecuada de expresar la posición de un punto es en coordenadas polares respecto al centro de gravedad mediante dos valores  $(r, t)$ . El valor de  $r$  es la distancia al centro de gravedad, mientras que el valor de  $t$  es la distancia según el arco de radio  $r$  contada a partir de un cierto origen. Este origen lo situamos para cada punto según una dirección que hacemos coincidente con el radio del punto, de manera que los valores de  $t$  son cero para todos los puntos. Es importante entender estas direcciones de origen de los valores de  $t$  como fijas para cada punto, de modo si sus coordenadas cambiasen (por ejemplo, porque se vuelven a medir y calcular las coordenadas) los valores de  $t$  ya no serían cero. De esta manera estos orígenes son unos radios bien definidos y que una vez fijados no dependen de las coordenadas de los puntos.

Expresado en este sistema de coordenadas la condición tomará otra forma distinta

$$F_1(r_1, t_1, r_2, t_2, \dots, r_n, t_n) = C_1.$$

Una vez linealizada en torno a los valores  $(r, t)$  que tienen los puntos, y tomando  $C_1$  igual al valor que arroja  $F_1$  para los valores  $(r_i, t_i)$  que efectivamente se tienen, queda de la forma

$$k_1 \Delta r_1 + k'_1 \Delta t_1 + k_2 \Delta r_2 + \dots + k'_n \Delta t_n = 0. \quad (15.14)$$

Al aplicar un factor de escala respecto al centro de gravedad los valores de  $r$  y  $t$  se multiplican por dicho factor, lo que significa que varían en proporción a sí mismos. Si el factor de escala aplicado es  $1 + \varepsilon$  la variación total es

$$\varepsilon (k_1 r_1 + k'_1 t_1 + \dots + k_n r_n + k'_n t_n).$$

Con el sistema de coordenadas elegido los valores de  $t$  son cero. Así, si la coordenada  $r$  del punto  $i$  varía una cantidad  $\Delta r_i$  la condición

## 15 Cambio de los parámetros fijos, II: Global

(15.14) implica

$$k_i \Delta r_i - \varepsilon (k_1 r_1 + k_2 r_2 + \dots + k_n r_n) = 0; \quad (15.15)$$

mientras que si la variación es en  $t_i$  la condición será

$$k'_i \Delta t_i - \varepsilon (k_1 r_1 + k_2 r_2 + \dots + k_n r_n) = 0; \quad (15.16)$$

En la ecuación (15.15) podemos aprovechar lo hecho en la demostración del óptimo para la traslación. Si todos los valores de  $r$  fuesen iguales esta ecuación sería la misma que la de una traslación. Al ser unos valores de  $r$  mayores que otros es como si se tratase de una traslación en la que cada punto está repetido  $r_i$  veces. De modo que podemos interpretarlo como una traslación con un total de  $\sum r_i$  puntos. En la sección anterior se demostró que el óptimo se consigue si cada  $k$  vale  $1 / \sum r_i$ . El punto  $i$  real es la suma de  $r_i$  puntos ficticios, por lo que el valor de  $k_i$  en (15.15) ha de ser igual a  $r_i / \sum r_i$ . Esto además completa la validez de dicha ecuación, ya que para que la analogía sea exacta es necesario que el primer sumando,  $k_i \Delta r_i$ , se corresponda con la suma de variaciones iguales  $\Delta_i$  en los  $r_i$  puntos ficticios, cada una de las cuales daría lugar a  $\frac{1}{\sum r_i} \Delta_i$ ; o sea que la suma total tiene que ser  $\frac{r_i}{\sum r_i} \Delta_i$ , como en efecto se obtiene tomando  $k_i = r_i / \sum r_i$ .

El óptimo para los valores de  $k$  obtenido en la traslación es que fuesen todos iguales, no necesariamente iguales a  $1/n$ . Esto último se tomó para que su suma fuese 1. Ahora es más cómo tomar  $k_i = r_i$  simplemente, sin dividir por  $\sum r_i$ .

Los coeficientes de la matriz M, que es lo que finalmente interesa, son, en la columna relativa a  $r_i$ ,

$$\frac{\Delta r_j}{\Delta r_i} = \frac{-\varepsilon r_j}{\Delta r_i} = \frac{-\frac{k_i \Delta r_i}{\sum r_i k_i} r_j}{\Delta r_i} = \frac{r_i r_j}{\sum r^2}.$$

Esto es la ecuación (15.4). En las filas de los  $t_j$  los coeficientes son cero ya que los valores no varían al escalar por  $(1 - \varepsilon)$ .

El hecho de que los  $k'_i$  no aparezcan en la ecuación (15.15) no significa que valgan cero, sino que pueden tomar cualquier valor. Esto se debe a que valgan lo que valgan  $k'_i t_i$  va a valer cero. Para determinar los valores de  $k'_i$  hay que emplear la ecuación (15.16). Esta ecuación determina lo que ha de variar cada coordenada  $r_j$  como consecuencia

## 15 Cambio de los parámetros fijos, II: Global

de una variación  $\Delta t_i$  en el punto  $i$ . Esta variación es

$$-\frac{k'_i r_j}{\sum r^2} \Delta t_i.$$

(Ya sea han sustituido los valores  $k_i$  por  $r_i$ ). Estos son, eliminando el factor  $\Delta t_i$ , los coeficientes de la matriz M en las filas correspondientes a  $r_j$ .

Habría que demostrar que la mínima traza de  $A\Sigma A^T$  se obtiene si todos los  $k'_i$  son cero. Esto es muy fácil, pero habremos optimizado primero los valores de  $k_i$  ignorando los  $k'_i$ , y obtenidos los primeros habremos demostrado que lo mejor para los segundos es que valgan cero. La demostración correcta exige determinar simultáneamente los  $k_i$  y  $k'_i$  para obtener el mínimo absoluto posible. En la sección siguiente se aborda el problema con una generalidad mucho mayor, obteniendo una solución que como casos particulares proporciona el mínimo para la escala y para el giro.

### **Demostración de óptimo general**

En el caso de poder aplicar un factor de escala, si una coordenada varía una cierta cantidad, haciendo que se rompa la condición  $F(r_1, \dots, t_n) = C$ , se multiplican todas las coordenadas por una constante  $(1 + \varepsilon)$ , lo que da lugar a una variación en cada coordenada de  $\varepsilon r_i$ , y hay que escoger  $\varepsilon$  de manera que la variación total de F anule la variación en la misma F que supuso la variación primera de una coordenada. Análogamente, en el giro la variación de las coordenadas es ahora cero para las  $r$  y proporcional a  $r$  para las  $t$ .

Sea cual sea la libertad del sistema, el ajuste que de acuerdo a la misma se efectúe sobre el conjunto de coordenadas para restaurar el valor original de la función F será para cada coordenada proporcional a la magnitud del ajuste y a una constante de proporcionalidad distinta para cada coordenada. En el caso de la traslación es 1 (o  $-1$ , según se mire), en el escalado proporcional a  $r$  y en el giro igual a cero para la dirección desde el origen del giro hasta el punto y proporcional a la distancia para la dirección perpendicular.

Sean  $x_1, \dots, x_n$  las coordenadas, en donde la letra  $x$  representa una magnitud abstracta, que puede ser la coordenada  $x$ , la  $y$ ,  $r$ ,  $t$  o cual-

## 15 Cambio de los parámetros fijos, II: Global

quier otra. La condición  $F(x_1, \dots, x_n) = C$  una vez linealizada se reduce a

$$k_1\Delta_1 + k_2\Delta_2 + \dots + k_n\Delta_n = 0, \quad (15.17)$$

en donde  $\Delta_i$  es una manera abreviada de escribir  $\Delta x_i$ .

Una variación en una coordenada en cantidad  $\Delta x_i$  tendrá que ser compensada, de manera que se siga cumpliendo (15.17), con una variación en todas las coordenadas, incluyendo la propia  $x_i$ , de acuerdo a la libertad del sistema que estemos considerando. Esta variación será, en primer orden, proporcional para cada coordenada a la magnitud de la corrección (es decir, a la magnitud de la traslación, escalado, giro... ). Sean  $\alpha_1, \alpha_2, \dots, \alpha_n$  dichas constantes de proporcionalidad, de manera que un ajuste de magnitud  $\varepsilon$  produzca una variación sobre las coordenadas de

$$-\alpha_1\varepsilon, -\alpha_2\varepsilon, \dots, -\alpha_n\varepsilon.$$

La variación en la condición (15.17) será entonces

$$-\varepsilon(k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n),$$

de donde se calcula  $\varepsilon$  para que la variación total sea cero:

$$k_i\Delta x_i - \varepsilon(k_1\alpha_1 + k_2\alpha_2 + \dots + k_n\alpha_n) = 0.$$

Se obtiene

$$\varepsilon = \frac{k_1\Delta x_i}{\sum \alpha_i k_i}. \quad (15.18)$$

Dado que los valores de  $k$  se pueden multiplicar todos por una constante arbitraria podemos elegirlos de manera que  $\sum \alpha_i k_i = 1$ . Si además recordamos que, en el caso más general, todas estas expresiones no son más que aproximaciones de primer orden, escribimos

$$\frac{\partial \varepsilon}{\partial \Delta x_i} = k_i.$$

La variación en cada coordenada es, en primer orden, el valor de  $\varepsilon$  multiplicado por su correspondiente  $-\alpha$ :

$$\frac{\partial \Delta_j}{\partial \Delta x_i} = -\alpha_j k_i. \quad (15.19)$$

## 15 Cambio de los parámetros fijos, II: Global

Estos son los coeficientes de la matriz  $M$ . La matriz  $M$  se puede escribir entonces como

$$M = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{pmatrix} (k_1 \ k_2 \ \dots \ k_n). \quad (15.20)$$

Recordamos la expresión a minimizar:

$$\text{Tr}(A\Sigma A^T) = \text{Tr}[(I - M)\Sigma(I - M)^T] = \text{Tr}(I) - 2\text{Tr}(M\Sigma) + \text{Tr}(M\Sigma M^T).$$

Como  $\text{Tr}(I)$  es fijo hay que hacer mínimo

$$-2\text{Tr}(M\Sigma) + \text{Tr}(M\Sigma M^T) = -2\text{Tr}(M\Sigma) + \text{Tr}(\Sigma M^T M).$$

El producto  $M^T M$  queda

$$\begin{pmatrix} k_1 \\ k_2 \\ \dots \\ k_n \end{pmatrix} (\alpha_1 \ \alpha_2 \ \dots \ \alpha_n) \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{pmatrix} (k_1 \ k_2 \ \dots \ k_n) = \\ = (\sum \alpha^2) \begin{pmatrix} k_1 \\ k_2 \\ \dots \\ k_n \end{pmatrix} (k_1 \ k_2 \ \dots \ k_n).$$

De modo que la traza a minimizar es

$$-2\text{Tr}((\alpha_i k_j)(\sigma_{ij})) + (\sum \alpha^2) \text{Tr}((\sigma_{ij})(k_i k_j)).$$

Llamamos  $a$  a la cantidad  $\sum \alpha^2$ . Tras operar queda la suma de las siguientes cantidades:

$$\begin{aligned} & (ak_1 - 2\alpha_1) (k_1\sigma_{11} + k_2\sigma_{12} + \dots + k_n\sigma_{1n}) \\ & (ak_2 - 2\alpha_2) (k_1\sigma_{21} + k_2\sigma_{22} + \dots + k_n\sigma_{2n}) \\ & \dots \\ & (ak_n - 2\alpha_n) (k_1\sigma_{n1} + k_2\sigma_{n2} + \dots + k_n\sigma_{nn}). \end{aligned}$$

## 15 Cambio de los parámetros fijos, II: Global

Esta suma es igual que la que apareció en la minimización de la traslación, salvo que allí se tenía "1" en lugar de cada  $\alpha_i$  y  $n$  en lugar de  $a$ . Si en la suma de entonces consideramos el primer punto repetido  $\alpha_1$  veces, el segundo  $\alpha_2$  veces, etc., y volvemos a calcular la traza, aparecerán  $\alpha_1$  líneas iguales, después  $\alpha_2$  líneas iguales, etc., y en cada una de ellas, en el paréntesis "grande" las cantidades  $k_1\sigma_{i1}$ ,  $k_2\sigma_{i2}$ , ...  $k_n\sigma_{in}$  aparecen respectivamente  $\alpha_1$ ,  $\alpha_2$ , ...  $\alpha_n$  veces. Además, el número total de puntos es ahora  $\sum \alpha$  en lugar de  $n$ , cantidad a la que llamaremos  $A$ . De modo que la suma quedaría

$$\begin{aligned} & (A\alpha_1k_1 - 2\alpha_1) (\alpha_1k_1\sigma_{11} + \alpha_2k_2\sigma_{12} + \dots + \alpha_nk_n\sigma_{1n}) \\ & (A\alpha_2k_2 - 2\alpha_2) (\alpha_1k_1\sigma_{21} + \alpha_2k_2\sigma_{22} + \dots + \alpha_nk_n\sigma_{2n}) \\ & \dots \\ & (A\alpha_nk_n - 2\alpha_n) (\alpha_1k_1\sigma_{n1} + \alpha_2k_2\sigma_{n2} + \dots + \alpha_nk_n\sigma_{nn}). \end{aligned}$$

Dado que el óptimo es que todos los valores de  $k$  sean iguales a  $1/A$ , los valores óptimos para las variables  $\alpha_i k_i$  son  $\alpha_i/A$ . Es decir, que el óptimo (mínimo) para

$$\begin{aligned} & (Ak_1 - 2\alpha_1) (k_1\sigma_{11} + k_2\sigma_{12} + \dots + k_n\sigma_{1n}) \\ & (Ak_2 - 2\alpha_2) (k_1\sigma_{21} + k_2\sigma_{22} + \dots + k_n\sigma_{2n}) \\ & \dots \\ & (Ak_n - 2\alpha_n) (k_1\sigma_{n1} + k_2\sigma_{n2} + \dots + k_n\sigma_{nn}). \end{aligned}$$

se alcanza si cada  $k_i$  es igual a  $\alpha_i/A$ .

Si escribimos  $k_i \rightarrow Bk'_i$  para una constante cualquiera, es decir, realizamos un cambio de variable, los valores óptimos para los nuevos  $k'_i$  serán los antiguos divididos por  $B$ , es decir,  $\alpha_i/(AB)$ . Pero mediante este cambio de variable cada línea de la suma queda multiplicada por un factor  $B$ , como resultado de llevar a cabo la sustitución en los valores de  $k$  segundo paréntesis, y en lugar de  $A$  aparece  $AB$ , a lo que llamaremos  $C$ . Multiplicar todas las líneas por una constante no afecta a qué valores de  $k$  hacen mínima la suma, así que la constante  $B$  multiplicando a todo podemos ignorarla, y hemos llegado a que los

## 15 Cambio de los parámetros fijos, II: Global

valores de  $k_i$  (volviendo a escribir  $k_i$  en lugar de  $k'_i$ ) que hacen mínimo

$$\begin{aligned} & (Ck_1 - 2\alpha_1) (k_1\sigma_{11} + k_2\sigma_{12} + \dots + k_n\sigma_{1n}) \\ & (Ck_2 - 2\alpha_2) (k_1\sigma_{21} + k_2\sigma_{22} + \dots + k_n\sigma_{2n}) \\ & \dots \\ & (Ck_n - 2\alpha_n) (k_1\sigma_{n1} + k_2\sigma_{n2} + \dots + k_n\sigma_{nn}). \end{aligned}$$

son  $\alpha_i/C$ , valga lo que valga  $C$ .

En el caso que nos ocupa  $C$  es igual a  $a = \sum \alpha^2$ , de modo que la solución es

$$k_i = \frac{\alpha_i}{\sum \alpha^2}. \quad (15.21)$$

Con estos valores de  $k$  se cumple  $\sum \alpha_i k_i = 1$ , como tenía que ser pues ya hemos supuesto esta relación en un punto anterior del desarrollo.

Alternativamente al razonamiento anterior se puede volver a minimizar la suma sin más que llevar a cabo una demostración paralela a la efectuada en su momento para la traslación: Si cada  $k_i$  es igual a  $\alpha_i/a$  la suma vale

$$-\frac{1}{a} \sum \alpha_i \alpha_j \sigma_{ij}, \quad (15.22)$$

mientras que si cada uno de ellos es igual a  $\alpha_i + \delta_i$  la suma anterior se incrementa en

$$a \sum \delta_i \delta_j \sigma_{ij},$$

que es una cantidad positiva, pues es la varianza de la variable  $a \sum \delta_i x_i$ .

La cantidad (15.22) es  $a$  veces la varianza de la variable  $(\alpha_1 x_1 + \dots + \alpha_n x_n)/a$  o bien  $1/a$  veces la de la variable  $(\alpha_1 x_1 + \dots + \alpha_n x_n)$ . Si volvemos a minimizar las varianzas de acuerdo a la libertad que estamos considerando, las varianzas quedarán como están, pues ya estábamos en el mínimo. Es decir, que esta cantidad que se resta se hace cero. Esto significa que si las varianzas de las variables  $x_1, x_2, \dots, x_n$  ya están reducidas a su mínimo posible de acuerdo a la libertad del sistema la varianza de aquella cantidad es cero, lo que equivale a decir que esa cantidad tiene un valor fijo. Es decir, la interpretación geométrica de la solución óptima de acuerdo a la libertad considerada es que el sistema queda fijado definiendo un valor para la cantidad

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n. \quad (15.23)$$

## 15 Cambio de los parámetros fijos, II: Global

La matriz A de cambio de varianzas es

$$A = I - \frac{1}{\sum \alpha^2} \begin{pmatrix} \alpha_1 \alpha_1 & \dots & \alpha_1 \alpha_n \\ \dots & \dots & \dots \\ \alpha_n \alpha_1 & \dots & \alpha_n \alpha_n \end{pmatrix}. \quad (15.24)$$

En la ecuación (15.23) se pueden multiplicar todas las constantes  $\alpha$  por una cantidad arbitraria, ya que es lo mismo decir que se fija un valor para esa cantidad que fijarlo para cualquier múltiplo suyo. Podemos entonces tomar los valores de los  $\alpha_i$  de manera que  $\sum \alpha_i^2 = 1$ , con lo que ese término se elimina de (15.24).

Es digno de señalar que la solución no depende de los valores de la matriz  $\Sigma$ .

La aplicación de esto a los casos de escala libre o giro libre es inmediata. Si la libertad consiste en poder aplicar un factor de escala respecto al centro de gravedad o respecto a cualquier otro punto, la variación de cada coordenada es proporcional a sí misma, expresadas las coordenadas respecto al punto origen del escalado. Por lo tanto  $\alpha_{x_i} = x_i$  y  $\alpha_{y_i} = y_i$ , lo que da la matriz ya conocida para cuando la escala es libre, con la ventaja de que ahora hemos demostrado que esta matriz también es válida cuando el origen del escalado es otro punto distinto del centro de gravedad.

Si la libertad consiste en poder aplicar un giro respecto a un punto las variaciones proporcionales de cada coordenada son  $\alpha_{x_i} = -y_i$ ,  $\alpha_{y_i} = x_i$ , lo que da lugar a la matriz ya encontrada para el giro.

La interpretación geométrica de la definición del sistema de coordenadas en uno y otro caso la obtenemos de (15.23). Si llamamos  $x_0, y_0$  a las coordenadas que tienen los puntos, y mediante  $x, y$  denotamos esas mismas magnitudes entendidas como variables (por ejemplo, si volvemos a medir el conjunto de puntos los valores  $(x, y)$  cambiarán mientras que los  $x_0$  e  $y_0$  siguen siendo los mismos) entonces las definiciones de los sistemas en la elección óptima para los casos de escala y giro libres son que

$$\sum (x_0 x + y_0 y),$$

$$\sum (-y_0 x + x_0 y)$$

tengan un valor dado. Este valor será obviamente el que tienen cuando las coordenadas son iguales a  $(x_0, y_0)$ , esto es,  $\sum r_0^2$  para la escala y 0 para el giro.

## 15 Cambio de los parámetros fijos, II: Global

Estas interpretaciones son las mismas que obtuvimos unas secciones atrás según el sentido común. En el caso de la escala es ligeramente distinta, pues era  $\sum(xx + yy) = r_0^2$ . Si escribimos  $x = x_0 + \delta_x$ , anál.  $y$ , ambos sumatorios son

$$\begin{aligned}\sum(x_0x + y_0y) &= \sum(x_0^2 + y_0^2) + \sum(x_0\delta_x + y_0\delta_y), \\ \sum(xx + yy) &= \sum(x_0^2 + y_0^2) + 2\sum(x_0\delta_x + y_0\delta_y) + \sum(\delta_x^2 + \delta_y^2),\end{aligned}$$

por lo que las condiciones equivalen a

$$\begin{aligned}\sum(x_0\delta_x + y_0\delta_y) &= 0, \\ 2\sum(x_0\delta_x + y_0\delta_y) + \sum(\delta_x^2 + \delta_y^2) &= 0.\end{aligned}$$

Ambas condiciones son iguales en primer orden, y dado que todo el desarrollo se basa en el estudio de las varianzas de unas variables en función de otras en aproximación lineal, ambas condiciones son iguales al efecto.

### Extensión a otras transformaciones

En todo el capítulo excepto en la última sección se han analizado exclusivamente precisiones en las coordenadas  $x, y$  de un conjunto de puntos que puede admitir traslación, rotación y escalado. El caso de la traslación se puede aplicar a cualquier situación en la que no exista un cero absoluto y los parámetros se pueden desplazar libremente, o bien existiendo ese cero estamos interesados en las precisiones en relación a un punto; en ningún momento de los desarrollos hicimos uso del significado de los parámetros « $x$ » e « $y$ ». En particular, se puede aplicar a una red altimétrica o más generalmente a las coordenadas  $z$  de un conjunto de puntos.

En cuanto a transformaciones más complejas, el resultado de la última sección resuelve todos los casos posibles siempre que sean linealizables.

Si la libertad de que goza el sistema se puede expresar en función de más de un parámetro, por ejemplo escala y giro a la vez, si el producto de las matrices  $M$  deducidas como óptimas para cada uno de los parámetros por separado es cero, lo que equivale a  $\sum \alpha_i \alpha_i' = 0$

## 15 Cambio de los parámetros fijos, II: Global

en donde los primeros  $\alpha_i$  son los de un parámetro y los segundos,  $\alpha'_i$ , los de otro, para cada posible pareja de parámetros, entonces esta descomposición de las libertades es ortogonal y la optimización por separado de cada una de ellas da lugar al óptimo conjunto. Esto es lo que sucedió para la traslación, escalado y giro. Si por el contrario esto no es así hay que llevar a cabo una optimización conjunta. Se resuelve en el capítulo «Interpretación geométrica».

Para calcular las nuevas precisiones a partir de las antiguas se necesita la matriz de estas últimas:  $\Sigma_{xx}$ , pero en ningún momento hemos supuesto que estas variables siguen una distribución normal. Lo visto en este capítulo se aplica pues a variables que pueden seguir cualquier distribución con la única diferencia de que la matriz  $\Sigma_{xx}$  no vendrá dada por la  $N^{-1}$  de un planteamiento mínimo cuadrático.

# 16 Estimación de la varianza

## Introducción

La propia desviación típica de un conjunto de observaciones puede considerarse un parámetro a estimar. En un primer momento parece que esto complica mucho el proceso de estimación, ya que estamos considerando como incógnita la propia distribución de los errores. Sin embargo un pequeño acercamiento al problema basta para ver que no cambia nada. Hasta ahora teníamos

$$g(x_1, \dots, x_n) \propto f_{x_1, \dots, x_n}(\ell_1, \dots, \ell_n)h(x_1, \dots, x_n).$$

La función  $f$  que siguen las observaciones depende del valor real de los parámetros, y por eso escribimos  $f_{x_1, \dots, x_n}(\ell_1, \dots, \ell_n)$ . Entonces los parámetros  $(x_1, \dots, x_n)$  no se diferencian en nada de  $\sigma$  o de cualquier otro parámetro que describa la distribución. Son, todos, valores en función de los cuales varía  $f(\ell_1, \dots, \ell_n)$ .

Hasta ahora no hemos empleado ningún tipo de letra particular para las desviaciones típicas. A partir de ahora, puesto que serán valores desconocidos que estimamos, emplearemos  $\sigma$  para los valores reales y  $\sigma$  para los ajustados. Todas las  $\sigma$  vistas hasta ahora eran siempre valores conocidos  $\sigma$ .

También hasta ahora multiplicábamos todas las ecuaciones de residuo por  $\sigma_0/\sigma_i$ , en donde  $\sigma_0$  era una constante cualquiera, obteniendo unos residuos  $v'$  cuyas observaciones  $\ell'$  seguían distribuciones normales con  $\sigma = \sigma_0$ . Ahora habrá en el ajuste, igual que antes, varios grupos de observaciones del mismo tipo (con un mismo  $\sigma_i$ ), de

## 16 Estimación de la varianza

desviaciones típicas  $\sigma_\alpha, \sigma_\beta, \dots, \sigma_\lambda$ , pero ahora consideramos esos valores como desconocidos. El caso más sencillo es aquél en el que existe un sólo conjunto de observaciones o bien damos por conocidas las relaciones de proporcionalidad entre los distintos  $\sigma_i$ , es decir, los cocientes  $\sigma_i/\sigma_j$ , aunque no conozcamos ningún  $\sigma_i$ . En estos casos podemos multiplicar cada ecuación por una cantidad proporcional al inverso de su desviación típica, y definimos  $\sigma_0$  como la constante de proporcionalidad, es decir, el producto del valor que multiplica a una ecuación por la desviación típica de su observación. De esta manera  $\sigma_0$  es un valor real desconocido,  $\sigma_0$ . Una definición equivalente de  $\sigma_0$  es la desviación típica que tendría una observación cuya ecuación de residuo multiplicaríamos por 1, es decir, a la que asignaríamos peso unidad. Si hemos dado peso unidad a las observaciones de un grupo con  $\sigma = \sigma_\alpha$ , entonces la estimación de  $\sigma_0$  coincide con la de  $\sigma_\alpha$ . Estudiaremos primero el caso de un  $\sigma_0$  único, y el caso general lo veremos sólo al final del capítulo. Las cantidades  $v$  y  $\varepsilon$  que aparezcan serán siempre normalizadas.

$\sigma_0$  es un parámetro que puede ser cualquiera de los  $x_i$ , por ejemplo  $x_n$ , pero no emplearemos esta notación, sino que consideraremos  $\sigma_0$  un parámetro a parte de los  $x_i$ . De esta manera el número  $n$  es el mismo estimemos o no  $\sigma_0$ .

### **Función $g(\sigma_0^2)$**

La estimación de  $\sigma_0$  es bastante distinta que la del resto de parámetros. En primer lugar, no es una magnitud de las grandes, sino de aquellas en las que su variabilidad es comparable al propio parámetro. El resultado puede variar significativamente según la función  $h$  que consideremos. Para la desviación típica, no saber nada se interpreta como que se considera igual de probable un valor entre 10 y 12, por ejemplo, que uno entre 20 y 24, o entre 100 y 120. Esto sucede con las variables que tienen un significado análogo a un factor de escala. Entonces la función es

$$h(\sigma_0) = 1/\sigma_0. \quad (16.1)$$

Este capítulo podría haberse llamado estimación de la desviación típica, pero se llama intencionadamente estimación de la varianza. Buscaremos estimadores centrados, función  $g$  e intervalos de confianza para la variable  $\sigma_0^2$ , y no para  $\sigma_0$ . Los estimadores centrados son

## 16 Estimación de la varianza

distintos según respecto a cuál de las dos lo consideremos centrado (si  $x$  es una variable cualquiera, no es lo mismo  $E[x^2]$  que  $E[x]^2$ ). La razón para considerar  $\sigma_0^2$  es que esta cantidad se define como  $E[(x - \mu_x)^2]$ , es decir, la esperanza se toma sobre los cuadrados, y el resultado es la varianza, mientras que la desviación típica es la raíz cuadrada de la varianza, que no es la esperanza de nada. Por esta razón la varianza cumple ciertas propiedades relativas a estimaciones centradas que no cumple la desviación típica. Por ejemplo, si se suman dos variables aleatorias independientes cualesquiera (no necesariamente normales), de varianzas  $\sigma_1^2$  y  $\sigma_2^2$ , la varianza de la variable que resulta es  $\sigma_1^2 + \sigma_2^2$ , de manera que, si para cada una de las distribuciones que se suman tenemos una estimación centrada para su varianza,  $\sigma_1^2$  y  $\sigma_2^2$ , entonces el valor  $\sigma_1^2 + \sigma_2^2$  es una estimación centrada de la varianza de la variable suma, mientras que la propiedad análoga para desviaciones típicas no se cumple. Otra propiedad es que si  $(x_1, \dots, x_k)$  son variables aleatorias de media cero y  $(c_1, \dots, c_k)$  constantes cualesquiera distintas de cero, la variable  $x_1^2/c_1 + \dots + x_k^2/c_k$  tiene una media que podemos estimar de manera centrada mediante  $\sigma_1^2/c_1 + \dots + \sigma_k^2/c_k$ . En particular podemos tomar  $c_1 = \sigma_1^2, \dots, c_k = \sigma_k^2$  y entonces la estimación centrada de la media es  $k$ .

Para trabajar con  $\sigma_0^2$  primero tenemos que hallar la distribución  $h(\sigma_0^2)$  partiendo de la de  $\sigma_0$ . Puede demostrarse que si  $h(\sigma_0) = 1/\sigma_0$  entonces  $h(\sigma_0^2) = 1/\sigma_0^2$ . Emplear una variable cuya representación escrita incluye  $^2$  da lugar a confusión. Por ejemplo,  $\int \sigma_0^2 d\sigma_0^2$  es igual a  $\sigma_0^4/2$ , ya que la variable es  $\sigma_0^2$ . Para evitarlo emplearemos la letra  $t$  como sinónimo de  $\sigma_0^2$ . Ahora se ve claro que  $\int t dt = t^2/2$ . Con esta notación,

$$h(\sigma_0^2) = \frac{1}{t}.$$

La función  $g$  según esta función  $h$  es

$$g(x_1, \dots, x_n, \sigma_0^2) \propto \frac{1}{t} \frac{1}{t^{m/2}} e^{-\frac{\sum \epsilon^2}{2t}} = \frac{1}{t^{\frac{m}{2}+1}} e^{-\frac{\sum \epsilon^2}{2t}}. \quad (16.2)$$

Si fijamos un valor para  $\sigma_0$ , el problema de estimar los parámetros  $(x_1, \dots, x_n)$  es la estimación mínimo cuadrática ya estudiada, cuyo resultado es independiente de  $\sigma_0$ . Por lo tanto el hecho de estimar  $\sigma_0^2$  no modifica en nada la estimación de  $X$ , y la solución mínimo cuadrática

## 16 Estimación de la varianza

ca sigue cumpliendo las mismas propiedades (máxima probabilidad, centrada directa y centrada inversa).

El valor de  $\sigma_0^2$  que maximiza la probabilidad es distinto de la estimación centrada, y como debe hacerse en estos casos se toma esta última. La estimación de máxima probabilidad la obtenemos fácilmente derivando  $g$  respecto de  $t$ , y se obtiene

$$\sigma_0^2 = \frac{\sum v^2}{m+2}.$$

Este resultado es el que, junto con los valores  $(x_1, \dots, x_n)$  de la estimación mínimo cuadrática, hace máximo el valor de  $g$ , por lo que las regiones  $(n+1)$ -dimensionales del espacio de puntos  $(x_1, \dots, x_n, \sigma_0^2)$  serán más pequeñas en torno a ese punto, para una misma probabilidad. Pero no estamos interesados en esas regiones conjuntas, sino en regiones independientes, por un lado para  $X_r$  y por otro para  $\sigma_0^2$ . El punto que maximiza  $g(\sigma_0^2)$  será otro.

Para obtener la estimación centrada, y también la de máxima probabilidad independiente de los demás parámetros, debemos primero hallar la función  $g(\sigma_0^2)$ . Para ello tenemos que calcular qué «área»  $n$ -dimensional tiene la función (16.2) en cada corte por un valor de  $t$ .

$$g(t) \propto \int_E \frac{1}{t^{\frac{m}{2}+1}} e^{-\frac{\sum \varepsilon^2}{2t}} dE.$$

El conjunto de posibles  $(\varepsilon_1, \dots, \varepsilon_m)$  es  $n$ -dimensional, y admitiendo la aproximación lineal es un subespacio de  $\mathbb{R}^m$ . El subespacio no pasa por el origen en general. Supongamos que sí pase por el origen. Entonces la función  $\exp(-\sum \varepsilon^2/(2t))$  es, salvo por una constante, una distribución normal multidimensional de dimensión  $n$ . La dimensión es  $n$  y no  $m$  ya que no todos los valores  $(\varepsilon_1, \dots, \varepsilon_n)$  son posibles, sino sólo un subespacio de dimensión  $n$ . Si quisiésemos podríamos buscar  $n$  variables combinaciones lineales de los  $\varepsilon$  y que estuviesen contenidas en el subespacio, de manera que el conjunto de posibles  $E$  vendría dado por cualquier combinación de las  $n$  nuevas variables. La integral de la exponencial de una normal multidimensional ya la hemos obtenido, y es

$$\int_E e^{-\frac{\sum \varepsilon^2}{2t}} dE = (2\pi)^{n/2} t^{n/2}.$$

## 16 Estimación de la varianza

Pero en general el subespacio  $E$  no pasa por el origen. Si tomamos un posible  $E$  cualquiera,  $E_0$ , podemos escribir  $E = E_0 + (E - E_0)$ ; es decir, una constante más un subespacio que pasa por el origen. Pero al elevar al cuadrado,

$$E^2 = E_0^2 + (E - E_0)^2 + 2E_0(E - E_0),$$

y aparece el término  $2E_0(E - E_0) = 2\sum \varepsilon_0(\varepsilon - \varepsilon_0)$ . Sin embargo, si tomamos  $E_0 = V$  ese término desaparece. En efecto,  $V$  es el punto de  $E$  para el cual  $\sum \varepsilon^2$  es mínimo, lo que significa que es el punto del subespacio  $E$  más próximo al origen, y en consecuencia el vector  $V$  es perpendicular al subespacio, lo que a su vez quiere decir que  $V$  es perpendicular a cualquier  $E - V$ , que expresado en función de las componentes de ambos vectores es  $\sum v(\varepsilon - v) = 0$ . Por lo tanto,

$$E^2 = V^2 + (E - V)^2. \quad (16.3)$$

De esta relación,

$$e^{-\frac{\sum \varepsilon^2}{2t}} = e^{-\frac{\sum v^2}{2t}} e^{-\frac{\sum (\varepsilon - v)^2}{2t}},$$

$$\int_E e^{-\frac{\sum \varepsilon^2}{2t}} = e^{-\frac{\sum v^2}{2t}} (2\pi)^{n/2} t^{n/2}.$$

Y entonces

$$g(t) \propto \frac{1}{t^{\frac{m}{2}+1}} e^{-\frac{\sum v^2}{2t}} t^{n/2} = \frac{1}{t^{\frac{m-n}{2}+1}} e^{-\frac{\sum v^2}{2t}}. \quad (16.4)$$

### Estimador de máxima probabilidad y centrado

El estimador de máxima probabilidad se obtiene derivando:

$$\sigma_0^2 = \frac{\sum v^2}{m - n + 2}.$$

El estimador centrado para  $\sigma_0^2$  es

$$E[\sigma_0^2] = \frac{\int_0^\infty \frac{t}{t^{\frac{n-m}{2}+1}} e^{-\frac{\sum v^2}{2t}} dt}{\int_0^\infty \frac{1}{t^{\frac{n-m}{2}+1}} e^{-\frac{\sum v^2}{2t}} dt}.$$

## 16 Estimación de la varianza

Estas integrales son fáciles de calcular. Hacemos el cambio de variable  $\sum v^2/(2t) = z$ , y entonces

$$\frac{1}{t} = \frac{2}{\sum v^2} z;$$

$$-\frac{\sum v^2}{2t^2} dt = dz \quad \Rightarrow \quad \frac{1}{t^2} dt = -\frac{2}{\sum v^2} dz.$$

Los límites de integración se invierten. Ahora bien, para una función cualquiera  $f$  se tiene que  $\int_a^b f = -\int_b^a f$ , por lo tanto podemos mantener los límites como están si cambiamos el signo, y el signo menos que aparece se anula con el que surge al sustituir  $dt$ , y finalmente queda

$$E[\sigma_0^2] = \frac{\int_0^\infty z^{\frac{m-n}{2}-2} e^{-z} dz}{\frac{2}{\sum v^2} \int_0^\infty z^{\frac{m-n}{2}-1} e^{-z} dz}.$$

Estas integrales son de la forma  $\int_0^\infty z^x e^{-z} dz$ , que llamaremos  $\Pi(x)$ . Integrando por partes se demuestra que  $\Pi(x) = x\Pi(x-1)$ , si  $x > 0$ , y entonces

$$E[\sigma_0^2] = \frac{1}{\frac{2}{\sum v^2} \left(\frac{m-n}{2} - 1\right)} = \frac{\sum v^2}{m-n-2}, \quad m-n > 2.$$

Si  $r = m-n = 2$  entonces  $E[\sigma_0^2] = \infty$ , y con más razón si  $r = 1$ . Finalmente,

$$\sigma_0^2 = E[\sigma_0^2] = \begin{cases} \frac{\sum v^2}{r-2}, & r > 2. \\ +\infty, & r \leq 2. \end{cases} \quad (16.5)$$

### Intervalos de confianza para la desviación típica

El paso siguiente es obtener intervalos de confianza para  $\sigma_0^2$ , o lo que es lo mismo, para  $\sigma_0$ , que serán los que queramos. Para ello ya no es suficiente la proporcionalidad (16.4), sino que hay que hallar la constante exacta. Al hacerlo aparece un término  $(\sum v^2)^r$  que es bastante engorroso. Para simplificar las fórmulas calcularemos los intervalos para la variable  $z = \sum v^2/t$ . Si queremos la probabilidad

## 16 Estimación de la varianza

para un intervalo de  $\sigma_0$ , digamos  $[a, b]$ , entonces el intervalo para  $z$  es  $[\sum v^2/b^2, \sum v^2/a^2]$ , y calculamos esta probabilidad con la expresión que obtenemos a continuación.

La función  $g(z)$  es

$$g(z) \propto z^{\frac{r}{2}-1} e^{-z/2}.$$

La integral de esta función es  $2(r/2 - 1)!$ , así que

$$g(z) = \frac{z^{\frac{r}{2}-1}}{2(\frac{r}{2}-1)!} e^{-z/2} = \chi_r^2(z).$$

Por lo tanto la variable  $\sum v^2/\sigma_0^2$  sigue una distribución  $\chi_r^2$ .

Los intervalos de confianza para la desviación típica serán de la forma  $[\sigma_0/h, h\sigma_0]$ , que además dan lugar a unos límites para el intervalo en  $z$  particularmente sencillos:

$$\frac{\sum v^2}{h^2\sigma_0^2} = \frac{r-2}{h^2},$$

y el intervalo en  $z$  es  $[\frac{r-2}{h^2}, h^2(r-2)]$ , y por lo tanto

$$P(\sigma_0 \in [\sigma_0/h, h\sigma_0]) = \chi_r^2(h^2(r-2)) - \chi_r^2\left(\frac{r-2}{h^2}\right). \quad (16.6)$$

### Estimación centrada geométrica

Por ser  $\sigma_0$  una variable de tipo escala se podría buscar un estimador centrado en el sentido de la media geométrica. La esperanza de una variable, en este caso de  $\sigma_0^2$ , tiene la siguiente interpretación: si realizamos muchos ajustes y vamos tomando aquéllos en los que se obtienen siempre los mismos valores observados (si eso fuese posible), y para esos ajustes realizamos la media aritmética de los valores verdaderos  $\sigma_0^2$ , el límite de esa media es  $E[\sigma_0^2]$ . Al operador análogo a la esperanza pero referido a la media geométrica podemos llamarlo  $G$ , y el valor es  $G[\sigma_0^2]$ . Con este estimador sí que es indiferente trabajar con  $\sigma_0^2$  o con  $\sigma_0$ , ya que  $G[\sigma_0^2] = (G[\sigma_0])^2$ . Podemos abordar el problema de su obtención de dos maneras distintas. Por un lado mediante la definición de la media geométrica como límite,

## 16 Estimación de la varianza

$$G[\sigma_0^2] = \lim_{p \rightarrow 0} \left( \frac{\int_0^\infty t^p g(t) dt}{\int_0^\infty g(t) dt} \right)^{1/p},$$

o bien trabajando con la variable  $\log(t)$ , y  $G[\sigma_0^2] = \exp(E[\log(t)])$ . En ambos casos se llega a las mismas integrales. El valor exacto viene dado por una expresión un tanto complicada, pero es prácticamente igual a

$$G[\sigma_0^2] = \frac{\sum v^2}{r-1}.$$

De todos modos, el mismo razonamiento por el cual tomamos  $E[\sigma_0^2]$  en lugar de  $E[\sigma_0]$  aconseja la elección de  $E[\sigma_0^2]$  mejor que  $G[\sigma_0^2]$ .

### Precisión en la estimación de $\sigma_0$

Para valores pequeños de  $r$  la precisión con la que se estima  $\sigma_0$  es muy mala. Así, para  $r = 5$  si queremos el intervalo del 90% de probabilidad tenemos que tomar  $h = 1,84$ . A continuación se muestran distintos valores de  $r$  y los valores de  $h$  necesarios en cada caso para obtener el intervalo del 90%.

$r$	5	10	20	30	50	100	200
$h$ (90%)	1,84	1,48	1,31	1,24	1,18	1,124	1,086

A la vista de estos valores concluimos que no se debe estimar  $\sigma_0$  si  $r < 30$ , salvo que el conocimiento que tengamos a priori sea tan impreciso que el valor estimado sea mejor. No debemos emplear la estimación nunca por debajo de  $r = 10$  (si el conocimiento a priori de  $\sigma_0$  es tan malo que aun así la estimación en estos casos es más precisa, entonces tenemos un problema).

### Modificación de las regiones de confianza para las magnitudes estimadas

El conocimiento con mayor o menor precisión de  $\sigma_0$  también afecta a las regiones de confianza para todos los valores ajustados vistos  $(x, l$  y  $v)$ . Recordamos que las obtuvimos a partir de la distribución que seguían los valores reales estimados, que vimos que era siempre

## 16 Estimación de la varianza

una normal multidimensional cuya matriz  $\Sigma$  era siempre la constante  $\sigma_0^2$  multiplicada por otra matriz; por ejemplo,  $\Sigma_{xx} = \sigma_0^2 N^{-1}$ . Ahora  $\sigma_0^2$  ya no es conocido, sino que sigue una cierta distribución. Por la definición de  $\sigma^2$  de una variable cualquiera como una esperanza, y por ser  $\sigma_0^2$  una estimación centrada, se cumple  $\Sigma_{xx} = \sigma_0^2 N^{-1}$ , y análogamente para las demás matrices. Sin embargo  $\Sigma_{xx}$ , etc., son matrices «directas», es decir, no son estimaciones, y siguen siendo  $\sigma_0^2 N^{-1}$ , etc. El desconocimiento de  $\sigma_0^2$ , y en consecuencia de las matrices  $\Sigma_{xx}$ , etc., no importa mucho, ya que estas matrices no tienen ningún interés tras el ajuste. Tienen interés en la fase de diseño, cuando se estudia el problema en sentido directo, para ver cuáles son las probabilidades de comisión de errores en los valores ajustados, y se trabajará (si es que existe la fase de diseño) con un valor concreto de  $\sigma_0$ .

$\sigma_0^2$  será igual a  $c\sigma_0^2$ , para una constante  $c$  desconocida, y también  $\sigma^2 = c\sigma^2$  para cualquier varianza estimada. Si conociésemos el valor de  $c$  sabríamos que la probabilidad de que los valores reales de un conjunto de  $d$  magnitudes estimadas estuviesen dentro de la elipse (o su equivalente  $d$ -dimensional) de semiejes  $k\sigma_i$ , para  $i = 1, \dots, d$ , sería  $\chi_d^2(k^2/c)$ .

El valor  $c = \sigma_0^2/\sigma_0^2$  sigue una distribución que podemos obtener a partir de la de  $\sigma_0^2$ . Sin embargo las expresiones serán más sencillas si definimos  $c$  como el inverso de ese valor, es decir,

$$c = \frac{\sigma_0^2}{\sigma_0^2} = \frac{\sum v^2}{(r-2)\sigma_0^2} = \frac{z}{r-2}.$$

Sabiendo que  $z \sim \chi_r^2$  se puede demostrar que

$$g(c) = (r-2)\chi_r^2((r-2)c),$$

y entonces la probabilidad para la región  $d$ -dimensional de semiejes  $k\sigma_i$  es

$$P_d(k) = \int_0^\infty \chi_d^2(ck^2)(r-2)\chi_r^2((r-2)c) dc. \quad (16.7)$$

Se cumple, como no podía ser de otra manera,  $\lim_{r \rightarrow \infty} P_d(k) = \chi_d^2(k^2)$ .

Normalmente estaremos interesados en  $d = 1$  (intervalos) y  $d = 2$  (elipses). Para  $d = 1$  la función  $P_d(k)$  está relacionada con la  $t$  de

## 16 Estimación de la varianza

Student de  $r$  grados de libertad:

$$P_1(k) = \int_{-k}^k t_r(x) dx, \quad k = \sqrt{\frac{r}{r-2}} k; \quad (16.8)$$

y para  $d = 2$  se puede demostrar que

$$P_2(k) = 1 - \left(1 + \frac{k^2}{r-2}\right)^{-r/2}. \quad (16.9)$$

A continuación mostramos las probabilidades  $P_1(k)$  y  $P_2(k)$  en tanto por ciento para distintos valores de  $r$ , así como el valor de  $k$  para el cual se tiene una probabilidad del 95,5 %.

$d = 1$

$k \setminus r$	10	20	30	50	100	200	$\infty$
1	71,0	69,6	69,1	68,8	68,5	68,4	68,3
1,5	87,5	87,0	86,9	86,8	86,7	86,7	86,6
2	95,1	95,2	95,3	95,3	95,4	95,4	95,45
2,3	97,2	97,5	97,6	97,7	97,8	97,8	97,9
$k(95,5\%)$	2,05	2,03	2,02	2,01	2,01	2,01	2,00

$d = 2$

$k \setminus r$	10	20	30	50	100	200	$\infty$
1	44,5	41,8	40,9	40,3	39,8	39,6	39,3
1,5	71,0	69,2	68,6	68,2	67,9	67,7	67,5
2	86,8	86,6	86,5	86,5	86,5	86,5	86,5
2,5	94,4	94,9	95,1	95,3	95,5	95,5	95,6
2,8	96,7	97,3	97,5	97,7	97,9	97,9	98,0
$k(95,5\%)$	2,62	2,56	2,54	2,52	2,50	2,50	2,49

Vemos que los valores de  $k$  para la probabilidad del 95,5 % apenas varían. Podemos tomar siempre  $k = 2$  para  $d = 1$  y  $k = 2,5$  para  $d = 2$ , salvo tal vez para  $d = 2$ ,  $r = 10$ , en donde la diferencia ya puede ser apreciable.

En la fórmula (16.7) el factor  $X_d^2(ck^2)$  es el que corresponde al tipo de regiones cuya probabilidad queremos obtener. Si quisiésemos por

## 16 Estimación de la varianza

ejemplo la probabilidad en un círculo de radio  $k$ , y denotando por  $C(k)$  la probabilidad para un radio  $k$  si  $\sigma_0$  fuese conocida e igual a  $\sigma_0$ , entonces en lugar de  $\chi_d^2(ck^2)$  deberíamos emplear  $C(\sqrt{ck})$ .

### Combinación de varias estimaciones de la varianza

Es muy habitual realizar un ajuste con un cierto número de redundancias,  $r_1$ ,\* obteniendo una estimación  $\sigma_{01}^2$ , y disponer de otro valor  $\sigma_{02}^2$  de otro ajuste con  $r_2$  redundancias. En estos casos podemos combinar ambas estimaciones para obtener un nuevo valor, que equivale a una estimación de un ajuste con  $r_1 + r_2$  redundancias. Para poder obrar así en primer lugar las dos estimaciones tienen que serlo del mismo valor  $\sigma_0^2$ . Esto significa que en los dos ajustes tenemos que haber asignado peso 1 a observaciones con la misma desviación típica (la desconocida  $\sigma_0$ ). Esta condición se cumple, por ejemplo, si en ambos ajustes hemos asignado peso 1 a las observaciones de lectura horizontal, y en ese caso además  $\sigma_0 = \sigma_\alpha$ ; o si hemos dividido ambos conjuntos de ecuaciones por el mismo valor. En resumen, dos ajustes tienen el mismo  $\sigma_0$  si hemos aplicado los mismos pesos a ecuaciones correspondientes a observaciones de igual, aunque desconocida, desviación típica.

Dos ajustes independientes podemos pensarlos, si así nos conviene, como uno único, ya que nada cambia. Simplemente las matrices  $A$  y  $N$  tendrán bloques de ceros, y los bloques correspondientes a cada uno de los ajustes originales no se operan nunca entre sí. Con esta interpretación obtenemos inmediatamente la estimación centrada conjunta para  $\sigma_0^2$ :

$$\sigma_0^2 = E[\sigma_0^2] = \frac{\sum v^2}{r-2} = \frac{(r_1-2)\sigma_{01}^2 + (r_2-2)\sigma_{02}^2}{r_1 + r_2 - 2}.$$

Y en general si disponemos de  $k$  valores  $\sigma_{0i}^2$  de otros tantos ajustes, entonces obtenemos una nueva estimación de la forma

$$\sigma_0^2 = \frac{\sum (r_i - 2)\sigma_{0i}^2}{\sum r_i - 2}. \quad (16.10)$$

\*Estamos modificando ligeramente la notación. Ahora  $r_i$  no se refiere a la redundancia de la observación  $\ell_i$ , sino a la redundancia total del ajuste  $i$ .

## 16 Estimación de la varianza

Este valor  $\sigma_0^2$  es la nueva estimación que sustituye a las anteriores, y por tanto es el que se tomará para calcular las regiones de confianza, con un valor de la redundancia  $r = \sum r_i$ .

**Aclaración sobre la estimación**  $\sigma_0^2 = \sum v/r$

Muchas veces se estima  $\sigma_0^2$  como  $\sum v^2/r$ . Está tan extendido este método que debemos dar aquí una explicación, aun a riesgo de crear confusión donde no debería haberla. Este valor surge en el estudio directo, y por tanto no sirve de nada en la estimación. Hemos demostrado (cap. 10) que, para unos valores reales concretos, cada  $v_i$  sigue una distribución  $N(0, \sigma_{v_i})$ , siendo  $\sigma_{v_i}^2 = r_i \sigma_0^2$ , de modo que  $E[v_i^2] = r_i \sigma_0^2$ . Y dado que  $\sum r_i = r$ , entonces

$$\frac{E[\sum v_i^2]}{r} = \sigma_0^2. \quad (16.11)$$

Más aún, puede demostrarse que la variable  $\sum v^2/\sigma_0^2$  sigue una distribución  $\chi_r^2$ . Este resultado debe compararse con el obtenido en la p. 209, en donde demostramos que  $\sum v^2/\sigma_0^2 \sim \chi_r^2$ . Sin embargo son dos resultados totalmente distintos. En el de entonces la variable es  $\sigma_0^2$ , mientras que  $\sum v^2$  es una cantidad fija; hemos obtenido unos valores de  $v$  y no sabemos el valor de  $\sigma_0^2$ , pero sabemos que sus probabilidades vienen dadas por  $\sum v^2/\sigma_0^2 \sim \chi_r^2$ . En el resultado de ahora partimos de unos valores reales, por lo tanto  $\sigma_0^2$  es fijo, y expresamos que si realizamos muchos conjuntos de observaciones los conjuntos de residuos que se obtendrán cumplan  $\sum v^2/\sigma_0^2 \sim \chi_r^2$ . Dado que  $\sigma_0^2 = \sum v^2/(r-2)$ ,

$$\text{resultado directo:} \quad \sigma_0^2 \text{ cumple } (r-2) \frac{\sigma_0^2}{\sigma_0^2} \sim \chi_r^2. \quad (16.12)$$

$$\text{resultado inverso:} \quad \sigma_0^2 \text{ cumple } (r-2) \frac{\sigma_0^2}{\sigma_0^2} \sim \chi_r^2. \quad (16.13)$$

Es otro caso en el que para el estimador mínimo cuadrático las funciones de distribución directas e inversas casi coinciden (aunque no sería así si hubiésemos considerado otra función  $h$ ). Sin embargo hay

## 16 Estimación de la varianza

una diferencia mayor que en los otros casos casi coincidentes vistos hasta ahora. En el resultado directo la variable,  $\sigma_0^2$ , aparece en el numerador, así que salvo por la constante  $(r - 2)/\sigma_0^2$  es la función  $f(\sigma_0^2)$ . En el resultado inverso, sin embargo, la variable está en el denominador, de modo que no es en absoluto  $g(\sigma_0^2)$ . Para obtener esta última tendríamos que efectuar un cambio de variable, llegando a una expresión algo más complicada (en realidad es al revés: primero obtuvimos la función  $g(\sigma_0^2)$ , y mediante el cambio de variable  $z = \sum v^2/\sigma_0^2$  llegamos a  $z \sim \chi_r^2$ ). Esta diferencia conlleva que las estimaciones centradas directa e inversa no coincidan. La estimación centrada directa es  $\sum v^2/r$ , mientras que la centrada inversa, o simplemente centrada, es  $\sum v^2/(r - 2)$ .

Debido a la casi coincidencia, la probabilidad de que el valor estimado sea menor que  $\sigma_0/h$  es la misma que la de que el valor real sea mayor que  $h\sigma_0$ , lo que en general no sería cierto.  $P(\sigma_0 < \sigma_0/h) = P(\sigma_0 > h\sigma_0) = 1 - \chi_r^2((r - 2)/h^2)$  (cf. (16.6)).

Si tomásemos  $\sigma_0^2$  igual a  $\sum v^2/r$ , además de no ser una estimación centrada los valores de  $k(95,5\%)$  de unas páginas atrás habría que multiplicarlos todos por  $r/(r - 2)$ , por lo que las diferencias respecto a los valores 2 y 2,5 serían considerables para pequeños valores de  $r$ .

$k(95\%)$

$d \setminus r$	10	20	30	50	100	200	$\infty$
1	2,29	2,14	2,09	2,06	2,03	2,02	2,00
2	2,93	2,70	2,62	2,57	2,53	2,51	2,49

### Estimación de varias varianzas en un ajuste

Hasta ahora hemos estudiado la estimación de un único  $\sigma_0$ . En el caso más general tenemos varios conjuntos de ecuaciones correspondientes a observaciones con desviaciones típicas  $\sigma_\alpha, \sigma_\beta, \dots, \sigma_\lambda$  desconocidas. El problema es bastante más complicado, ya que la determinación de los valores ajustados de los demás parámetros no es independiente de la de las distintas  $\sigma$ , ya que si la relación entre las desviaciones típicas de dos conjuntos de ecuaciones cambia entonces la estimación de los valores ajustados cambia. La manera de obtener la solución es mediante un proceso iterativo. Tomamos unos valores aproximados  $\sigma_{\alpha_0}, \sigma_{\beta_0}, \dots, \sigma_{\lambda_0}$  y multiplicamos las ecuaciones

## 16 Estimación de la varianza

de residuo por cantidades inversamente proporcionales a ellas. Estas cantidades serán  $\sigma_{0\alpha}/\sigma_\alpha, \dots, \sigma_{0\lambda}/\sigma_\lambda$ . El problema es pues estimar  $\sigma_{0\alpha}, \sigma_{0\beta}, \dots, \sigma_{0\lambda}$ . Una vez resuelto el ajuste calculamos unos valores aproximados para  $\sigma_{0\alpha}, \sigma_{0\beta}, \dots, \sigma_{0\lambda}$  de la forma

$$\sigma_{0\alpha_0} = \sqrt{\frac{\sum v_\alpha^2}{r_\alpha - 2}}, \quad \sigma_{0\beta_0} = \sqrt{\frac{\sum v_\beta^2}{r_\beta - 2}}, \quad \dots, \quad \sigma_{0\lambda_0} = \sqrt{\frac{\sum v_\lambda^2}{r_\lambda - 2}}. \quad (16.14)$$

Para cada desviación típica el sumatorio está extendido a los residuos de sus observaciones correspondientes, y la redundancia por la que se divide es la suma de las redundancias de ese mismo conjunto de observaciones. El cociente de estos valores obtenidos entre las constantes por las que hemos multiplicado cada conjunto de ecuaciones proporciona unos nuevos valores  $\sigma_{\alpha_0}, \sigma_{\beta_0}, \dots, \sigma_{\lambda_0}$ . Tras varios ajustes el proceso converge.

# 17 Atajos

El capítulo 7 deja claro el método a seguir para plantear las ecuaciones de residuo: se escogen tantos parámetros como sean necesarios para expresar todas las magnitudes observadas y se plantean dichas expresiones. Posteriormente se linealiza y se resuelve. Sucede que a veces el número de parámetros es muy grande pero mediante una adecuada agrupación de residuos se pueden disminuir considerablemente. Otras veces la linealización es laboriosa y es más fácil sustituir la fórmula exacta antes de linealizar por otra que puede servir sin introducir error importante. Por último, otras veces lo más complicado de todo el ajuste es la obtención de unos valores aproximados suficientemente buenos como para que el proceso iterativo converja a la solución correcta.

## **Agrupación exacta de residuos, I. Transformación de coordenadas**

En el cálculo de una transformación de coordenadas entre dos conjuntos de puntos las observaciones son las coordenadas de todos los puntos en los dos sistemas. Los parámetros necesarios para expresar esas observaciones son los parámetros de la transformación más las coordenadas de todos los puntos en alguno de los dos sistemas. Si se quiere por ejemplo calcular una transformación de semejanza plana y el número de puntos para ello es de 40, el número de parámetros necesarios es  $4 + 2 \cdot 40$ , y la expresión de los residuos de cada una de las coordenadas de un punto en ambos sistemas en función de estos

## 17 Atajos

parámetros es

$$\begin{aligned}v_{x_1} &= x_1 - x_1, \\v_{y_1} &= y_1 - y_1, \\v_{x_2} &= x_2 - \lambda(\cos \alpha(x_1 - T_x) - \sin \alpha(y_1 - T_y)), \\v_{y_2} &= y_2 - \lambda(\sin \alpha(x_1 - T_x) + \cos \alpha(y_1 - T_y)).\end{aligned}$$

La interpretación geométrica del problema permite agrupar estos cuatro residuos en uno solo. Sea  $p$  el transformado al sistema 2 de un punto del sistema 1 y  $q$  el correspondiente punto en el sistema 2, y sea  $d$  la distancia entre ambos puntos. Supongamos que sabemos de antemano que el factor de escala será  $\approx 1$  y que las precisiones de los puntos son similares en ambos sistemas. En estas circunstancias el punto ajustado en coordenadas del sistema 2 estará en el punto medio entre  $p$  y  $q$ , y la suma de los cuadrados de los residuos asociados a ese punto será  $2(d/2)^2$ . Para el punto en el sistema 2 está claro que  $v_{x_2}^2 + v_{y_2}^2 = (d/2)^2$ . Para el punto 1, si bien las componentes  $v_{x_1}$  y  $v_{y_1}$  no son las componentes  $(x, y)$  del segmento  $d/2$  que hay entre  $p$  y el punto ajustado, debido al giro de la transformación, la suma de los cuadrados es la misma.

Si las precisiones de los puntos en un sistema y otro son distintas o el factor de escala es muy distinto de 1 el punto ajustado en el sistema 2 no estará en el punto medio de  $p$  y  $q$  sino a una cierta relación de distancias de uno y otro, en función de la relación entre las precisiones de ambos: la del punto  $q$  es la precisión del punto en el sistema 2 y la del punto  $p$  la del sistema 1 multiplicada por el factor de escala que pasa de 1 a 2. Si esa relación de distancias es de  $k$ , entre  $p$  y el punto, y  $1 - k$ , entre el punto y  $q$ , se tiene que

$$v_{x_1}^2 + v_{y_1}^2 + v_{x_2}^2 + v_{y_2}^2 = \left( \frac{1}{\lambda^2} k^2 + (1 - k)^2 \right) d^2 = k' d^2$$

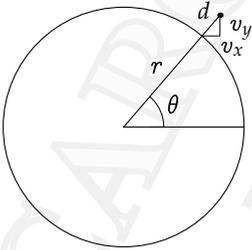
para cierta constante  $k'$ . Si las precisiones de los puntos son similares dentro de cada sistema o al menos mantienen para cada punto una misma relación de precisiones entre uno y otro sistema, la constante  $k'$  será la misma para todos los puntos y su valor exacto no importa ya que la solución mínimo cuadrática va a ser en cualquier caso la que haga mínimo  $\sum d^2$ .

Para expresar  $d^2$  para cada punto son suficientes los parámetros de transformación:

$$d^2 = \left( x_2 - \lambda(\cos \alpha(x_1 - T_x) - \sin \alpha(y_1 - T_y)) \right)^2 + \left( y_2 - \lambda(\sin \alpha(x_1 - T_x) + \cos \alpha(y_1 - T_y)) \right)^2.$$

Con lo que de los 84 parámetros originales se han eliminado 80. Esta expresión se puede simplificar si se consideran coordenadas respecto al centro de gravedad, como se vio en el capítulo 12.

## II. Circunferencia



Vamos a ver otro ejemplo de eliminación de los parámetros coordenada. Supongamos que se midieron las coordenadas de un conjunto de puntos sobre una circunferencia con el objeto de definirla. Suponemos que la medida de cada coordenada sigue una distribución normal y es independiente de la otra coordenada y de las medidas de los otros puntos.

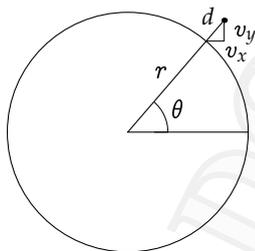
Para expresar cada uno de los residuos, que en este ejemplo son los correspondientes a las coordenadas de los puntos, son necesarios parámetros que definan la posición de cada punto medido sobre la circunferencia. Para eso se pueden tomar las coordenadas del centro de la circunferencia, su radio y un ángulo  $\theta$  para cada punto contado a partir de una dirección fija, por ejemplo el sentido positivo del eje  $x$ . De esta manera las coordenadas ajustadas de cada punto son

$$x_i = x_c + r \cos \theta_i, \quad y_i = y_c + r \sin \theta_i,$$

siendo  $(x_c, y_c)$  las coordenadas del centro de la circunferencia y  $r$  su radio. Si el número de puntos medidos es  $n$  el número de parámetros necesarios para determinar el problema es pues  $3 + n$ .

Sin embargo podemos eliminar los parámetros  $\theta_i$  si observamos que el punto ajustado para cada punto observado será evidentemente el punto de la circunferencia que más próximo le quede, y esto es la

## 17 Atajos



intersección de la circunferencia con la recta que une su centro con el punto observado (siendo más precisos, con el punto con coordenadas iguales a las observadas, ya que el punto observado se encuentra por hipótesis sobre la circunferencia). Si la distancia del punto observado a la circunferencia es  $d$  se tendrá  $v_x^2 + v_y^2 = d^2$ . Sea además  $r'$  la distancia por coordenadas del punto observado  $(x, y)$  al centro

de la circunferencia. Se tiene entonces

$$v_x^2 + v_y^2 = d^2 = (r' - r)^2 = \left( \sqrt{(x - x_c)^2 + (y - y_c)^2} - r \right)^2,$$

y el parámetro  $\theta$  ha desaparecido.

### Agrupación modificada de residuos, I. Circunferencia

Podemos intentar modificar la última ecuación de la siguiente forma:

$$\begin{aligned} d^2 &= (x - x_c)^2 + (y - y_c)^2 - 2r\sqrt{(x - x_c)^2 + (y - y_c)^2} + r^2 \\ &\approx (x - x_c)^2 + (y - y_c)^2 - 2r^2 + r^2 = (x - x_c)^2 + (y - y_c)^2 - r^2. \end{aligned}$$

Pero así llegamos a una cantidad que es negativa si el punto observado queda dentro de la circunferencia ajustada y más negativa cuanto mayor sea  $r$ , lo que no puede ser para el cuadrado de un residuo. La razón de que una pequeña aproximación dé lugar a este resultado erróneo es que en la expresión del término independiente de un residuo, lo que llamamos  $L$ , no se pueden llevar a cabo simplificaciones, como ya se explicó en su momento. En este caso se ha modificado un término que participa en el valor de  $L$ .

Es posible una simplificación de la siguiente manera: Si el punto observado queda exactamente sobre la circunferencia ajustada se tendrá  $r'^2 - r^2 = 0$ , si queda dentro se tendrá  $r'^2 - r^2 < 0$  y si queda fuera  $r'^2 - r^2 > 0$ . Podemos entonces tomar estas cantidades como si fuesen los residuos a minimizar y obtener la solución que minimice

$\Sigma (r'^2 - r^2)^2$ . Para compararla con la solución mínimo cuadrática hay que relacionar las cantidades que se minimizan en uno y otro caso:

$$(r'^2 - r^2)^2 = ((r' - r)(r' + r))^2.$$

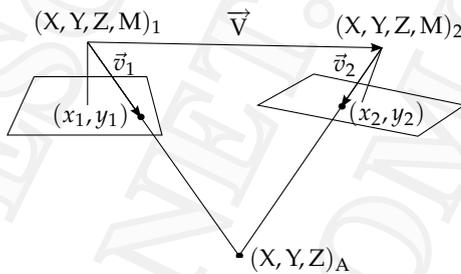
La solución mínimo cuadrática rigurosa minimiza  $\Sigma (r' - r)^2$ , de modo que la solución que proponemos ahora es como la solución mínimo cuadrática aplicando a cada punto un peso igual a  $(r' + r)^2$ . Como los puntos están sobre la circunferencia las cantidades  $(r' + r)$  serán prácticamente iguales para todos los puntos, por lo que no habrá diferencia apreciable entre ambas soluciones.

Expresado en función de los parámetros  $(x_c, y_c, r)$ , las cantidades que hacen de residuos en esta solución son

$$(x - x_c)^2 + (y - y_c)^2 - r^2,$$

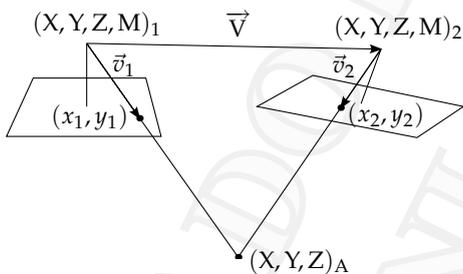
expresión que es más sencilla que los residuos originales y en la que no intervienen raíces cuadradas.

## II. Coplanaridad



Para el cálculo de la orientación relativa de dos fotografías en fotogrametría se tienen como observaciones las fotocoordenadas de los puntos comunes en una y otra foto. La orientación de cada una de las dos fotografías viene dada por las coordenadas

de su centro de proyección y su matriz de rotación, de modo que se tienen  $(X, Y, Z, M)_1$  y  $(X, Y, Z, M)_2$ .  $M_1$  y  $M_2$  son las matrices de rotación, cada una de las cuales se expresa en función de tres parámetros independientes. Se pueden tomar como valores fijos arbitrarios  $(X, Y, Z)_1 = (0, 0, 0)$ ,  $M_1 = I$  y  $X_2 = 1$ , por ejemplo. Hay que obtener entonces  $Y_2, Z_2$  y  $M_2$ .



Los parámetros necesarios para expresar todos los residuos son los cinco necesarios para  $(Y_2, Z_2, M_2)$  más las coordenadas  $(X, Y, Z)$  de cada punto. Si el número de puntos es  $n$  esto significa  $5 + 3n$ . Sin embargo se pueden agru-

par los cuatro residuos correspondientes a cada punto (dos fotocoordenadas en la primera foto y otras dos en la segunda) en una única cantidad. Sean  $(x_1, y_1)$  las fotocoordenadas en la primera foto,  $(x_2, y_2)$  en la segunda y  $f$  la distancia desde el centro de proyección de las fotografías al plano de las mismas (distancia focal). Los vectores  $\vec{V}$ ,  $\vec{v}_1$  y  $\vec{v}_2$  deberían ser coplanarios y por ello su producto mixto igual a cero. La diferencia a cero de dicho producto se puede tomar como un residuo combinado de los cuatro residuos. Es lo que se conoce como condición de coplanaridad.

Para plantear el producto mixto hay que expresar los vectores en un mismo sistema de coordenadas, el que sea. Si escogemos el de la primera fotografía (recuérdese que se ha hecho  $M_1 = I$ ) los anteriores vectores tienen por coordenadas  $(X_2, Y_2, Z_2)$ ,  $(x_1, y_1, -f)$  y  $(x_2, y_2, -f)M_2$ . El producto mixto viene expresado por el determinante formado por los tres vectores. En este determinante ya no aparecen las coordenadas del punto, por lo que una vez más se han eliminado del conjunto de parámetros necesarios para la solución las coordenadas de los puntos.

Un análisis geométrico muestra que si los planos de las fotografías coinciden la minimización de los determinantes equivale a la solución mínimo cuadrática.

### Falsa linealización, I. Transformación proyectiva

La transformación proyectiva es la transformación plana más general que transforma rectas en rectas. Viene dada por el siguiente par de fórmulas:

$$\begin{aligned}x' &= \frac{a_1x + a_2y + a_3}{c_1x + c_2y + 1}, \\y' &= \frac{b_1x + b_2y + b_3}{c_1x + c_2y + 1}.\end{aligned}\tag{17.1}$$

Para calcular los parámetros de la transformación es necesario conocer al menos cuatro puntos en uno y otro sistemas. Si se tienen más se puede proceder a un ajuste mínimo cuadrático. Los residuos para cada punto son en rigor  $x - x'$ ,  $y - y'$ ,  $x' - x''$  e  $y' - y''$ , pero según lo visto unas secciones atrás podemos sustituirlos por  $x' - x''$  e  $y' - y''$  siendo  $(x'', y'')$  el transformado del punto observado  $(x, y)$  de acuerdo a los parámetros de la transformación.

Con esto eliminamos los parámetros coordenada de todos los puntos restando para el ajuste únicamente los ocho parámetros de la proyección. Sin embargo las ecuaciones de residuo siguen sin ser lineales ni, aun sin serlo, admiten una solución tan sencilla como en el caso de la semejanza. Las ecuaciones se pueden linealizar fácilmente en un proceso al que podemos llamar falsa linealización. No porque la linealización sea falsa o inexacta, sino porque son falsos los residuos derivan de ello:

$$\begin{aligned}x'(c_1x + c_2y + 1) &= a_1x + a_2y + a_3, \\y'(c_1x + c_2y + 1) &= b_1x + b_2y + b_3;\end{aligned}$$

de donde

$$\begin{aligned}v_x &= xa_1 + ya_2 + a_3 - xx'c_1 - yy'c_2 - x', \\v_y &= xb_1 + yb_2 + b_3 - xy'c_1 - yy'c_2 - y'.\end{aligned}\tag{17.2}$$

Las cantidades anteriores son *falsos residuos* porque son iguales a los residuos originales  $x' - x''$ ,  $y' - y''$  multiplicados por la cantidad  $c_1x + c_2y + 1$ . Minimizar los falsos residuos (e.d., la suma de sus cuadrados) equivale a aplicar pesos a los residuos originales. Si la proyectividad es «pequeña», en el sentido de que se aparta poco de una transformación afín y con ello las cantidades  $c_1x + c_2y$  son siempre muy pequeñas, los pesos aplicados serán parecidos para todos los puntos y el resultado del ajuste será prácticamente igual que el mínimo cuadrático riguroso. Aun no siendo así puede no merecer la pena realizar el ajuste sin linealizar; en cada caso habrá que analizar la geometría de los puntos para ver si es permisible la linealización o no.

**II. Intersección directa fotogramétrica**

Vamos a ver otro ejemplo de falsa linealización. Supongamos que disponemos de dos fotografías de orientación conocida y queremos calcular las coordenadas de un punto que aparece en ambas. Sean  $(x_1, y_1)$  y  $(x_2, y_2)$  las fotocoordenadas del punto en una y otra fotografías. Se pueden calcular las rectas que pasan por cada uno de esos puntos y el centro de proyección correspondiente y tomar como solución el punto medio de la perpendicular común a ambas. Si el eje óptico de las fotografías es aproximadamente vertical se puede tomar en lugar de la perpendicular común el segmento horizontal de menor longitud que enlaza ambas rectas. Además de ser más fácil el cálculo de este segmento que la perpendicular común, la horizontal mínima responde mejor al principio de minimización de residuos, pues las medidas se efectúan sobre los planos de las fotografías, que son horizontales. La solución mínimo cuadrática no estará exactamente en el punto medio de ese segmento sino ligeramente más alejada de los puntos de toma, ya que aunque el segmento horizontal, en una posición más alejada del mínimo, sea mayor, al estar más alejado su proyección sobre las fotografías será menor. Esto se cumplirá si la separación del segmento mínimo es muy pequeña, ya que la propia longitud del segmento apenas variará, por estar en el entorno del mínimo, pero sí lo hará su proyección sobre los planos de las fotografías.

Si una fotografía está muy girada respecto a la otra no se puede adoptar la solución anterior. La solución mínimo cuadrática minimiza los residuos de las fotocoordenadas  $x_1, y_1, x_2$  e  $y_2$ :

$$\begin{aligned} x_1 &= -f \frac{m_{11}(X - X_1) + m_{12}(Y - Y_1) + m_{13}(Z - Z_1)}{m_{31}(X - X_1) + m_{32}(Y - Y_1) + m_{33}(Z - Z_1)}, \\ y_1 &= -f \frac{m_{11}(X - X_1) + m_{12}(Y - Y_1) + m_{13}(Z - Z_1)}{m_{31}(X - X_1) + m_{32}(Y - Y_1) + m_{33}(Z - Z_1)}, \end{aligned} \quad (17.3)$$

y otro par análogo para las fotocoordenadas en la segunda fotografía. En estas ecuaciones las únicas incógnitas son  $X, Y, Z$ . Todo lo demás son valores conocidos u observados.

Pasando los denominadores al otro miembro y desarrollando todo, y agrupando los números que multipliquen a  $X, Y$  o  $Z$ , se llegará

a expresiones de la forma

$$\begin{aligned}v_{1x} &= a_1 - (b_1X + c_1Y + d_1Z), \\v_{1y} &= a_2 - (b_2X + c_2Y + d_2Z), \\v_{2x} &= a_3 - (b_3X + c_3Y + d_3Z), \\v_{2y} &= a_4 - (b_4X + c_4Y + d_4Z).\end{aligned}\tag{17.4}$$

La solución por mínimos cuadrados de este sistema dará un punto ajustado  $(X, Y, Z)$ .

Los residuos anteriores son los residuos originales multiplicados por el denominador de las fracciones en (17.3). Este denominador es la coordenada  $Z$  del punto en el sistema de la fotografía; es decir, lo que se suele llamar la distancia del punto a la fotografía. Por tanto lo que se minimiza no son los residuos sobre la fotografía sino los segmentos paralelos a esos pero situados donde el punto  $(X, Y, Z)$ , y es por tanto el análogo al mínimo segmento horizontal entre las rectas proyectivas en el caso de ejes ópticos verticales. Si el punto está a una distancia parecida de ambas fotografías no habrá diferencia apreciable entre una y otra solución. Aun cuando se encuentre a la mitad de distancia de una fotografía que de la otra la solución será perfectamente válida. Más aún, es muy probable que si una fotografía se encuentra más cerca del objeto que la otra sus parámetros de orientación se conozcan con una precisión proporcionalmente mejor, por lo que la solución mínimo cuadrática rigurosa resulta ser la de la falsa linealización.

Valga esto para recordar una vez más que la solución mínimo cuadrática no se emplea muchas veces porque sea la más adecuada desde un punto de vista teórico sino porque es la más fácil de calcular, y si resulta que en algunos problemas encontramos otras soluciones más fáciles que aquella no hay razón para no emplearlas mientras arrojen soluciones correctas, y en algunos casos pueden ser incluso más correctas que la solución mínimo cuadrática.

### Linealización forzada, I. Rotación 3D

Vamos a abordar ahora el problema de obtener valores aproximados cuando ello es difícil y ver cómo se puede hacer más fácil mediante uno de esos atajos.

## 17 Atajos

Supongamos que queremos obtener los parámetros  $a, b, c$  de una transformación plana definida de la siguiente manera:

$$\begin{aligned}x' &= xa + yb + xyab + x^2yc(b^2 - 2a^2), \\y' &= xc + ya + xyac + y^2xb(c^2 - 2a^2).\end{aligned}$$

Si por la naturaleza del problema las componentes de segundo orden y mayores van a ser muy pequeñas se pueden calcular valores aproximados para  $a, b$  y  $c$  empleando solamente las componentes de primer orden de la transformación, como si las de segundo orden no existiesen. Si esto no es así habrá que buscar otra solución. Una manera posible de hallar los parámetros consiste en calcular de manera independiente cada una de las combinaciones de los parámetros  $a, b, c$  que aparecen en las ecuaciones de transformación. En este caso esto significa hacer

$$\begin{aligned}a &= A, & b &= B, & c &= C, & ab &= D, \\ac &= E, & c(b^2 - 2a^2) &= F, & b(c^2 - 2a^2) &= G.\end{aligned}$$

Tomando 4 o más puntos se resuelve el sistema lineal por mínimos cuadrados:

$$\begin{aligned}x'_1 &= x_1A + y_1B + x_1y_1D + x_1^2y_1F, \\y'_1 &= x_1C + y_1A + x_1y_1E + y_1^2x_1G, \\x'_2 &= x_2A + y_2B + x_2y_2D + x_2^2y_2F, \\&\text{etc.}\end{aligned}$$

El sistema es lineal porque las incógnitas son  $A, B, \dots G$ . Lo demás son las coordenadas de los puntos; para cada punto dos ecuaciones.

Una vez resuelto se puede tomar  $a = A, b = B$  y  $c = C$ . Estos valores no pueden ser la solución final ya que eso equivale a ignorar las relaciones entre los parámetros  $A \dots G$ . Empleando otro de estos en lugar de, o junto con,  $A, B$  o  $C$  para la estimación de  $a, b$  y  $c$  se obtendrían otros valores para estos últimos. Los valores de estos tres parámetros, se obtengan como se obtengan, han de tomarse como punto de partida para una solución rigurosa.

Al contrario de las linealizaciones que dan lugar a falsos residuos, que son transformaciones matemáticas exactas, la linealización forzada no se puede emplear para sustituir las ecuaciones originales por

otras y resolver el sistema nuevo en lugar del original. Es sin embargo un método muy útil para obtener valores aproximados en multitud de problemas en los que las relaciones entre observaciones y parámetros son complicadas y no hay manera de saber a priori cuáles pueden ser aproximadamente los valores de los parámetros.

Uno de esos problemas es la obtención de la rotación tridimensional que pasa de un conjunto de puntos a otro. La solución óptima a este problema se vio en el capítulo 13, pero para poder calcular esa solución es necesaria una matriz de giro aproximada. Ya se apuntó en su momento la linealización forzada como solución aunque no se mencionara esta expresión. Se calculan los nueve números de la matriz como si fuesen independientes:

$$\begin{aligned}x'_1 &= r_{11}x_1 - r_{12}y_1 + r_{13}z_1, \\y'_1 &= r_{21}x_1 - r_{22}y_1 + r_{23}z_1, \\z'_1 &= r_{31}x_1 - r_{32}y_1 + r_{33}z_1.\end{aligned}$$

Con tres puntos conocidos en uno y otro sistemas se pueden obtener los nueve valores. Obsérvese que las ecuaciones para  $x'$ ,  $y'$  y  $z'$  son independientes entre sí, de modo que no es necesario resolver un sistema de 9 ecuaciones con 9 incógnitas sino 3 sistemas con 3 incógnitas cada uno. Si se dispone de más puntos se puede llevar a cabo un ajuste mínimo cuadrático.

Una vez obtenidos los nueve valores hay que proceder a modificarlos de manera que la matriz que forman sea una matriz de rotación. Habrá que procurar que las modificaciones sean pequeñas. Se desarrolla a continuación un posible procedimiento. Sean

$$f_1 = (r_{11}, r_{12}, r_{13}), \quad f_2 = (r_{21}, r_{22}, r_{23}), \quad f_3 = (r_{31}, r_{32}, r_{33}).$$

En primer lugar ajustamos el módulo de  $f_3$  a 1, dividiendo cada una de sus componentes por  $|f_3|$ . A continuación sumaremos a  $f_1$  y  $f_2$  sendos múltiplos de  $f_3$ , ya modificado, de manera que el resultado pase a ser perpendicular a  $f_3$ :

$$f_1 \rightarrow f'_1 = f_1 + k_1 f_3, \quad f_2 \rightarrow f'_2 = f_2 + k_2 f_3;$$

$$0 = f'_1 f_3 = f_1 f_3 + k_1 f_3 f_3 = f_1 f_3 + k_1,$$

de donde  $k_1 = -f_1 f_3$  y análogamente  $k_2 = -f_2 f_3$ .

## 17 Atajos

Posteriormente hacemos el módulo de  $f'_2$  igual a 1, dividiendo cada componente por su módulo. Una vez hecho esto modificamos  $f'_1$  sumándole un múltiplo de  $f'_2$  de manera que el resultado sea ortogonal a  $f'_2$ . Tal modificación no altera la ortogonalidad entre  $f'_1$  y  $f_3$  ya que si tanto  $f'_1$  como  $f'_2$  son perpendiculares a  $f_3$ , cualquier combinación lineal de ellos también lo es:

$$f'_1 \rightarrow f''_1 = f'_1 + k_3 f'_2 \quad \Rightarrow \quad 0 = f''_1 f'_2 = f'_1 f'_2 + k_3, \quad k_3 = -f'_1 f'_2.$$

Por último ajustamos el módulo de  $f''_1$  a 1.

## II. Coplanaridad

La linealización forzada se emplea para obtener los valores iniciales para una orientación relativa de dos fotografías. En primer lugar se agrupan residuos mediante la condición de coplanaridad como se explicó anteriormente. Como resultado de este agrupamiento aparecen los residuos agrupados:  $v =$

$$\begin{aligned} &+(m_{12}Z - m_{13}Y)p_x q_x &+(m_{22}Z - m_{23}Y)p_x q_y &-(m_{32}Z - m_{33}Y)p_x \\ &+(m_{13}X - m_{11}Z)p_y q_x &+(m_{23}X - m_{21}Z)p_y q_y &-(m_{33}X - m_{31}Z)p_y \\ &-(m_{11}Y - m_{12}X)q_x &-(m_{21}Y - m_{21}X)q_y &+(m_{31}Y - m_{32}X) \end{aligned} \quad (17.5)$$

En donde  $(X, Y, Z)$  son las coordenadas de un centro de proyección respecto al otro, en el sistema de coordenadas de una de las dos fotografías, a la que llamaremos la *primera*, siendo la otra la *segunda*;  $m_{11} \dots m_{33}$  son las componentes de la matriz de rotación que aplicada de la forma  $x' = Mx$  transforma un vector  $x$  del sistema de coordenadas de la primera fotografía al correspondiente  $x'$  en el sistema de la segunda. Por último  $(p_x, p_y)$  son las fotocoordenadas  $(x, y)$  del punto en la primera fotografía divididas por la focal  $f$  y análogamente  $(q_x, q_y)$  las fotocoordenadas en la segunda fotografía divididas por  $f$ . Se tiene una ecuación de residuo como esta para cada punto.

La linealización forzada en este problema consiste en tratar las 9 cantidades  $(m_{12}Z - m_{13}Y) \dots (m_{31}Y - m_{32}X)$  como si fuesen independientes:  $L_1 \dots L_9$ . Las ecuaciones de residuo son entonces lineales y se resuelve el ajuste mínimo cuadrático sin necesidad de obtener unos valores aproximados previos, lo que sería difícil y es precisamente la razón de que se emplee la linealización forzada.

En la orientación relativa de dos fotografías la escala es libre, por lo que hay que fijar arbitrariamente algún parámetro que defina la escala. En el ejemplo de unas secciones atrás se fijó  $X = 1$ . Esta libertad se ve reflejada en las ecuaciones de residuo en el hecho de que si se multiplican las coordenadas  $X, Y, Z$  por un número todos los residuos se multiplican por ese mismo número. En el ajuste tras la linealización forzada los parámetros son  $L_1 \dots L_9$ , que al igual que  $X, Y, Z$  se multiplican todos por una constante si  $X, Y, Z$  se multiplican por esa misma constante. Por lo tanto se pueden multiplicar todos por una constante variando así la escala del modelo y hay que fijar uno de ellos para el ajuste, por ejemplo  $L_6 = 1$ , quedando ocho parámetros para calcular. La constante por la que se multiplican todos los valores de  $L$  puede ser incluso negativa, al menos de cara al ajuste, por lo que no tiene sentido plantear si las coordenadas  $(X, Y, Z)$  son del fotograma segundo respecto al primero o viceversa.

En la linealización forzada aparecen dos problemas. Uno de ellos es que una pequeña variación de uno de los parámetros de verdad (los originales antes de la linealización) puede provocar una variación muy grande de uno de los parámetros recombinados (los llamaremos así porque cada uno de ellos es una combinación de los originales) o viceversa. Cuando esto es así suele haber una división por un número cercano a cero. Este no es un problema exclusivo de la linealización forzada y tiene que ver en general con una elección de parámetros adecuada a la geometría del problema. En la sección anterior, por ejemplo, en donde se tomó como referencia la fila tercera de la matriz obtenida para modificar las filas primera y segunda, esto podría dar lugar a una alteración muy grande de la matriz si resulta que la fila tercera tenía un módulo muy pequeño. Es mejor por tanto tomar como referencia la fila con mayor módulo. En el ejemplo que nos ocupa ahora la elección problemática es la del parámetro  $L$  que se fija. Hemos propuesto a modo de ejemplo  $L_6 = 1$ . Esto está bien siempre que  $L_6$  no sea cero o un valor muy pequeño. Esto sucederá si el vector  $(X, Y, Z)$  es aproximadamente proporcional a  $(0, 1, 0)$ , además de otros casos según los valores de  $m_{33}$  y  $m_{31}$ . Si  $L_6$  está muy cercano a cero su precisión en proporción a sí mismo es pésima, lo que equivale a que fijado  $L_6 = 1$  la precisión de los demás valores de  $L$ , que resultarían muy grandes, es muy mala; tanto que pueden aparecer problemas numéricos en el ajuste.

En esos casos habría que fijar otro valor de  $L$ . El problema es que no podemos saber a priori cuáles serán grandes y cuáles pequeños a partir de las fotocoordenadas de los puntos comunes a menos que realicemos ciertas suposiciones previas, como por ejemplo que los planos de ambas fotografías son aproximadamente paralelos. Si estas suposiciones no son posibles habrá que decidir el parámetro fijo según se va resolviendo el sistema de las ecuaciones normales, reordenando filas y columnas según sea necesario para evitar dividir por un valor muy pequeño. No se puede dar el caso de que todos los valores de  $L$  sean pequeños y estén mal definidos en proporción a sí mismos ya que  $\sum L_i^2 = 2(X^2 + Y^2 + Z^2)$ , como se comprueba fácilmente, por lo que al menos uno está bien definido.

El otro problema de la linealización forzada es mucho más grave que el anterior. Al fin y al cabo el problema anterior siempre se puede resolver realizando el ajuste varias veces, tomando cada vez como fijo un parámetro  $L$  distinto. Como resultado de la linealización forzada el número de parámetros aumenta. En este caso pasa de cinco a ocho (una vez fijado uno de los parámetros  $L$ ). Los parámetros recombinados forman un conjunto  $n'$ -dimensional, del cual solamente un subconjunto  $n$ -dimensional equivale exactamente a  $n$  valores de los parámetros originales. Sea  $\Psi$  el conjunto de todas las  $n'$ -uplas de valores de parámetros recombinados. En este ejemplo esto quiere decir el conjunto de todas las 9-uplas  $L_1, L_2, \dots, L_9$  (que tiene dimensión 8 porque uno de ellos es fijo). Sea  $\Omega$  el subconjunto de  $n'$ -uplas de  $\Psi$  que corresponden exactamente a una  $n$ -upla de los parámetros originales. En este ejemplo esto significa el subconjunto de dimensión 5 de  $\Psi$  cada uno de cuyos elementos se corresponde exactamente con unos valores de  $X, Y, Z, M$ .

Si se ha medido una serie de puntos en las dos fotografías y las medidas estuviesen libres de error, si se resuelve el sistema para los  $L_i$ , se obtendría una solución que hace los residuos iguales a cero y que se corresponden con unos valores de  $X, Y, Z, M$ ; es decir, una solución «posible», esto es, una solución para los  $L_i$  que pertenece a  $\Omega$ . Supongamos ahora que el número de puntos medidos es de solamente cinco. El sistema de ecuaciones de residuo para resolver los  $L_i$  consta de cinco ecuaciones y ocho incógnitas por lo que la solución para los  $L_i$  es un conjunto de dimensión 3. En todo ese conjunto de infinitas 9-uplas habrá una única que se pueda hacer corresponder con unos valores  $X, Y, Z, M$ , es decir  $\in \Omega$ ; las demás son soluciones imposibles. Si va-

mos aumentando el número de puntos el conjunto de soluciones posibles para los  $L_i$  va disminuyendo de dimensión, hasta que al llegar a ocho puntos comunes se reduce a una única solución, la perteneciente a  $\Omega$ . Ahora bien, habrá unas posiciones o superficies particulares en el espacio para los puntos sexto, séptimo, etc. de manera que si esos puntos están en esas posiciones las ecuaciones de residuo que añaden se sigan cumpliendo para todo el conjunto 3-dimensional de soluciones  $L_i$  que se obtiene a partir de los cinco primeros puntos, o al menos para un subconjunto de dimensión 2 o 1 de ese conjunto. Si los puntos se sitúan en esas posiciones, por muchos puntos que añadamos no podremos disminuir la dimensión del conjunto de  $L_i$  a un único punto: la solución verdadera  $\in \Omega$ . Este conjunto de posiciones malas serán rectas o superficies, de modo que un punto «cualquiera» no estará nunca exactamente sobre una de ellas y por tanto ocho puntos siempre permitirían determinar la solución única. Pero para que eso fuese así sería necesario que las medidas estuviesen libres de error. En la realidad, si los puntos se sitúan cerca de las posiciones malas, los pequeños errores de medida hacen que la solución que se obtenga para los  $L_i$  sea una de ese conjunto 1, 2 o 3-dimensional que puede estar muy alejado de la solución  $\in \Omega$ .

La problemática planteada no es resultado de divagaciones puramente teóricas. Al contrario, son tantas las posibles configuraciones malas para los puntos medidos, debido a las tres dimensiones de margen entre  $\Omega$  y  $\Psi$ , y los puntos en fotogrametría tienden a acercarse a esas configuraciones debido a que por lo general se separan poco de una superficie plana, que es muy común que incluso con 10 o 12 puntos medidos los valores obtenidos para  $L_1 \dots L_9$  estén tan alejados de una 9-upla posible que no se pueda con ellos determinar los valores de  $X, Y, Z, M$ .

Entre las ecuaciones de residuo expresadas en función de los parámetros originales y la linealización total puede haber expresiones intermedias que conserven algunas de las relaciones entre parámetros y que por esta razón no den lugar a soluciones malas. Cuanto más se disminuya la dimensión de  $\Psi$  menor será el conjunto de configuraciones malas para las observaciones. En el problema de resolver la orientación relativa mediante las ecuaciones de residuo (17.5) se puede aplicar la linealización forzada y además de fijar el valor de

## 17 Atajos

uno de los nueve parámetros tener en cuenta que

$$\begin{vmatrix} L_1 & L_2 & L_3 \\ L_4 & L_5 & L_6 \\ L_7 & L_8 & L_9 \end{vmatrix} = 0.$$

Esto permite desarrollar una solución con 7 puntos medidos en la que se resuelve el sistema de las ecuaciones de residuo más una ecuación de grado 3. Lo que excede ahora la dimensión de  $\Psi$  a la de  $\Omega$  es 2, disminuyendo por ello el conjunto de configuraciones malas, y pudiéndose dar el caso de que la configuración de los puntos no permita resolver el problema mediante la linealización forzada total pero sí mediante este método de tratar solamente 7 de los  $L_i$  como independientes en lugar de 8.

Estos métodos intermedios suelen ser bastante complicados, más cuanto más queramos reducir la dimensión de  $\Psi$ , y específicos para cada problema. La obtención de unos valores aproximados iniciales puede convertirse en la parte más complicada del ajuste, pudiendo ser necesario terminar recurriendo a la fuerza bruta: ir probando varios valores iniciales hasta que con unos de ellos el ajuste converja.

# 18 Interpretación geométrica

## La solución mínimo cuadrática

La mayoría de los elementos que intervienen en una estimación mínimo cuadrática tienen un significado geométrico claro. En el espacio  $\mathbb{R}^m$  de valores observados sólo un subconjunto de dimensión  $n$  es posible. Si sustituimos las relaciones exactas por la aproximación lineal ese subconjunto es lineal. Podemos situar el origen de coordenadas de  $\mathbb{R}^m$  en el punto observado, el punto real u otro punto posible. En estos dos últimos casos el subconjunto de valores posibles pasa por el origen, mientras que en el primero generalmente no es así. Si estamos analizando el proceso de estimación, en el que mantenemos fijo  $\mathcal{L}$ , puede ser más cómodo situar el origen en ese punto, mientras que si estudiamos el problema desde el punto de vista directo lo más adecuado tal vez sea emplear el punto  $L_r$ . En la figura 18.1 se muestran varios vectores que intervienen en la estimación, con el origen del sistema de coordenadas en  $\mathcal{L}$ .

La figura representa una situación en la que  $m = 3$  y  $n = 2$ . Llamamos  $\Omega$  al subconjunto de valores reales posibles. Dado un punto  $L_r \in \Omega$ , la cantidad  $\sum \varepsilon^2$  es  $\overline{L_r \mathcal{L}^2}$ . La solución mínimo cuadrática es entonces el punto  $L_{aj} \in \Omega$  tal que  $\overline{L_{aj} \mathcal{L}}$  es mínimo, es decir, el pie de la perpendicular a  $\Omega$  trazada desde  $\mathcal{L}$ . Un punto  $(l_1, \dots, l_n)$  de  $\Omega$  linealizado lo podemos escribir como

$$L_x = L_0 + \Delta x_1 \vec{x}_1 + \dots + \Delta x_n \vec{x}_n. \quad (18.1)$$

## 18 Interpretación geométrica

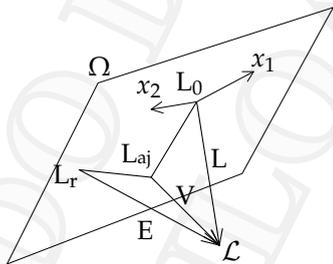


Fig. 18.1

Los vectores  $\vec{x}$  pertenecen al espacio  $\mathbb{R}^m$  y tienen por lo tanto  $m$  componentes:

$$\vec{x}_1 = (x_{1l_1}, x_{1l_2}, \dots, x_{1l_m}).$$

Y análogamente para  $\vec{x}_2, \dots, \vec{x}_n$ , y se cumple

$$\frac{\Delta l_j}{\Delta x_j} = x_{j l_i}.$$

A cada una de estas componentes las llamaremos  $a_{ij}$ . Los vectores  $\vec{x}$  no son en general ni ortogonales ni del mismo módulo. La ecuación (18.1) es una expresión paramétrica de los  $L_x$  posibles en función de  $n$  parámetros:  $(x_1, \dots, x_n)$ . Es una expresión compacta formada por vectores. Si escribimos por separado la ecuación correspondiente a cada  $L_x$  el sistema es el siguiente:

$$\begin{aligned} l_1 &= l_{10} + \Delta x_1 a_{11} + \Delta x_2 a_{12} + \dots + \Delta x_n a_{1n} \\ l_2 &= l_{20} + \Delta x_1 a_{21} + \Delta x_2 a_{22} + \dots + \Delta x_n a_{2n} \\ &\vdots \\ l_m &= l_{m0} + \Delta x_1 a_{m1} + \Delta x_2 a_{m2} + \dots + \Delta x_n a_{mn} \end{aligned}$$

que a su vez podemos escribir matricialmente como

$$L_x = L_0 + A(\Delta x_1 \ \dots \ \Delta x_n)^T. \quad (18.1a)$$

El vector  $V$  es perpendicular a  $\Omega$ , lo que significa que es perpendicular a todos los vectores contenidos en  $\Omega$ . Para ello es necesario y

## 18 Interpretación geométrica

suficiente que  $V$  sea perpendicular a  $n$  vectores linealmente independientes contenidos en  $\Omega$ . Los vectores  $\vec{x}_1, \dots, \vec{x}_n$  cumplen esa condición. La condición de perpendicularidad entre dos vectores se puede expresar mediante el producto escalar, de modo que  $V$  es el vector que satisfice

$$(a_{11}, a_{21}, \dots, a_{m1})^T V = 0, \quad \dots \quad (a_{1n}, a_{2n}, \dots, a_{mn})^T V = 0.$$

Que en forma matricial es  $A^T V = 0$ .

Hay muchos vectores que satisfacen esa condición, y  $V$  es el único de entre ellos que tiene por extremo el punto  $\mathcal{L}$  y el origen en un punto de  $\Omega$ , en concreto en  $L_{aj}$ , de modo que

$$V = \mathcal{L} - (L_0 + A(\Delta x_1 \ \dots \ \Delta x_n)^T) = L - AX_\Delta$$

para algunos valores  $(\Delta x_1, \dots, \Delta x_n)$ , de donde

$$\begin{aligned} A^T(L - AX_\Delta) = 0 &\Rightarrow A^T L - A^T A X_\Delta = 0 \\ &\Rightarrow X_\Delta = N^{-1} A^T L. \end{aligned}$$

El punto  $L_0$  es cualquier punto de  $\Omega$ . Si tomamos  $L_0 = L_r$  entonces

$$X_\Delta = X_{aj} - X_r = N^{-1} A^T E, \tag{18.2}$$

y este es el error en la estimación.

Una vez obtenido  $X_\Delta$  calculamos  $L_{aj}$  y  $V$ . Si sustituimos  $X_\Delta$  por su expresión en función de  $A$  y  $L$ , las expresiones para  $L_{aj}$  y  $V$  son

$$\begin{aligned} L_{aj} &= L_0 + AN^{-1}A^T L, \\ V &= \mathcal{L} - L_{aj} = L - AN^{-1}A^T L = (I - AN^{-1}A^T)L. \end{aligned} \tag{18.3}$$

Desde  $\mathcal{L}$  sólo se puede trazar un vector perpendicular a  $\Omega$ , pero desde  $L_{aj}$  se pueden trazar infinitos. Todos esos vectores forman el subespacio perpendicular a  $\Omega$  pasando por  $L_{aj}$ , que llamaremos  $\bar{\Omega}$ . Tiene dimensión  $m - n$ . En la figura 18.1  $m - n = 1$ , y el subespacio es la recta que pasa por  $L_{aj}$  y  $\mathcal{L}$ . Normalmente será de dimensión mayor. Es en todo caso el subespacio perpendicular a  $\Omega$  por  $\mathcal{L}$ , y corta a  $\Omega$  en un único punto  $L_{aj}$ .

## 18 Interpretación geométrica

La proyección de  $L$  sobre  $\Omega$  es  $L_{aj} - L_0$ , y sobre  $\bar{\Omega}$  es  $V$ . El vector  $V$  en función de  $L$  viene dado por (18.3), así que la matriz  $(I - AN^{-1}A^T)$  es la que transforma  $L$  en su proyección sobre  $\bar{\Omega}$ . Ahora bien, las columnas de  $A$  son los vectores  $(\vec{x}_1, \dots, \vec{x}_n)$ , que son una base de  $\Omega$ , de modo que dada una matriz cuyas columnas forman una base de un subespacio, la matriz

$$(I - A(A^T A)^{-1} A^T) \quad (18.4)$$

es la que proyecta cualquier vector sobre el subespacio ortogonal.

La propiedad la hemos deducido solamente para vectores  $L = \mathcal{L} - L_0$  siendo  $L_0 \in \Omega$ . Cualquier vector que no sea paralelo a  $\Omega$  será igual a  $kL$  para algún  $L$  de los anteriores y un número real  $k$ , y mediante la matriz se transforma en  $kV$ , por lo que sigue siendo válida. Finalmente, si  $L$  es paralelo a  $\Omega$  entonces es de la forma  $AX$  para algún vector  $X$ , y su producto por la matriz de proyección es

$$(I - A(A^T A)^{-1} A^T)AX = AX - AX = 0,$$

de modo que la matriz es válida para cualquier vector.

La proyección de  $L$  sobre  $\Omega$  es  $L - V$ , igual a

$$(A(A^T A)^{-1} A^T)L,$$

por lo que  $A(A^T A)^{-1} A^T$  es la matriz que proyecta un vector sobre el subespacio generado por las columnas de  $A$ .

### Funciones $g(E)$ , $g(X_r)$ y $g(L_r)$

Por la ortogonalidad entre  $V$  y  $\Omega$  se cumple

$$E^2 = (\mathcal{L} - L_r)^2 = V^2 + (L_{aj} - L_r)^2. \quad (18.5)$$

La función  $g(E)$ , definida para los puntos de  $\Omega$ , depende de  $\varepsilon_1^2 + \dots + \varepsilon_m^2 = E^2$ , y es por lo tanto simétrica respecto a  $L_{aj}$ , por lo que la estimación mínimo cuadrática es centrada. Sea  $(L_{aj}, \vec{y}_1, \dots, \vec{y}_n)$  un sistema de coordenadas ortonormal centrado en  $L_{aj}$  y contenido en  $\Omega$ . Si suponemos  $\sigma_0 = 1$  entonces

$$g(Y) \propto e^{-\frac{1}{2}(Y^2 + V^2)} = e^{-\frac{1}{2}V^2} e^{-\frac{1}{2}Y^2} \propto e^{-\frac{1}{2}Y^2}.$$

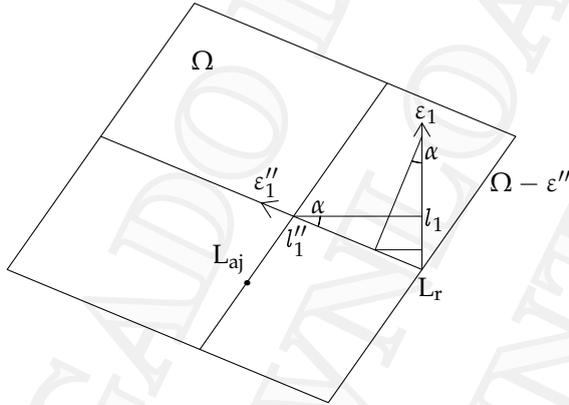


Fig. 18.2

Expresando los vectores  $\vec{y}$  en función de los vectores  $\vec{x}$ , es decir, de la matriz  $A$ , podríamos llegar a las expresiones para  $g(X_r)$  y  $g(L_r)$  deducidas en el capítulo 10.

### Contribución de cada error a la solución ajustada

Vamos a estudiar ahora el problema desde el punto de vista directo. La figura 18.2 representa el espacio  $\mathbb{R}^m$  de posibles valores observados centrado en el valor real  $L_r$ . Las variables  $\epsilon_1, \dots, \epsilon_m$  son independientes y siguen distribuciones normales  $(0, 1)$  (supondremos  $\sigma_0 = 1$ ). Entonces cualquier giro de los ejes dará lugar a otras variables  $\epsilon_1'', \dots, \epsilon_m''$  también independientes y también siguiendo cada una una distribución  $N(0, 1)$ .

Podemos tomar  $\epsilon_1''$  de manera que siga la dirección de la proyección de  $\epsilon_1$  sobre  $\Omega$ . El conjunto de vectores de  $\Omega$  perpendiculares a  $\epsilon_1''$  es el subespacio  $\Omega - \epsilon_1''$ , de dimensión  $n - 1$ . Las componentes  $\epsilon_2'', \dots, \epsilon_n''$ , al ser perpendiculares a  $\epsilon_1''$ , se podrán descomponer a su vez en una componente en el espacio  $\bar{\Omega}$  y otra en  $\Omega - \epsilon_1''$ . Por lo tanto la proyección de cada una de ellas sobre  $\Omega$  se encuentra en un punto del espacio  $\Omega - \epsilon_1''$ , y entonces la proyección del vector  $E$  sobre  $\Omega$ , el punto  $L_{aj}$ , será la componente  $\epsilon_1''$  más un vector de  $\Omega - \epsilon_1''$ , como se

## 18 Interpretación geométrica

muestra en la figura, y la componente  $l_1''$  del vector  $L_{aj}$  coincide con la observada.

El vector  $L_{aj}$  tiene unas componentes  $(l_1, \dots, l_m)$ , cada una de las cuales es la proyección de  $L_{aj}$  sobre el eje correspondiente. El subespacio  $\Omega - \varepsilon_1''$  es perpendicular a  $\varepsilon_1$ , o lo que es lo mismo a  $l_1$ , de modo que la proyección de  $L_{aj}$  sobre este eje es la misma que la de su componente  $l_1'' = \ell_1''$ .

$$\overline{L_{aj} l_1} = \overline{L_{aj} l_1''} \sin \alpha.$$

Y puesto que la variable  $\ell_1''$  sigue una distribución  $N(0, 1)$ , entonces  $l_1 \sim N(0, \sin \alpha)$ .

El valor  $\sin \alpha$  es fácil de calcular. La proyección de un vector de la forma  $(\varepsilon_1, 0, \dots, 0)$  sobre  $\Omega$  sigue la dirección de  $\varepsilon_1''$  y tiene módulo  $\sin \alpha \varepsilon_1$ . A su vez la componente  $\varepsilon_1$  de este vector es su proyección sobre el eje  $\varepsilon_1$ , y mide  $\sin^2 \alpha \varepsilon_1$ . Si tomamos  $\varepsilon_1 = 1$  entonces la proyección sobre  $\Omega$  del vector es

$$\vec{\varepsilon}_{1\Omega} = AN^{-1}A^T \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

que es la primera columna de la matriz de proyección  $AN^{-1}A^T$ , a la que llamaremos  $P_\Omega$ , y su primera componente el elemento 1,1 de la matriz. Si en lugar de  $\varepsilon_1 = 1$  tomamos  $\varepsilon_i = 1$ , su proyección sobre  $\Omega$  es la columna  $i$  y la proyección de éste sobre el eje  $\varepsilon_i$  es el elemento  $i,i$  de la matriz. Por lo tanto

$$\sigma_{\varepsilon_i}^2 = \sin^2 \alpha_i = p_{\Omega ii}.$$

La proyección de un vector  $(0, \dots, \varepsilon_i, 0, \dots, 0)$  sobre  $\Omega$  es lo que contribuye el error  $\varepsilon_i$  a la solución ajustada, es decir, al error en la solución ajustada. Las distintas contribuciones se suman vectorialmente y se obtiene el punto  $L_{aj}$ . El vector  $\vec{\varepsilon}_{1\Omega}$  sigue como ya vimos una distribución  $N(0, \sin \alpha_i)$ . Si a su vez la proyectamos sobre un eje  $\varepsilon_j$ , el vector resultante sigue una distribución  $N(0, \sin \alpha_{ji})$ , en donde  $\alpha_{ji}$  no es ningún ángulo, simplemente una notación convencional, y es la componente  $l_j$  generada por un error unitario  $\varepsilon_i$ . De igual manera que como obtuvimos  $\sin \alpha_i^2$  vemos que  $\sin \alpha_{ji}$  es el elemento  $j,i$  de

## 18 Interpretación geométrica

$P_{\Omega}$ . Si sumamos para cada componente  $l_j$  las contribuciones debidas a cada  $\varepsilon_i$  obtenemos

$$L_{aj} - L_r = AN^{-1}A^T E,$$

que ya sabíamos.

### Matrices $\Sigma_{ll}$ y $\Sigma_{vv}$

Los errores  $\varepsilon_1, \dots, \varepsilon_n$  son estadísticamente independientes, y por tanto también sus contribuciones a  $l_i$  y  $l_j$ . Aplicando la transmisión de varianzas obtenemos

$$\sigma_{l_i l_j} = p_{\Omega i1} p_{\Omega j1} + \dots + p_{\Omega im} p_{\Omega jm}.$$

La matriz  $P_{\Omega}$  es simétrica, así que esta suma por filas es igual a la misma suma por columnas. Pero ya vimos que la columna  $i$  es  $\vec{\varepsilon}_{i\Omega}$ , y la suma es entonces el producto escalar  $\vec{\varepsilon}_{i\Omega} \vec{\varepsilon}_{j\Omega}$ , que a su vez es el producto del módulo de  $\vec{\varepsilon}_{i\Omega}$  por la proyección de  $\vec{\varepsilon}_{j\Omega}$  sobre  $\vec{\varepsilon}_{i\Omega}$ . En la figura 18.2 vemos que las proyecciones de un vector  $L_x$  cualquiera de  $\Omega$  sobre  $\vec{\varepsilon}_{i\Omega}$  y sobre el eje  $\varepsilon_i$ , que llamaremos simplemente  $y, z$ , están en la relación

$$z = y \operatorname{sen} \alpha_i.$$

Pero la proyección de  $\vec{\varepsilon}_{j\Omega}$  sobre  $\varepsilon_i$  es  $\sin \alpha_{ij} = p_{\Omega ij}$ , y entonces

$$\sigma_{l_i l_j} = |\vec{\varepsilon}_{i\Omega}| \frac{p_{\Omega ij}}{\operatorname{sen} \alpha_i} = p_{\Omega ij},$$

así que  $\Sigma_{ll} = P_{\Omega}$ . Con pocas operaciones más podríamos demostrar que  $\Sigma_{vv} = I - \Sigma_{ll}$ , y en particular  $\sigma_{v_i}^2 = 1 - \operatorname{sen}^2 \alpha_i = \operatorname{cos}^2 \alpha_i$ .

### Redundancias parciales

Vamos a estudiar ahora la redundancia parcial de una observación, definida como  $E[v_i]_{\varepsilon_i} / \varepsilon_i$ .

La figura 18.3 representa la situación en la que  $\varepsilon_1$  es fijo. El conjunto de posibles valores observados es el subespacio  $\varepsilon_1 = \text{cte}$ , de dimensión  $m - 1$ , que corta a  $\Omega$  en un subespacio paralelo a  $\Omega - \varepsilon_1$ .

## 18 Interpretación geométrica

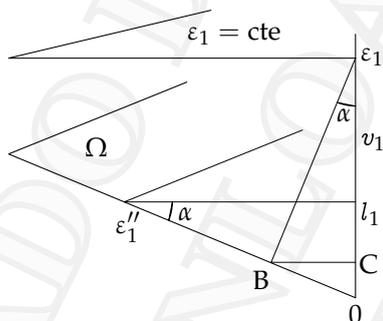


Fig. 18.3

Ahora  $\varepsilon_1''$  sigue una distribución normal centrada en el punto B, a una distancia  $\varepsilon_1 \sin \alpha$  del origen. A su vez  $l_1$  sigue una distribución centrada en el punto C:  $\varepsilon_1 \sin^2 \alpha$ . El segmento  $l_1 \varepsilon_1$  es  $v_1$ , que por lo tanto sigue una distribución normal centrada en

$$\varepsilon_1 - \varepsilon_1 \sin^2 \alpha = \varepsilon_1 \cos^2 \alpha,$$

por lo que  $r_i = \cos^2 \alpha_i$ .

Sea  $\beta_i = 90^\circ - \alpha_i$  el ángulo que el eje  $\varepsilon_i$  forma con el subespacio  $\Omega$  (definido como el mínimo ángulo entre el eje y las rectas de  $\Omega$ ). Puede demostrarse que  $\sum \cos^2 \beta_i = n$ , siendo  $n$  la dimensión del subespacio. Entonces

$$\sum r_i = \sum \cos^2 \alpha_i = m - \sum \sin^2 \alpha_i = m - \sum \cos^2 \beta_i = m - n = r.$$

Vemos que si  $\Omega$  está próximo a ser perpendicular al eje  $\varepsilon_i$  entonces  $r_i$  es próximo a 1, mientras que si está próximo al propio eje entonces  $r_i$  es muy pequeño. Si el eje estuviese contenido en  $\Omega$  significaría que cualquier valor de  $l_i$  es posible independientemente del valor de las demás observaciones, y entonces  $r_i = 0$ , mientras que si es perpendicular entonces sólo es posible el valor  $l_i$ , lo que significa que se está observando una magnitud de valor conocido, y  $r_i = 1$ . El subespacio perpendicular a un  $\varepsilon_i$  tiene dimensión  $m - 1$ . El subespacio perpendicular simultáneamente a  $k$  ejes es la intersección de los perpendiculares a cada uno de ellos, y tiene dimensión  $m - k$ . Puesto que la dimensión de  $\Omega$  es  $m - r$ , sólo puede haber  $r$  observaciones con redundancia muy próxima a 1. Por otra parte sólo puede haber  $n$  ejes

contenidos en  $\Omega$ , por lo que sólo puede haber  $n$  observaciones con una redundancia muy próxima a cero. Estas dos propiedades están formuladas de manera más precisa por  $\sum r_i = r$ .

### Óptimos para la definición global de traslación, escala, etc.

Según se vio en el capítulo 15, cuando un sistema goza de una cierta libertad, por ejemplo libertad de giro, la aplicación de dicha libertad modifica las coordenadas de una manera que para pequeñas variaciones podemos aproximar linealmente. Así, el efecto de un giro grande o pequeño sobre las coordenadas de los puntos no es lineal, pero si el giro es muy pequeño es aproximadamente  $\propto (y, x)$  para un punto de coordenadas  $(x, y)$  respecto al centro de aplicación del giro. A estas constantes de proporcionalidad las llamamos  $\alpha$ . Los parámetros los denotamos todos de manera abstracta por la letra  $x$ .

$$(\Delta x_1, \Delta x_2, \dots, \Delta x_n) \propto \vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n). \quad (18.6)$$

Decir que una aplicación de la libertad del sistema provoca en los parámetros una variación proporcional a  $\vec{\alpha}$  equivale a decir que a los parámetros se les pueden sumar múltiplos arbitrarios de  $\vec{\alpha}$ . En el caso de traslación libre, por ejemplo,  $\vec{\alpha} = (1, 0, \dots, 1, 0)$  para la  $x$  y  $\vec{\alpha} = (0, 1, \dots, 0, 1)$  para la  $y$  (es este caso hay dos grados de libertad). En el de escala libre

$$\vec{\alpha} = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)_0,$$

en donde los valores no han de entenderse como variables, sino como valores fijos. Es lo que indica el subíndice 0.

Podemos representar el elipsoide de precisión de todo el conjunto de parámetros, definido por la matriz de varianzas  $\Sigma$ , en un sistema de  $n$  ejes de coordenadas que representan de manera abstracta a cada parámetro. Si por ejemplo los parámetros son las coordenadas  $(x, y, z)$  de 6 puntos habrá un total de 18 ejes. Este elipsoide tiene dimensión  $n - 1$ , y por tanto en el espacio  $n$ -dimensional es «plano». La razón es que al gozar el sistema de un grado de libertad, esta ha quedado fijada tomando como fijo algún valor. Por ejemplo, si se fija  $x_1 = 1000$ , el parámetro  $x_1$  no tiene variación y el elipsoide no se extiende en la dimensión del eje  $x_1$ . Podrá ser de dimensión  $n$  si se trata de un subconjunto de parámetros de un total mayor en el cual la libertad se fijó

## 18 Interpretación geométrica

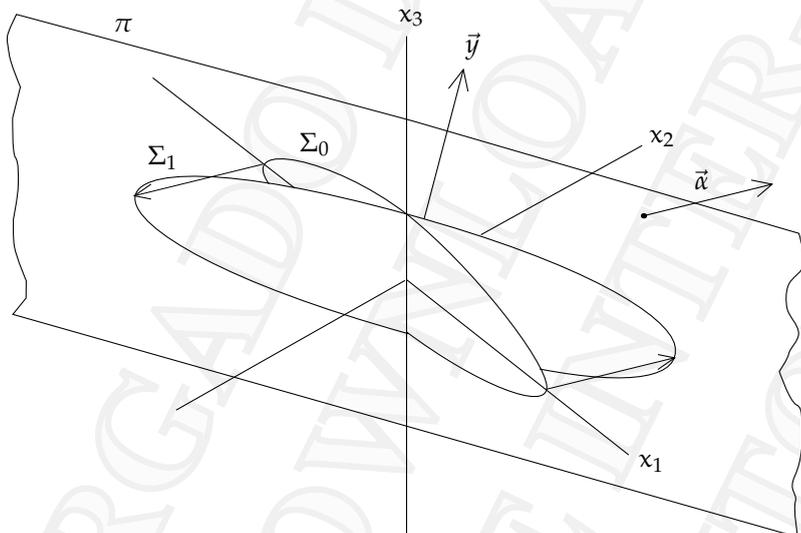


Fig. 18.4: Cambio de  $\Sigma$  de acuerdo a la manera de fijar la libertad del sistema

de una manera en la que los parámetros del subconjunto seleccionado no intervinieron. Por ejemplo, un subconjunto de puntos de una red fijada en posición mediante la coordenada de un punto que no pertenece al subconjunto en cuestión.

Para fijar el sistema de acuerdo a la libertad de la que goza se dará un valor a una cierta función de los parámetros,  $F(x_1, \dots, x_n)$ , que en el entorno del punto  $(x_1, \dots, x_n)_0$  (los valores que realmente tienen los parámetros en el caso concreto que nos ocupe) se aproximará a una relación lineal

$$k_1x_1 + k_2x_2 + \dots + k_nx_n = C. \quad (18.7)$$

Esta combinación lineal la podemos entender como una nueva variable  $y$ . Fijar su valor significa gráficamente que el elipsoide representado por  $\Sigma$  no se extiende en la dirección del vector

$$\vec{y} = (k_1, k_2, \dots, k_n).$$

En la figura 18.4  $\Sigma_0$  representa la matriz de varianzas de la que se parte. Se ha obtenido en un ajuste en el que se ha fijado el pará-

metro  $x_2$ , y por eso en la figura puede verse que el elipsoide no se extiende en esa dirección. El sistema de coordenadas se redefine de acuerdo a una condición que una vez linealizada en es como (18.7), por lo que la nueva matriz de varianzas no tendrá extensión en la dirección del vector  $\vec{y} = (k_1, k_2, \dots, k_n)$ , lo que significa que estará contenida en el espacio de dimensión  $n - 1$  perpendicular a  $\vec{y}$  (en la figura es  $\pi$ ). Como la aplicación de la libertad del sistema da lugar a una variación de los parámetros proporcional a  $\vec{a}$ , de acuerdo a (18.6), la definición de la misma según una u otra condición puede dar lugar a una mayor o menor precisión en la componente según  $\vec{a}$  del vector de parámetros  $(x_1, \dots, x_n)$ , de modo que la matriz  $\Sigma$ , o el elipsoide que la representa, puede desplazarse en esa dirección pero no en otra. El resultado de esto es que la matriz  $\Sigma_1$  según la nueva definición del sistema de coordenadas es la proyección sobre  $\pi$  del elipsoide  $\Sigma_0$  según la dirección de  $\vec{a}$ .

Para obtener el elipsoide más pequeño habrá entonces que proyectar  $\Sigma_0$  sobre un espacio  $\pi$  perpendicular a  $\vec{a}$ , lo que implica que  $\vec{y}$  ha de tomarse igual a  $\vec{a}$ . El nuevo elipsoide será igual al original restando (eliminando) su componente según la dirección de  $\vec{a}$ . La transformación que elimina la componente  $\vec{a}$  de un vector  $(x_1, \dots, x_n)$  se puede obtener como sigue: Se gira el sistema de coordenadas de manera que el eje  $x_1$  pase a coincidir con la dirección de  $\vec{a}$ ; posteriormente se hace la  $x_1 = 0$ , y por último se vuelve a girar el sistema para reponerlo en su orientación original. Sea para ello  $Q$  una matriz ortogonal con  $\vec{a}$  como primera fila. Tomaremos los valores  $(\alpha_1, \dots, \alpha_n)$  de manera que el módulo de  $\vec{a}$  sea igual a 1. Con esto se tiene

$$Q\vec{a} = (1, 0, 0, \dots, 0),$$

lo que significa que la que originalmente era la dirección  $\alpha$  es ahora la dirección  $x_1$ . Hacer 0 ahora la primera coordenada significa multiplicar por una matriz igual que la identidad pero con un 0 en la posición 1,1, lo que equivale a eliminar directamente en  $Q$  la primera fila. Por último se multiplica por  $Q^{-1} = Q^T$  para deshacer el giro:

$$Q^T \left( Q - \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 0 \end{pmatrix} \right) = I - \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_n \end{pmatrix} (\alpha_1 \ \alpha_2 \ \dots \ \alpha_n).$$

## 18 Interpretación geométrica

Por tanto la solución para la matriz  $A$  óptima para el cambio en la definición del sistema de coordenadas de acuerdo a la libertad del sistema es

$$A = I - M = I - (\alpha_i \alpha_j), \quad \sum \alpha_i^2 = 1. \quad (18.8)$$

Si el sistema goza de más de una libertad la aplicación de cada una de ellas dará lugar a variaciones en las coordenadas proporcionales cada una en primer orden a un cierto vector:

$$\begin{aligned} &(\alpha_1, \alpha_2, \dots, \alpha_n), \\ &(\beta_1, \beta_2, \dots, \beta_n), \\ &\dots \\ &(\lambda_1, \lambda_2, \dots, \lambda_n). \end{aligned}$$

Si estos vectores son todos perpendiculares entre sí, lo que significa que las distintas libertades son independientes, la matriz  $\Sigma$  mínima se consigue eliminando sucesivamente las componente según la dirección de  $\vec{\alpha}, \vec{\beta}, \dots, \vec{\lambda}$ :

$$A = I - (\alpha_i \alpha_j) - (\beta_i \beta_j) - \dots - (\lambda_i \lambda_j).$$

Sea

$$\Gamma = \begin{pmatrix} \alpha_1 & \beta_1 & \dots & \lambda_1 \\ \alpha_2 & \beta_2 & \dots & \lambda_2 \\ & & \dots & \\ \alpha_n & \beta_n & \dots & \lambda_n \end{pmatrix}.$$

La anterior matriz  $A$  es

$$A = I - \Gamma^T \Gamma.$$

Si las libertades no son todas independientes entre sí la eliminación de todas sus componentes, o lo que es lo mismo, la proyección sobre el espacio  $\pi$  perpendicular a todas ellas, viene dada por (18.4). En este caso:

$$A = I - \Gamma(\Gamma^T \Gamma)^{-1} \Gamma^T. \quad (18.9)$$

# 19 Estimadores para otras distribuciones, I

En los capítulos precedentes nos hemos ocupado del estudio de diversos problemas de estimación bajo la suposición de que las observaciones siguen distribuciones normales. Algunas veces la diferencia de la distribución de las observaciones respecto a una normal es tan grande que los resultados de la estimación mínimo cuadrática no son correctos, y se hace necesario buscar los estimadores centrados de acuerdo a las distribuciones que siguen las observaciones.

El primer paso es obtener la distribución. No es una tarea sencilla, pero es la parte esencial del proceso, ya que la distribución determina el estimador. Es necesario disponer de muchos datos, miles de observaciones. La parte más característica de las distribuciones, y que más influye en el estimador resultante, son los extremos. Pero son la parte más difícil de obtener porque contienen menos observaciones. Supongamos por ejemplo un conjunto de 1000 observaciones. En la zona de la distribución que contiene el 2% extremo de los errores habrá sólo unas 20, insuficiente para obtener la distribución.

Normalmente los datos existen en gran cantidad, pero suele suceder que quien se ocupa de diseñar estimadores no dispone de ellos, y quien sí los tiene no se preocupa de su análisis.

## **Obtención de valores siguiendo la distribución de los errores**

Idealmente los datos deben de ser errores, pero es muy raro tener la suerte de conocer los valores verdaderos. En un primer momento

## 19 Estimadores para otras distribuciones, I

parece que lo siguiente más deseable son varios conjuntos de observaciones, cada uno de los cuales consista en la repetición muchas veces de la medida de una magnitud. De esta manera el valor verdadero de cada magnitud se puede determinar con bastante precisión, y es casi como si lo conociésemos. Pero a veces esto no es así, ya que varias repeticiones de una misma medida pueden estar muy correladas. Si el conjunto de errores de observación sigue una distribución  $f$ , los errores en las repeticiones de una medida puede ser que sigan esa misma distribución o no. Esto último sucede si en las repeticiones de una observación hay una componente constante, sistemática, y entonces la distribución de esos errores está centrada en un valor  $\varepsilon_1$  (desconocido), y su variabilidad es menor que la de  $f$ , ya que la componente  $\varepsilon_1$  es constante. Esta componente puede ser un error de identificación (puntería) debido al propio objeto visado, por ejemplo debido a que una parte está más iluminada que la otra y no apreciamos correctamente el centro, o a que estamos midiendo un punto erróneo (otra esquina, otra mata). En el conjunto total de observaciones cada una tendrá una componente  $\varepsilon_{i1}$ , que será simplemente una componente más del error. Algunas veces será positiva, otra negativa, otras mayor y otras menor, y junto con los demás errores da lugar a  $f$ . Entonces hay veces en que no se puede obtener el valor verdadero mediante un estimador centrado a partir de las medidas repetidas, ya que en su lugar obtendríamos  $l + \varepsilon_1$ , y los errores estimados serían en general menores, al no incluir la componente  $\varepsilon_1$ .

En consecuencia, muchas veces sólo se pueden conseguir estimaciones de los errores como resultado de un ajuste, lo que complica mucho la obtención de su distribución, ya que las propias estimaciones de los errores se basan en el estimador, que a su vez se basa en la distribución supuesta. Por ello los primeros resultados son sólo provisionales. Se obtiene una primera distribución, en función de ella se determina el estimador y se realizan los ajustes, y las estimaciones de los errores determinan otra distribución más afinada, y así sucesivamente. Es un proceso que requiere muchos datos y lleva tiempo.

La estimación de los errores de cara a obtener la distribución no es la misma que las estimaciones centradas. Vimos que si una observación tiene una redundancia pequeña entonces la estimación centrada del error (el residuo) es en general menor que el error. Exactamente  $E[v_j]_{\varepsilon_i} / \varepsilon_i = r_i$ , por definición. Podemos entonces pensar en sustituir

## 19 Estimadores para otras distribuciones, I

$v$  por  $E[|\varepsilon|]$ . Para el estimador mínimo cuadrático este valor es

$$E[|\varepsilon|] = \sigma_\varepsilon \sqrt{\frac{2}{\pi}} e^{-\frac{v^2}{2\sigma_\varepsilon^2}} + |v| \int_{-\frac{v}{\sigma_\varepsilon}}^{\frac{v}{\sigma_\varepsilon}} N_{(0,1)}(x) dx.$$

Más sencillo es  $E[\varepsilon^2]$ . Sea cual sea el estimador, si  $v$  es una estimación centrada entonces  $g(\varepsilon)$  es una distribución centrada en  $v$ , cuya varianza  $\sigma_\varepsilon$  habremos estimado de manera centrada como  $\sigma_\varepsilon$ . Entonces,

$$\begin{aligned} E[\varepsilon^2] &= E[(v + (\varepsilon - v))^2] \\ &= v^2 + 2vE[\varepsilon - v] + E[(\varepsilon - v)^2] = v^2 + 0 + \sigma_\varepsilon^2, \end{aligned}$$

de donde

$$\sqrt{E[\varepsilon^2]} = \sqrt{v^2 + \sigma_\varepsilon^2}. \quad (19.1)$$

Pero tanto este estimador como  $E[|\varepsilon|]$  no responden al problema de obtener la distribución de los errores. Una vez más se hace necesario formular el problema con precisión. Queremos obtener la distribución de los errores, y lo haremos mediante ciertas estimaciones  $\tilde{\varepsilon}$ . Por lo tanto lo que queremos exactamente es que esos  $\tilde{\varepsilon}$  sigan la distribución de los errores. Entonces

(19.2) *Hay que encontrar un estimador  $\tilde{\varepsilon}$  tal que  $f_{\tilde{\varepsilon}} = f_\varepsilon$ .*

Es entonces un problema de tipo directo. Podríamos haber escrito  $f(\tilde{\varepsilon}) = f(\varepsilon)$ , pero esta notación, de la que hemos hecho uso alguna vez, puede ser ambigua, ya que empleamos la letra  $f$  para dos distribuciones distintas. Cuando sea necesario especificar a qué función  $f$  nos referimos emplearemos un subíndice, y si no hay lugar a ambigüedad indicaremos la variable entre paréntesis como hasta ahora. La solución del problema para el estimador mínimo cuadrático es muy sencilla, ya que sabemos que  $v_i \sim N(0, \sqrt{r_i}\sigma_{\varepsilon_i})$ , y entonces  $v_i/\sqrt{r_i}$  es el estimador buscado

$$\tilde{\varepsilon}_i = \frac{v_i}{\sqrt{r_i}} \sim N(0, \sigma_{\varepsilon_i}). \quad (19.3)$$

Claro que si estamos buscando la distribución de los residuos es precisamente porque ésta no será una normal, y entonces no se aplicará el estimador mínimo cuadrático. Un cierto error seguirá (habremos

## 19 Estimadores para otras distribuciones, I

supuesto) una distribución  $f_\varepsilon$ . A partir de esta distribución, de las de los demás errores y de la fórmula que permite calcular  $v$  podemos obtener la distribución  $f_v$ . Habrá que modificar el valor de  $v$  de manera que el resultante siga la distribución  $f_\varepsilon$ . Se puede demostrar que una solución es

$$\tilde{\varepsilon} = F_\varepsilon^{-1}(F_v(v)), \quad (19.4)$$

en donde  $F$  es la función de distribución, integral de la función de densidad.

Esta fórmula es difícil de aplicar de manera exacta en la práctica, porque la distribución  $f_v$  suele ser muy difícil de hallar, aunque sí es factible obtener una distribución aproximada mediante simulación, como veremos más adelante. Volveremos sobre este problema al final del próximo capítulo, cuando hayamos estudiado la relación de los residuos con los errores.

### El parámetro $\tau$

Sea como fuere, supongamos que tenemos un conjunto de valores  $\tilde{\varepsilon}$ . Tendremos que «normalizarlos»,  $\tilde{\varepsilon}' = \tilde{\varepsilon}/\sigma_{\tilde{\varepsilon}}$ . Pero aquí surge una primera diferencia importante respecto a la distribución normal u otras similares. Cuando la densidad de probabilidad no disminuye tan rápido como en una distribución normal el valor de  $\sigma$  puede hacerse muy grande (puede incluso ser infinito), y deja de ser útil como indicador de precisión y como valor de referencia. Por ello es necesario buscar otra magnitud como definición de la precisión, a la que llamaremos  $\tau$ . Una definición muy útil es la siguiente:

(19.5)  $\tau$  es el valor tal que el intervalo  $[-2\tau, 2\tau]$  contiene el 95,5% de probabilidad.

Tomaremos  $\tilde{\varepsilon}' = \tilde{\varepsilon}/\tau_{\tilde{\varepsilon}}$ . Será necesario entonces obtener una estimación de  $\tau$  para cada  $\tilde{\varepsilon}$ , problema que estudiaremos más adelante.

### Agrupación de los datos en intervalos

Dado el conjunto de valores  $\tilde{\varepsilon}'$ , a los que de ahora en adelante llamaremos simplemente  $\tilde{\varepsilon}$ , tenemos que ajustar una distribución a ellos. Ajustar directamente una función de densidad es muy difícil, ya

## 19 Estimadores para otras distribuciones, I

que hay que preocuparse de que la integral no sea infinito, entre otros detalles. Es más fácil calcular primero la función  $F$ . Por otra parte, normalmente  $f$  será simétrica, y entonces podemos trabajar con  $|\tilde{\varepsilon}|$ . Su función de distribución es  $\mathcal{F} = 2F - 1$ . Podemos escribir  $\tau$  en función de  $\mathcal{F}$ :

$$\tau = \mathcal{F}^{-1}(0,955). \quad (19.6)$$

Para obtener  $\mathcal{F}$  agruparemos los datos en intervalos. A la hora de contar el número de valores cada error no cuenta como 1, sino como su redundancia parcial; es decir, que el número de valores en un intervalo  $I$  ha de tomarse como  $\sum_{|\tilde{\varepsilon}| \in I} r$ . La lógica de esta manera de contar los errores se pone de manifiesto con un ejemplo. Supongamos que hemos medido los cuatro ángulos de un cuadrilátero y efectuado el ajuste, obteniendo cuatro residuos iguales. Tendremos 4 valores de  $|\tilde{\varepsilon}|$  iguales, así que realmente sólo hemos estimado «1» residuo. Cada  $|\tilde{\varepsilon}|$  cuenta como  $1/4$ . Sin embargo, si hemos medido cuatro veces una misma magnitud (sin error sistemático en las medidas), entonces tenemos cuatro residuos, pero sabemos, por ejemplo si hemos tomado la media, que su suma es igual a 0, o si hemos tomado la mediana que uno de ellos vale cero. Sea cual sea el estimador aplicado, los cuatro residuos cumplirán una condición, con lo que realmente hemos estimado 3. Debe entenderse que de la información que son  $m$  observaciones,  $n$  (no unas  $n$  observaciones en particular, sino, digamos, una cantidad  $n$  de información) se emplea en la estimación de los parámetros, y queda  $r$  para estimar residuos.

Para valores pequeños de  $|\tilde{\varepsilon}|$  tomaremos el número de intervalos que se requiera para la adecuada construcción de  $\mathcal{F}$ . Hacia los valores más altos llegará un momento en el que no podremos tomar los intervalos tan finos como sería deseable, y tendremos que limitarnos a agruparlos según su cantidad permita. En la formación de los intervalos interviene mucho la experiencia y la manera de tomarlos de cada persona. Un ejemplo puede ser el siguiente:

$\frac{x}{\mathcal{F}}$	0	0,2	0,4	0,6	0,8	1	
	1	0,789	0,599	0,448	0,328	0,238	
$\frac{x}{\mathcal{F}}$	1,2	1,4	1,6	1,8	2	2,5	3,2
	0,172	0,120	0,088	0,062	0,045	0,020	0,011

Este es un ejemplo real, con aproximadamente 12 000 valores (es

## 19 Estimadores para otras distribuciones, I

decir,  $\sum r = 12\,000$ ). Se muestra  $\bar{\mathcal{F}} = 1 - \mathcal{F}$  porque es más fácil trabajar con esta función que con  $\mathcal{F}$ , ya que para valores grandes de  $x$  (la variable) lo significativo es la diferencia a 1 de  $\mathcal{F}$ . También resulta más fácil encontrar funciones que tiendan a 0 que funciones que tiendan a 1. Estas últimas suelen tener la forma  $1 - k$ , en donde  $k$  es una función que tiende a cero. Una excepción importante es la función arcotangente. Puede observarse que  $\bar{\mathcal{F}}(2) = 0,045$ , de acuerdo a la definición de  $\tau$ .

### Ajuste de una función de distribución

Una vez que tenemos los intervalos hay que buscar una función  $\bar{\mathcal{F}}$  que se ajuste a ellos. Esta función debe cumplir una serie de características:

1.  $\bar{\mathcal{F}}(0) = 1$ .
2.  $\lim_{x \rightarrow \infty} \bar{\mathcal{F}}(x) = 0$ .
3. Es estrictamente decreciente.
4.  $\bar{\mathcal{F}}'$  es estrictamente creciente ( $\bar{\mathcal{F}}'' > 0$ ).
5. A partir de un cierto punto la segunda derivada es decreciente ( $\bar{\mathcal{F}}''' < 0$ ).
6.  $\bar{\mathcal{F}}''(0) = 0$ .

De estas condiciones sólo las dos primeras y la de ser decreciente, no necesariamente de manera estricta, son las que tiene que cumplir la función para poder ser una  $\bar{\mathcal{F}}$ . La imposición de que sea estrictamente decreciente significa que no puede haber intervalos con probabilidad cero, y las otras tres condiciones las cumplen las distribuciones cuya función de densidad tiene forma de campana.

Para ajustar una función  $\bar{\mathcal{F}}$  buscaremos entre funciones conocidas. Algunas funciones convexas ( $\bar{\mathcal{F}}'' > 0$ ) decrecientes que tienden a 0 son las siguientes:

$$1 - \frac{2}{\pi} \arctan x, \quad e^{-q(x)}, \quad \frac{1}{q(x)}.$$

## 19 Estimadores para otras distribuciones, I

La segunda y la tercera de ellas para algunos polinomios  $q$ . De estas tres sólo cumplen todas las condiciones la primera y algunos casos particulares de las otras dos.

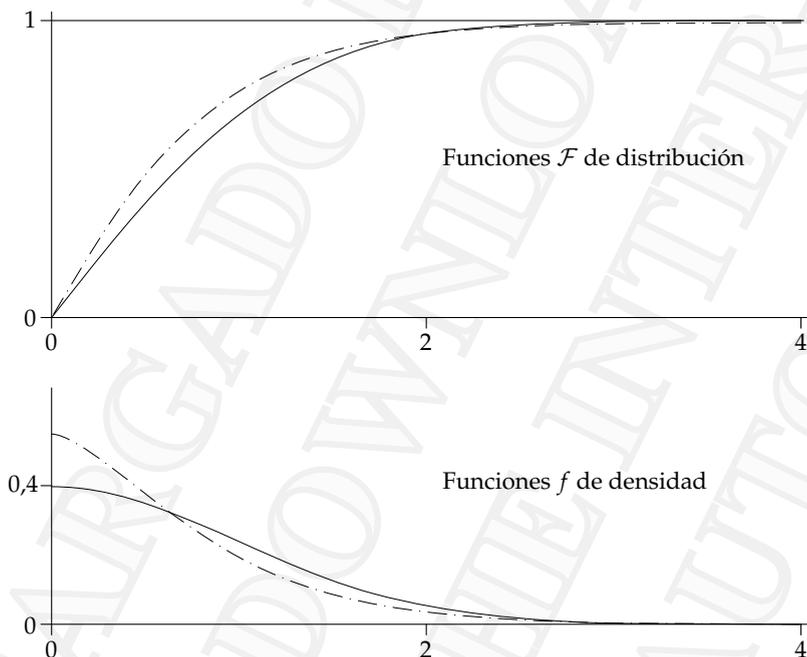
En lugar de buscar una función que cumpla todas las condiciones y se adapte a los datos en todo el recorrido puede ser más sencillo construir una función a tramos. La función resultante por supuesto tiene que ser continua. Es importante que la primera derivada también sea continua, y por cuestiones de tipo práctico también conviene que sea continua la segunda. Esto aumenta el número de parámetros necesarios, ya que en cada punto unión de dos tramos habrá tres parámetros que se necesiten sólo para cumplir la continuidad de  $\overline{F}$ ,  $\overline{F}'$  y  $\overline{F}''$ . En este ejemplo concreto se ajustó el tramo  $[0 - 3,2]$  mediante una función del tipo  $\exp(-(ax + bx^2 + cx^3 + dx^4))$ . Cuando buscamos una distribución distinta de la normal suele ser, como así era en este caso, porque el número de observaciones con errores grandes es mayor que en la normal, es decir, que  $\overline{F}$  no se aproxima tan rápidamente a 0. Por ello no es correcto que el tramo final de  $\overline{F}$  venga dado por una función exponencial. Podemos emplear una función del tipo  $1/x$  o  $1/x^2$ , por ejemplo. En este caso se tomó

$$\frac{b}{(c+x)^a}.$$

Los tres parámetros quedan determinados por la continuidad de  $\overline{F}$  y sus derivadas en el punto 3,2 (la función  $b/(c+ax)$  no sirve porque tiene sólo dos parámetros, ya que siempre podemos tomar  $a = 1$ ). No quedan parámetros para ajustar la distribución a los datos, pero es que tampoco hay intervalos por encima de 3,2. La única finalidad de esta función era construir una cola adecuada. Lo deseable sería disponer de más datos y poder tener intervalos fiables hasta por lo menos  $4,5\tau$ , y poder así construir una cola que se ajustase realmente a esta distribución, y no como se hizo en esta ocasión, tomar simplemente una cola estándar. Si hubiésemos querido un decrecimiento más lento podríamos haber tomado  $b/(a + \log(x+c))$ . Posteriormente el punto de unión de la cola al tramo principal se pasó al punto 3, por cuestiones de poca importancia. Los detalles del ajuste se encuentran en el libro de ejercicios.

Para que la función exponencial cumpla  $\overline{F}''(0) = 0$  tiene que ser  $a^2 = 2b$ . Puede imponerse esta condición o ajustar el tramo inicial

## 19 Estimadores para otras distribuciones, I



mediante otra función. En este caso se optó por la segunda opción, ajustando una función adecuada en el tramo  $[0,0,1]$ . Las funciones más sencillas que sirven para el tramo inicial son

$$1 - \arctan x, \quad 1 - \operatorname{sen} x, \quad 1 - x + x^3.$$

Como además se incluirán constantes en ellas, por ejemplo  $1 - ax + cx^3$ , los tres tipos de funciones son casi idénticos. En este caso se eligió el tercero. Son necesarios al menos tres parámetros para la continuidad en el punto de unión con el siguiente tramo, y se tomaron  $a - bx + cx^3$ . Al resolverlo el valor  $a$  no resultó exactamente igual a 1, por lo que se multiplicó toda la función  $\overline{\mathcal{F}}$  por  $1/a$ .

En esta página se muestra la comparación de las funciones  $\mathcal{F}$  y  $f$  de una distribución normal y otra empírica en la que  $\mathcal{F}$  no se acerca tan rápidamente a 1. En ambas se toma  $\tau = 1$ . En la normal  $\tau = \sigma$ . La distribución empírica se muestra con trazo discontinuo. Las funciones de distribución se cruzan en  $x = 2$ , de acuerdo a la definición

## 19 Estimadores para otras distribuciones, I

de  $\tau$ . Puede verse que la distribución empírica contiene más probabilidad en el intervalo  $[-\tau, \tau]$ .

### **Función $g$ y estimador de máxima probabilidad**

Una vez que tenemos la función  $f$  el siguiente paso es hallar la función  $g$ , punto en el cual interviene ya un conjunto de observaciones concreto. Emplearemos las relaciones lineales aproximadas entre parámetros y observaciones:  $V = L - AX_{\Delta}$ . Denotaremos por  $(-a_i -)$  la fila  $i$  de la matriz  $A$ . Para una observación  $\ell_i$  tenemos

$$f_{x_1, \dots, x_n}(\varepsilon_1) = f(L_i - (-a_i -)X_{\Delta}),$$

La función  $g$  es

$$g \propto \prod_i f(L_i - (-a_i -)X_{\Delta}). \quad (19.7)$$

Para encontrar el estimador centrado para un parámetro  $x_i$  tendríamos que obtener la función  $g(x_i)$ . Esta función es, para cada valor de la variable  $x_i$ , la densidad de probabilidad acumulada en la función  $g(x_1, \dots, x_n)$ . Esto se traduce en la siguiente integral:

$$g(x_i) = \int \dots \int g(x_1, \dots, x_n) dx_1 \dots dx_n,$$

en donde las variables de integración son todas excepto  $x_i$ .

Rara vez se pueden calcular estas integrales de modo exacto. Los métodos numéricos crecen enormemente de complejidad con el número de variables, y en consecuencia están limitados a un número pequeño de ellas. Por ello no se suele calcular el estimador centrado y se emplea el de máxima probabilidad, que es mucho más fácil de obtener.

Hay que minimizar el producto (19.7), o lo que es lo mismo, maximizar la suma

$$\sum_i -\log(f(L_i - (-a_i -)X_{\Delta})). \quad (19.8)$$

La función  $-\log f$  la denotamos por  $\psi$  (cf. p. 53). La derivada de  $\psi$  respecto de cada uno de los parámetros la obtenemos componiendo

## 19 Estimadores para otras distribuciones, I

$d\psi/d\varepsilon$  con la derivada de  $\varepsilon$  respecto de cada parámetro.

$$\frac{\partial\psi(\varepsilon_i)}{\partial x_j} = \left(\frac{d\psi}{d\varepsilon}\right)_{\varepsilon_i} \frac{\partial\varepsilon_i}{\partial x_j} = \psi'(L_i - (-a_i - )X_\Delta)a_{ij}.$$

La derivada del sumatorio (19.8) respecto de cada parámetro ha de ser igual a cero, lo que genera el sistema de ecuaciones

$$\begin{aligned} \psi'(L_1 - (-a_1 - )X_\Delta)a_{11} + \dots + \psi'(L_m - (-a_m - )X_\Delta)a_{m1} &= 0 \\ &\vdots \\ \psi'(L_1 - (-a_1 - )X_\Delta)a_{1n} + \dots + \psi'(L_m - (-a_m - )X_\Delta)a_{mn} &= 0 \end{aligned} \quad (19.9)$$

Este sistema se resuelve a su vez linealizando. Para ello tomamos un punto aproximado, que por sencillez será  $X_{\Delta 0} = 0$ . El valor aproximado de  $\varepsilon$  es  $\varepsilon_{i0} = L_i$ , de modo que

$$\psi'(L_i - (-a_i - )X_\Delta) \approx \psi'(L_i) - \sum_j \psi''(L_i) \frac{\partial\varepsilon}{\partial x_j} \Delta x_j,$$

en donde las derivadas parciales son  $a_{ij}$ . La primera ecuación queda

$$\sum_i \psi'(L_i)a_{i1} - \left( \sum_i \psi''(L_i)a_{i1}a_{i1}\Delta x_1 + \dots + \sum_i \psi''(L_i)a_{i1}a_{in}\Delta x_n \right) \approx 0$$

y análogamente las demás.

El conjunto de ecuaciones es casi igual que el de las ecuaciones normales del método de mínimos cuadrados. Los sumatorios que multiplican a cada  $\Delta x_j$  son un elemento de la matriz  $A^T A$  multiplicado por  $\psi''(L_i)$ . Estas últimas cantidades las podemos incluir en los elementos  $a_{ij}$  multiplicando cada uno de ellos por  $\sqrt{\psi''(L_i)}$ . Es una multiplicación de la matriz  $A$  por filas, análoga a la que transforma  $A$  (or.) en  $A$  (no.). A la matriz resultante la llamaremos  $A'$ , y  $\Psi'$  será el vector columna  $\psi'(L_i)$ . Entonces el sistema de ecuaciones es

$$A'^T A' X_\Delta \approx A'^T \Psi'.$$

---

\*Una vez planteado el sistema las variables son siempre los valores ajustados, por eso empleamos ya  $\varepsilon$  en lugar de  $\varepsilon$ .

## 19 Estimadores para otras distribuciones, I

La matriz  $A^T A'$  es la homóloga de la  $N$  del método de mínimos cuadrados, y por esa razón la llamamos también  $N$ . Entonces

$$NX_{\Delta} \approx A^T \Psi', \quad X_{\Delta} \approx N^{-1} A^T \Psi'. \quad (19.10)$$

Ahora bien, a esta solución aproximada se llega a través de dos linealizaciones. La primera es la misma que la del método de mínimos cuadrados, y es la que, para un  $X_0$  próximo a los valores reales, permite sustituir a todos los efectos las relaciones exactas por  $V = L - AX_{\Delta}$ . La segunda sin embargo requiere una cercanía a la solución mucho mayor, porque atañe a la función  $\psi'$ , cuyas variaciones se producen en el orden de magnitud de los residuos. Es decir, que si el valor ajustado de un residuo es, por ejemplo,  $v = 0,6$ , no se puede empezar el proceso de linealización en un punto en el que  $v = 20$ , en general, ya que el proceso no convergerá. Por ello se calcula previamente una solución aproximada por el método de mínimos cuadrados, que estará mucho más próxima a la solución final que los valores iniciales.

### Funciones $\psi'$

Vamos a ver cuál es la función  $\psi'$  en función de  $\bar{\mathcal{F}}$ .

$$f = -\bar{\mathcal{F}}'/2, \quad \psi = -\log(-\bar{\mathcal{F}}') + \log 2, \quad \psi' = -\frac{\bar{\mathcal{F}}''}{\bar{\mathcal{F}}'}$$

Para valores negativos de la variable obtenemos  $\psi'$  como  $-\psi'(-x)$ . También se puede obtener  $\psi''$  en función de las derivadas de  $\bar{\mathcal{F}}$ , pero es más cómodo derivar directamente  $\psi'$ . Ahora vemos la razón de buscar funciones  $\bar{\mathcal{F}}$  con segunda derivada continua. Esta segunda derivada aparece en el sistema de ecuaciones (19.9), y de ser discontinua podría el sistema no tener solución, lo que equivale a que el máximo de probabilidad estaría, para alguna de las observaciones, en uno de los puntos de discontinuidad. En el sistema linealizado aparece  $\bar{\mathcal{F}}'''$  (en  $\psi''$ ), que no será continua en general, pero eso no supone ningún inconveniente. Puede comprobarse gráficamente que en la solución de una ecuación mediante un proceso iterativo linealizando no causa ningún problema que en algún punto la función tenga una discontinuidad en su derivada.

## 19 Estimadores para otras distribuciones, I

A continuación se muestra la función  $\psi'$  para algunas de las funciones más habituales en la construcción de distribuciones.  $q(x)$  o simplemente  $q$  denota una función cualquiera, pero pensando sobre todo en un polinomio.

$\overline{\mathcal{F}}$	$\psi'$	$\overline{\mathcal{F}}$	$\psi'$
$b \operatorname{sen} ax$	$a \tan ax$	$b \operatorname{arctan} ax$	$\frac{2a^2x}{1 + (bx)^2}$
$q(x)$	$-q''(x)/q'(x)$	$1/q(x)$	$\frac{2q'^2 - q''q}{qq'}$
$e^{-q(x)}$	$\frac{q'' - q'^2}{q'}$	$\frac{b}{(c+x)^a}$	$\frac{a+1}{c+x}$

### Solución alternativa. Función de peso

El sistema (19.9) se puede resolver mediante un método iterativo distinto. Las cantidades  $\psi'(v_i)$  las escribimos como  $\frac{\psi'(v_i)}{v_i}v_i$ , con lo que las ecuaciones quedan

$$\frac{\psi'(v_1)}{v_1}v_1a_{1j} + \dots + \frac{\psi'(v_m)}{v_m}v_ma_{mj} = 0.$$

Si las cantidades  $\psi'(v_i)/v_i$  fuesen constantes,  $c_i$ , este sistema sería el mismo que el del método de los mínimos cuadrados, multiplicando los pesos originales de las observaciones por  $c_i$ . Podemos no obstante tomar para esas cantidades un valor constante aproximado, tomando para  $v_i$  el valor aproximado  $L_i$ . La función  $\psi'(x)/x$  recibe el nombre de función peso, y se denota por  $p$ . Cumple  $p(-x) = p(x)$ . Dividimos cada fila de la matriz  $A$  y del vector  $L$  por  $\sqrt{p(L_i)}$ , obteniendo unas nuevas matrices  $A'$  y  $L'$ . Volvemos a llamar  $N$  a  $A'^T A'$ , y entonces

$$NX_{\Delta} \approx A'^T L', \quad X_{\Delta} \approx N^{-1} A'^T L'. \quad (19.11)$$

Será necesario realizar varias iteraciones. La convergencia es más lenta que en el método de linealización.

**Estudio de la función  $\psi$**

La solución de máxima probabilidad no es centrada, y habrá que revisar la solución para verificar que es correcta, especialmente las observaciones con residuo elevado y las que tienen parámetros en común con ellas. Es interesante estudiar el comportamiento del estimador en ejemplos sencillos y representativos.

En primer lugar hay que ver cómo es la función  $\psi$ , cuya suma hay que minimizar. Siempre tiende a infinito, ya que el valor de la función de densidad tiende a cero. En el estimador mínimo cuadrático es  $x^2$ , que tiende a infinito de manera cada vez más rápida. Por eso este estimador tiende a evitar residuos muy altos, en lógica con la distribución normal supuesta. La siguiente manera de tender a infinito de modo más lento es con derivada constante o bien tendiendo a un valor constante. Es el caso de  $f \propto \exp(-|x|)$ , que para un conjunto de medidas de una misma magnitud da como resultado la mediana. Para esta distribución la función  $\bar{F}$  es  $e^{-x}$ . Para cualquier  $\bar{F}$  con un decrecimiento asintótico menor (exactamente,  $\lim_{x \rightarrow \infty} \bar{F}/e^{-x} = 0$ ) la derivada de  $\psi$  tiende a cero, de donde viene la propiedad de su estimador asociado de «eliminar» observaciones para las que se ha estimado un error muy grande, y se obtiene un resultado prácticamente igual que si esas observaciones no existiesen. Por eso estos estimadores se llaman robustos.

En la figura 19.3 se muestra la gráfica de una función  $\psi$  de un estimador robusto. Vamos a analizar qué solución proporciona el es-

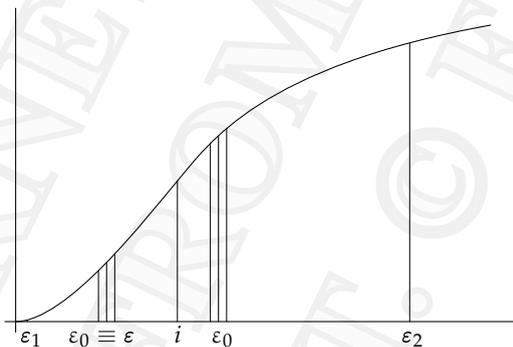


Fig. 19.3: Función  $\psi$  tipo para distribuciones reales

## 19 Estimadores para otras distribuciones, I

timador de máxima probabilidad para una función de ese tipo en el caso de dos observaciones  $\ell_1, \ell_2$  de una magnitud  $l$ . Supongamos  $\ell_1 > \ell_2$ , y sea  $\varepsilon_0 = (\ell_1 - \ell_2)/2$ . Si tomamos como solución la media aritmética entonces  $\varepsilon_1 = -\varepsilon_2 = \varepsilon_0$ , y tenemos

$$\psi(\varepsilon_1) + \psi(\varepsilon_2) = 2\psi(\varepsilon_0).$$

La solución es el punto que minimice  $\psi(\varepsilon_1) + \psi(\varepsilon_2)$ . Vamos a estudiar cómo se comporta esta cantidad cuando variamos  $l$  respecto a la media aritmética. Los valores absolutos de los errores serán  $\varepsilon_0 + \Delta\varepsilon$  y  $\varepsilon_0 - \Delta\varepsilon$ . Si  $\varepsilon_0$  es pequeño estará en la parte convexa de la función  $\psi$ . Entonces

$$\psi(\varepsilon_0 + \Delta\varepsilon) + \psi(\varepsilon_0 - \Delta\varepsilon) > 2\psi(\varepsilon_0),$$

de modo que la media aritmética será la solución. Si  $\varepsilon_0$  es muy grande estará en la parte cóncava de la gráfica, y ocurre lo contrario, que una separación respecto a la media aritmética disminuye la suma  $\psi(\varepsilon_1) + \psi(\varepsilon_2)$ . Si vamos separando cada vez más el valor  $l$  respecto a la media aritmética, para el residuo que aumenta en valor absoluto, supongamos  $\varepsilon_2$ , el valor de  $\psi'$  disminuye indefinidamente, tendiendo a cero. Para  $\varepsilon_1$  la derivada  $\psi'$  aumenta en un principio, pero tras pasar el punto de inflexión empieza a disminuir, y cuando el residuo tiende a cero  $\psi'$  también tiende a cero, si hemos tomado una función para la cual  $f'(0) = 0$ . Entonces llegará un punto para  $\varepsilon_1$  próximo a cero en el cual  $\psi'(|\varepsilon_1|) = \psi'(|\varepsilon_2|)$ . Si seguimos separando más el valor  $l$  entonces  $\psi(\varepsilon_1) + \psi(\varepsilon_2)$  ya no disminuirá, sino que aumentará, y entonces aquél punto es la solución. En la gráfica 19.3 se muestran ambos casos. Normalmente la disminución de la derivada a partir del punto de inflexión  $i$  es más lenta hacia valores mayores que hacia valores menores, lo que implica que sólo hay un mínimo para  $\varepsilon_0 \leq i$ .

El punto  $i$  es el error a partir del cual hay más probabilidad de cometer un error muy grande y otro pequeño que de cometer dos errores de igual magnitud y que sumen lo mismo que los anteriores, de acuerdo a la distribución supuesta, y por eso la solución más probable deja de ser la media aritmética. La función  $\psi$  más sencilla con una forma similar a la de la gráfica 19.3 es  $\log(1 + 50x^2)$ , que se corresponde con una función  $\mathcal{F}$  de tipo arcotangente. La constante 50 es la que hace  $\tau = 1$ . Para esta función el punto  $i$  es  $\approx 0,1414$ , lo que muestra los peligros de los estimadores de máxima probabilidad. Es más habitual valores de  $i$  entre 2 y 3.

## 19 Estimadores para otras distribuciones, I

Cuando  $\ell_1 - \ell_2$  es muy grande el estimador de máxima probabilidad expone dos soluciones indistintas, una próxima a  $\ell_1$  y otra próxima a  $\ell_2$ . Cuál de ellas se obtenga depende del valor inicial para el proceso iterativo. Podría incluso obtenerse el resultado de mínima probabilidad (la media), dado que también es un extremo relativo. Puesto que el problema es simétrico respecto a  $\ell_1$  y  $\ell_2$  el estimador centrado será siempre la media. Cuando hay más observaciones es más raro obtener un mínimo de probabilidad en lugar de un máximo. No obstante siempre puede haber pares de observaciones prácticamente independientes de las demás, y no es raro que la solución ajustada sea para algunas variables un mínimo relativo de probabilidad en lugar de un máximo. Por otra parte, puesto que en esos pares de variables no podemos saber cuál de las dos observaciones es la buena, la mejor solución será siempre tomar el punto medio, aunque sea un mínimo en lugar de un máximo. Partiendo de las relaciones lineales  $V = L - AX$ , el estimador mínimo cuadrático es la solución de un sistema de ecuaciones lineales, y es por tanto único. Sin embargo si empleamos otro estimador el sistema de ecuaciones ya no es lineal, y puede tener muchas soluciones, aunque normalmente estarán muy próximas entre sí, salvo en los casos de pares o conjuntos de puntos con una única redundancia, en los que si hay una observación con un error muy grande es imposible saber cuál de ellas es. La solución a la que se llegue depende del valor inicial aproximado. El método de la función de peso es menos proclive a obtener mínimos en lugar de máximos.

Supongamos ahora tres medidas en lugar de dos, y en particular el caso en el que dos medidas son iguales y la otra distinta. Ocurrirá algo similar al caso de dos medidas; la solución será aproximadamente la media aritmética hasta una cierta separación en las observaciones, punto a partir del cual el punto ajustado se aproxima rápidamente a las observaciones iguales. Puede verse que para las funciones  $\psi$  que estamos analizando la solución para errores pequeños está ligeramente más próxima a las medidas iguales que la media aritmética. Según va aumentando la separación entre las observaciones iguales y la distinta, el error ajustado para las medidas iguales aumenta cada vez más despacio, mientras que para la distinta lo hace cada vez más rápido; es decir, que el punto ajustado se separa cada vez más despacio de las medidas iguales. Cuando el residuo mayor llega al punto  $i$ , si aumentamos más la separación entre las observaciones el punto

## 19 Estimadores para otras distribuciones, I

ajustado empieza a acercarse a las iguales. Este caso no da lugar nunca al problema de dos máximos relativos y un mínimo intermedio.

Si las tres observaciones estuviesen muy separadas entre sí, lo que significaría que hay al menos dos observaciones con un error muy grande, entonces habría tres máximos relativos de probabilidad, próximos cada uno a una de las observaciones. Si se tiene un conjunto mayor de observaciones podrá haber más o menos observaciones con un error elevado. Si hay muy pocas en relación a las redundancias habrá una única solución de máxima probabilidad, si bien es una cuestión que depende también mucho de la configuración relativa de las observaciones con más error. Según vaya aumentando el número de observaciones con error elevado irán apareciendo varias soluciones localmente óptimas, siendo imposible discernir en torno a cuál de ellas se encuentran los valores verdaderos.

La función  $\psi'$  en su límite hacia infinito está siempre por encima de  $1/x$ . Se muestran a continuación funciones  $\bar{\mathcal{F}}$  que tienden a 0 muy despacio junto con sus correspondientes funciones  $\psi'$ :

$\bar{\mathcal{F}}$	$\psi'$
$\frac{1}{x}$	$\frac{2}{x}$
$\frac{1}{\log x}$	$\frac{1}{x} + \frac{2}{x \log x}$
$\frac{1}{\log \log x}$	$\frac{1}{x} + \frac{1}{x \log x} + \frac{2}{x(\log x) \log \log x}$

No existe ninguna función  $\bar{\mathcal{F}}$  para la cual  $\psi' = 1/x$ . Si fuese así entonces sería  $\psi = \log x$  y  $f \propto e^{-\psi} = 1/x$ , que no puede ser función de densidad porque su integral es infinito. Sin embargo esta función la hemos empleado con el significado de no saber nada para valores que pueden ser muy pequeños o muy grandes (la desviación típica), y eso mismo se puede interpretar para valores de errores grandes, debidos a equivocaciones o mal funcionamiento de aparatos (errores groseros), aunque en este caso realmente no es así; hay un valor de referencia, que es  $\tau$ , y parece claro que la probabilidad de cometer un error entre  $10\tau$  y  $20\tau$  es mayor que la de uno entre  $100\tau$  y  $200\tau$ , y este a su vez mayor que uno entre  $1000\tau$  y  $2000\tau$ .

**Descentralidad de la estimación de máxima probabilidad**

En problemas con muy pocos parámetros sí que es posible obtener la estimación centrada, y se puede comparar con la que se obtiene mediante la estimación de máxima probabilidad. A continuación se muestra un ejemplo basándose en una función  $\overline{F}$  del tipo siguiente:

$$\overline{F}(x) = \begin{cases} b \arctan ax, & x < 0,2. \\ ke^{-(ax+bx^2+cx^3)}, & 0,2 \leq x < 3. \\ \frac{b}{(x^2 + ax + c)^{0,4}}, & x \geq 3. \end{cases}$$

con los siguientes valores para las constantes:

$$b \arctan ax: \quad b \approx 0,7868, \quad a \approx 1,3550.$$

$$ke^{-(ax+bx^2+cx^3)}: \quad k \approx 1,002344, \quad a = 1,1, \quad b = 0,414, \quad c = -0,094.$$

$$\frac{b}{(x^2 + ax + c)^{0,4}}: \quad b \approx 0,007985, \quad a \approx -4,882, \quad c \approx 6,074.$$

Los detalles sobre su obtención y justificación se pueden encontrar en el libro de ejercicios. Puede comprobarse que  $\overline{F}(2) = 0,045$ . El decrecimiento asintótico de esta función es  $\sim b/x^{0,8}$ .

Se han tomado varios casos representativos, todos ellos consistentes en tres observaciones de una misma magnitud. La función  $g(x)$  es simplemente proporcional al producto  $\prod f(\ell_i - x)$ . La constante de proporcionalidad se obtiene como siempre evaluando la integral, y una vez obtenida ésta la estimación centrada viene dada por  $\int_{\mathbb{R}} xg(x)dx$ . Como la función  $f$  está definida a tramos, y es lo habitual, es un tanto pesado obtener los límites de cada tramo para la función  $g$  y su expresión exacta en cada uno de ellos, como se pudo ver en el capítulo 3 en el que se calculó para una distribución  $f$  triangular. Es mucho más sencillo obtener  $g(x)$  para un conjunto de valores  $x$  tomados a intervalos regulares, y de todas formas las integrales habrá que evaluarlas mediante métodos numéricos aproximados, así que no se pierde nada.

En las páginas siguientes se muestran las gráficas de las funciones  $g$  centradas en la solución de máxima probabilidad, y comparadas con una normal y la  $f$  de este problema, ambas para  $\tau = 1/\sqrt{3}$ .

19 Estimadores para otras distribuciones, I

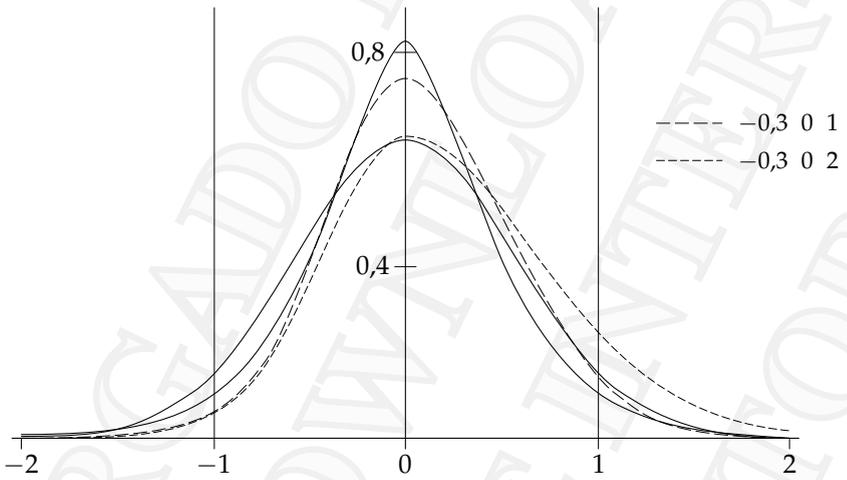


Fig. 19.4 : Funciones  $g$  para estimaciones de máxima probabilidad, I

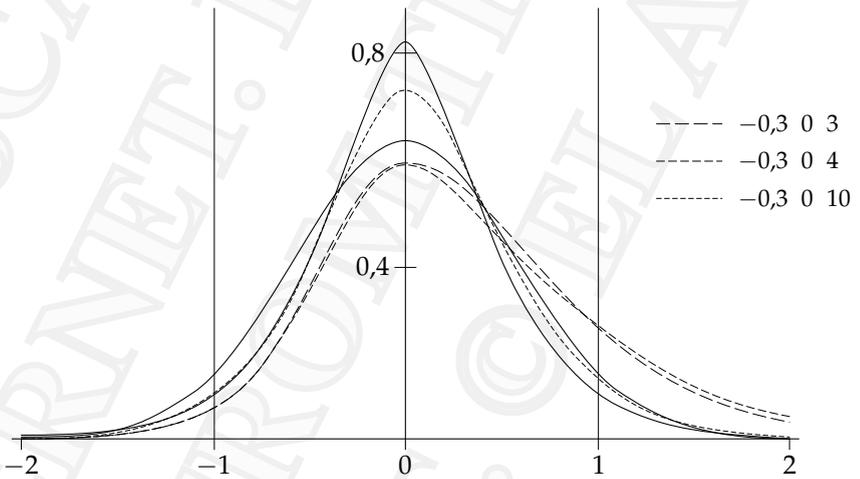


Fig. 19.5 : Funciones  $g$  para estimaciones de máxima probabilidad, II

19 Estimadores para otras distribuciones, I

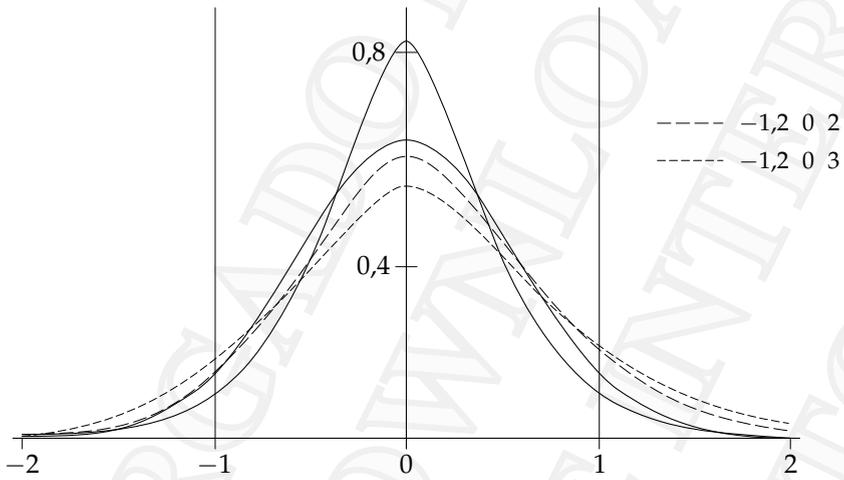


Fig. 19.6 : Funciones  $g$  para estimaciones de máxima probabilidad, III

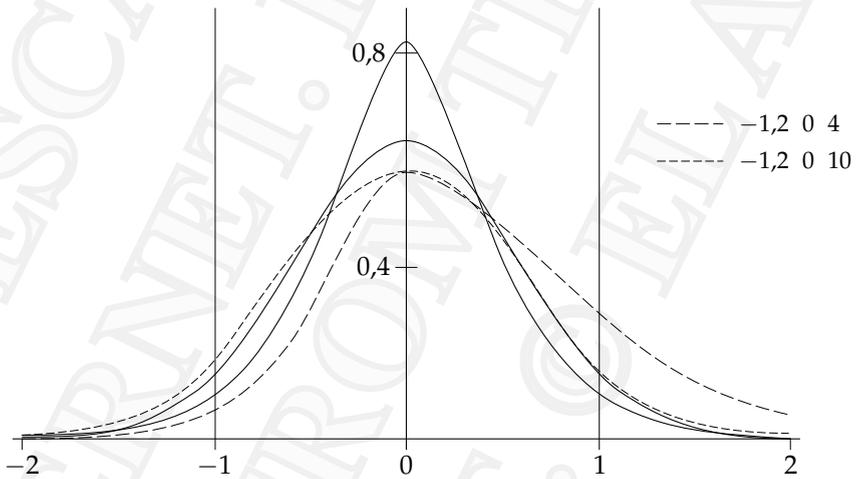


Fig. 19.7 : Funciones  $g$  para estimaciones de máxima probabilidad, IV

## 19 Estimadores para otras distribuciones, I

A continuación se muestra a la izquierda el conjunto de valores observados en cada caso y a la derecha la diferencia entre la estimación centrada y la solución de máxima probabilidad.

-0,3 0 1	0,081	-0,3 0 2	0,184
-0,3 0 3	0,247	-0,3 0 4	0,274
-0,3 0 10	0,038		
-1,2 0 2	0,082	-1,2 0 3	0,087
-1,2 0 4	0,311	-1,2 0 10	-0,022

Puede verse que el descentrado es apreciable. Los valores de  $\tau_x$  estarán en torno a  $1/\sqrt{3}$  ó  $1/\sqrt{2}$ , lo que significa en algunos casos un descentrado de  $\frac{1}{2}\tau_x$ .

### Diseño de un estimador más centrado

Habrá que buscar un estimador que proporcione resultados más centrados. La manera más sencilla consiste en resolver el problema mediante el método de la función de peso, pero modificando esa función. La función  $p$  para la distribución  $\overline{\mathcal{F}}$  que estamos empleando se muestra en la gráfica 19.8, junto con otras funciones. Analizando los casos anteriores se ve que es necesario reducir el peso de los residuos muy pequeños. La función  $p$  aumenta mucho según se aproxima a 0, y tendería a infinito si no fuese por el pequeño tramo en  $\overline{\mathcal{F}}$  de tipo arcotangente al principio. La función  $p$ , y en consecuencia la solución de máxima probabilidad, es muy sensible a variaciones muy pequeñas de  $\overline{\mathcal{F}}$  cerca del origen. A la función de pesos modificada la llamaremos  $p_e$  (función de pesos del estimador). Una solución que mejora considerablemente los resultados consiste en tomar para  $x < x_{\text{mín}}$  el valor  $p_e(x) = p(x_{\text{mín}})$ . Para el ejemplo que estamos siguiendo un buen valor de  $x_{\text{mín}}$  es 0,7. Normalmente la función  $p_e$  se podrá mejorar más mediante pequeños ajustes, y es un proceso que necesita tiempo. Una modificación habitual es la disminución del peso para los residuos muy grandes. En efecto, vemos que en el caso  $-1,2 \ 0 \ 10$ , cuyos residuos ajustados son  $-0,71 \ 0,49 \ 10,49$ , el descentrado es en sentido contrario. Al disminuir el peso para residuos pequeños este descentrado aumentará, y aparecerá uno con ese mismo sentido también en el caso  $-0,3 \ 0 \ 10$ .

## 19 Estimadores para otras distribuciones, I

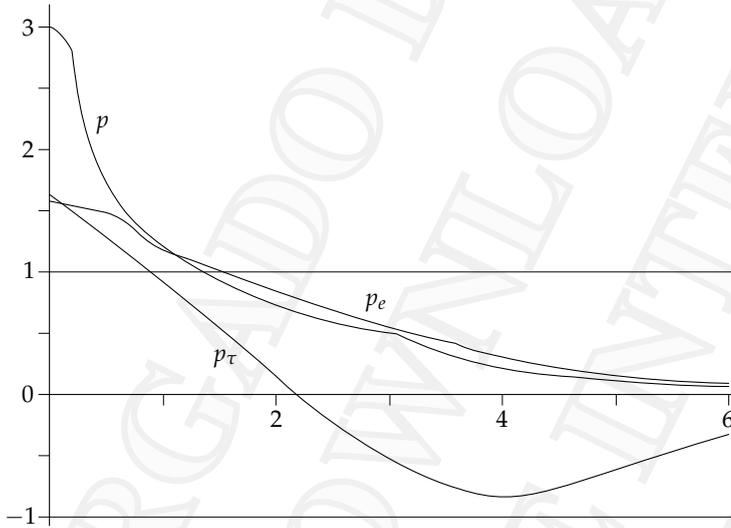


Fig. 19.8: Funciones de peso:  $p$ ,  $p_e$  y  $p_\tau$

La función  $p_e$  finalmente ajustada se muestra en la gráfica 19.8, comparada con la función original. En la comparación de estas gráficas, así como en el análisis de una función individualmente, hay que tener en cuenta que una función de peso la podemos multiplicar por una constante arbitraria. El último tramo de la función  $p_e$ , a partir de  $x = 6$ , es de tipo  $b/x^2$ . Los nuevos valores de descentrado son los siguientes:

-0,3	0	1	$\approx 0$	-0,3	0	2	0,007
-0,3	0	3	0,008	-0,3	0	4	0,016
-0,3	0	10	-0,001				
-1,2	0	2	-0,013	-1,2	0	3	-0,015
-1,2	0	4	-0,004	-1,2	0	10	0,003

Los resultados se pueden afinar más, pero hay que tener en cuenta que los casos seleccionados no son todos los posibles. De hecho algunos valores inicialmente conseguidos eran ligeramente mejores, pero finalmente se ajustó la función  $p_e$  teniendo en cuenta los casos de

## 19 Estimadores para otras distribuciones, I

error mayor intermedios entre 4 y 10, para llegar a los valores mostrados. El resultado de la estimación es muy sensible a pequeñas variaciones en la función de pesos, y la pequeña diferencia que hay entre  $p$  y  $p_e$  en gran parte de su dominio supone una mejora apreciable en el centrado de la estimación. El mayor descentrado en valor absoluto, incluyendo los valores no mostrados, es 0,016. Se ha reducido por tanto el descentrado máximo a la vigésima parte, estando ahora los valores muy por debajo de  $\tau_x/10$ .

Los casos estudiados son bastante representativos aunque se limiten a tres observaciones. Si se tienen por ejemplo cinco observaciones, normalmente podremos separarlas en dos conjuntos, uno de tres observaciones y otro de dos, de manera que las soluciones que proporciona el estimador para uno y otro estén muy próximas entre sí. Sean  $x_1$  y  $x_2$  esas soluciones. El valor  $x$  que se obtenga estará comprendido entre ambos. A su vez tanto  $x_1$  como  $x_2$  estarán próximos a las estimaciones centradas para cada conjunto, pues son los casos contemplados en el diseño de la función  $p_e$ , y la solución centrada estará entre ambas soluciones centradas, y en definitiva próxima a  $x$ . Cuando más aumenta el número de redundancias en relación a las observaciones más fácil es que un estimador, cualquiera, sea centrado. Podrá haber más parámetros y observaciones, relacionados de diversas maneras, pero muchos problemas son localmente aproximados a un pequeño número de parámetros con un número tampoco muy grande de observaciones.

La función  $p_e$  no da lugar a la estimación de máxima probabilidad para la función  $\overline{\mathcal{F}}$  de la que se ha derivado, pero sí para alguna otra función  $\overline{\mathcal{F}}$ .<sup>\*</sup> Entonces sucederá también que cuando el número de observaciones con un error grande es muy alto en relación a las redundancias el estimador encontrará varias soluciones, dependiendo del punto aproximado de partida. No obstante, cuando el

---

<sup>\*</sup>Vimos que una función de peso  $p$  no puede ser en su último tramo igual a  $1/x^2$ , ya que se correspondería con una función  $f$  de integral infinito. Sin embargo sí puede ser  $(1 + \varepsilon)/x^2$  para cualquier  $\varepsilon > 0$ , ya que entonces  $f \propto 1/x^{1+\varepsilon}$  en ese tramo, que tiene integral finita, y existe una constante de proporcionalidad tal que  $\int f = 1$ . Puesto que una función de peso la podemos multiplicar por cualquier constante, entonces si  $p$  tiene una cola de tipo  $b/x^2$  siempre será la función  $p$  de alguna función  $\overline{\mathcal{F}}$ . Lo que no se puede es diseñar una función  $\overline{\mathcal{F}}$ , derivar su función  $p$ , y el tramo final sustituirlo por  $1/x^2$ , porque entonces habrá que multiplicar  $p$  por alguna constante  $k > 1$  para obtener una función  $\overline{\mathcal{F}}$  de la que derive, lo que equivale a sustituir  $f$  por  $f^k$ , trastocando la función original.

número de observaciones es muy grande esta acumulación de errores grandes sólo se podrá dar en muy pocos lugares (exactamente,  $\binom{m}{k} \mathcal{F}(x)^k (1 - \mathcal{F})^{m-k}$  es la probabilidad de que  $k$  observaciones de entre  $m$  superen el error  $x$ ). Si no es así y los errores elevados son más abundantes de lo esperado, entonces es que el valor supuesto de  $\tau$  no es correcto, o tal vez toda la función  $f$ .

**Simulación del caso directo**

Con más de un parámetro se hace difícil calcular la estimación centrada y con ello la diferencia respecto a ella de la solución obtenida mediante  $p_e$ . Sin embargo es muy sencillo estudiar el proceso contrario, es decir, el directo, y partiendo de unos valores reales generar conjuntos de observaciones de acuerdo a la función  $F$ , y calcular los valores ajustados que se obtendrían. Si el número de parámetros es pequeño se puede generar de manera sistemática los conjuntos de observaciones siguiendo distribuciones  $F$ . Supongamos por ejemplo que estamos simulando un problema con tres parámetros y cinco observaciones. Dividimos el intervalo  $[0, 1]$  de manera regular en un número de valores, por ejemplo 30. Los valores serán  $\frac{1}{60}, \frac{3}{60}, \dots, \frac{59}{60}$ . Para cada uno de esos valores  $y$  calculamos  $F^{-1}(y)$ , y ese valor lo tomamos como  $l_1$ , según se muestra en la gráfica:

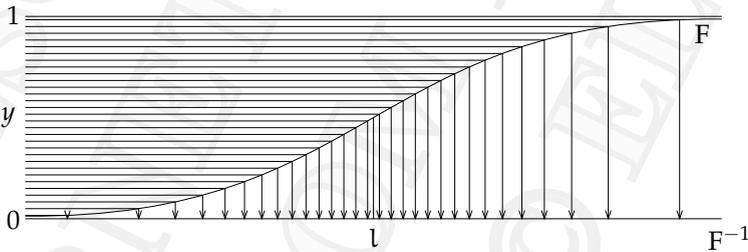


Fig. 19.9: Obtención de valores siguiendo una distribución  $F$

Realizamos esta operación para las cinco observaciones, y obtenemos un total de  $30^5 = 72\,900\,000$  ! combinaciones posibles de valores observados, para cada una de las cuales tendremos unos valores ajustados. Se pueden reducir a la mitad si nos damos cuenta de

## 19 Estimadores para otras distribuciones, I

que los resultados para unos valores  $(-\varepsilon_1, \dots, -\varepsilon_n)$  serán exactamente los mismos que para  $(\varepsilon_1, \dots, \varepsilon_n)$  pero cambiados de signo tanto los residuos como los errores en la estimación,  $x - \hat{x}$ , debido a la simetría de las funciones  $f$ . También debido esta simetría cualquier estimador, sea bueno o malo, será centrado en sentido directo, lo que muestra el poco valor que tiene el que un estimador sea centrado desde el punto de vista directo.

Cuando el número de observaciones es mayor está claro que es imposible proceder de este modo. Se puede pensar en reducir el número de valores considerados para cada observación, pero de hecho 30 ya es un número bastante pequeño, ya que el máximo error que se generará es  $F^{-1}(1/30)$ , y  $\frac{1}{30} \approx 0,033$ , por lo que apenas superamos  $2\tau$ . Una solución que se puede emplear para un conjunto mayor de observaciones y que además es mucho más realista consiste en dividir el intervalo  $[0, 1]$  en tantos intervalos regulares como observaciones se vayan a necesitar en total, es decir, el número  $m$  de observaciones multiplicado por el número de simulaciones, y los valores obtenidos como  $F_0^{-1}$ \* agruparlos de manera aleatoria de  $m$  en  $m$ . Para cada conjunto de  $m$  errores distribuimos sus valores entre las observaciones de manera también aleatoria, obteniendo así un conjunto de valores observados. El reparto aleatorio se puede efectuar antes o después de aplicar  $F_0^{-1}$ .

Si sólo vamos a realizar un pequeño número de simulaciones se puede obtener cada valor de error generando de manera aleatoria (se entiende que de acuerdo a una distribución uniforme) un número  $y$  entre 0 y 1 y tomando  $F^{-1}(y)$ . Finalmente, se pueden tomar los errores a discreción del analista con la finalidad de estudiar cómo se comporta el estimador en una circunstancia determinada.

En los análisis de las simulaciones hay que tener siempre presente que se trata de problemas directos, y no de problemas inversos que son los que realmente conforman el proceso de estimación. Es muy fácil tender a fijarse en casos formados por observaciones con errores pequeños salvo una con un error muy elevado, especialmente cuando los errores se deciden «a dedo», sin tener en cuenta que ese mismo conjunto de valores observados se podría haber obtenido para otros valores reales sin un error tan marcado en una observación. Una solución de cara al análisis consiste en agrupar las simulaciones según la

---

\* $F_0$  es la función  $F$  centrada en 0.

## 19 Estimadores para otras distribuciones, I

similitud de los valores  $\ell_i - \ell_j$ , pero sólo es factible obtener conjuntos representativos cuando el número de observaciones es muy pequeño. El anterior análisis sesgado favorece los estimadores de máxima probabilidad frente a los centrados. Así, volviendo a los ejemplos de tres observaciones de una magnitud, cuando los valores observados son  $-0,3 \ 0 \ 4$  la solución de máxima probabilidad que se obtuvo fue 0,05, mientras que la centrada fue 0,34. Los estimadores de máxima probabilidad tienden a proporcionar soluciones en las que aparentemente se han «detectado» y «eliminado» los errores grandes, dejando una solución muy limpia. Muchas veces será así, pero también muchas otras estará alejándose del valor real.

### Definición directa de la matriz de pesos

Cuando no se puede obtener  $\overline{\mathcal{F}}$ , por falta de datos suficientes, o bien aunque se pueda obtener porque quien debe hacerlo no sabe o no quiere, se define  $p_e$  directamente. Si se hace así hay que ser cautos en la elección de la función  $p_e$ , aunque muchas veces se hace exactamente lo contrario, y se escoge una función con una caída muy brusca en torno a un valor medio-alto, por ejemplo 3. Las funciones de este tipo, más aún que las de máxima probabilidad, tienden a ajustarse muy bien a un conjunto de observaciones ignorando las discrepantes con esa solución.

Siempre se podrá obtener un punto ajustado para el cual hay un conjunto relativamente amplio de residuos muy pequeños y unos cuantos residuos bastante elevados, pero esa no es la solución.



## 20 Estimadores para otras distribuciones, II

### Regiones de confianza. Introducción

Una vez obtenidos unos valores ajustados hay que calcular los intervalos de confianza y tal vez también la desviación típica de referencia, o más exactamente la medida de dispersión de referencia:  $\tau_0$ . En primer lugar supondremos que esta última no la estimamos, y es un valor conocido  $\tau_0$ . Los intervalos o regiones de confianza vienen dados por la integral de la función  $g$ , cuya obtención no será factible en muchas ocasiones. Habrá que buscar entonces alguna solución aproximada. El problema tiene dos partes. La primera es la estimación de los valores de  $\tau$ . Nos centraremos primero en los parámetros. Hay que obtener por tanto unos valores  $(\tau_{x_1}, \dots, \tau_{x_n})$ , y al ser posible una matriz  $T_{xx} = (\tau_{ij})$ , en la que  $\tau_{ii}^2 = \tau_i^2$ . Habrá que ver qué significado tienen los valores  $\tau_{ij}$  para  $i \neq j$ . La matriz no la llamaremos  $T_{xx}$  sino  $T_{xx}$ . Puesto que se tratará de una estimación deberíamos llamarla  $T_{xx}$ , pero como la matriz exacta es inalcanzable no hace falta distinguir entre ambas, y  $T_{xx}$  se referirá siempre a la matriz que hayamos obtenido. La segunda parte se refiere a qué tipo de distribución emplear para obtener los intervalos y regiones de confianza, puesto que la distribución exacta no la podemos obtener. Esta es una cuestión aparentemente más complicada pero que se resuelve de una manera muy sencilla.

### Obtención de $T_{xx}$ . Función $p_\tau$

Respecto a la primera parte del problema, la solución más sencilla

## 20 Estimadores para otras distribuciones, II

llo seguramente sea obtener  $T$  como  $A_\tau^T A'_\tau$  para alguna matriz  $A'_\tau$ . Por  $A'_\tau$  denotamos la matriz  $A$  divididas las filas por  $\sqrt{p_\tau}$  para alguna función  $p_\tau$ . Para encontrar una función  $p_\tau$  adecuada emplearemos una vez más como casos representativos los de tres observaciones de una magnitud, en los que podemos calcular el valor real  $\tau_x$ . De estos, aquellos en los que el residuo mayor está en torno a 10 pueden tomarse también como casos de dos observaciones, ya que esa observación ha quedado prácticamente eliminada por el peso  $p_e$  y apenas influye en el valor de  $\tau_x$ . A continuación se muestran varias ternas de observaciones con sus correspondientes  $\tau_x$ . Se muestran los residuos aproximados en lugar de los valores observados porque son aquellos los significativos, y sólo difieren de estos en un desplazamiento.

$v_1$	$v_2$	$v_3$	$\tau_x$	$v_1$	$v_2$	$v_3$	$\tau_x$
-0,5	0	0,5	0,47	-1	0	1	0,55
-0,5	0,2	0,8	0,50	-0,7	-0,4	1,6	0,56
-0,8	-0,5	2,5	0,70	-0,6	-0,3	3,6	0,76
-0,4	0,1	4,9	0,68	-0,3	0	6	0,62
-0,3	0	7	0,60	-0,2	0,1	10,1	0,59
-1,3	-0,1	1,9	0,68	-1,4	-0,2	2,8	0,79
-1,1	0,1	4,1	0,82	-0,9	0,3	5,3	0,72
-0,8	0,4	6,4	0,67	-0,7	0,5	10,5	0,65
-2	0	2	0,79	-2	0	3	0,96

Estos valores de  $\tau_x$  son los que hacen la probabilidad del intervalo  $[x - 2\tau_x, x + 2\tau_x]$  igual a 0,955. Es decir, que se ha tomado como centro del intervalo el punto  $x$  obtenido mediante la función  $p_e$  y no la estimación exactamente centrada. Debe hacerse así ya que los intervalos y regiones de confianza se darán centrados en los valores ajustados, y no en las estimaciones centradas que normalmente no conoceremos. Si empleásemos el valor ajustado de las estimaciones de máxima probabilidad algunos de los anteriores valores de  $\tau_x$  serían considerablemente mayores.

Se puede probar a tomar las funciones  $p$  y  $p_e$  para obtener la matriz  $A'_\tau$ , pero los resultados son mejorables. En la página siguiente se muestran algunas gráficas de funciones  $\mathcal{G}_x(x) = P\left(\frac{|x-x|}{\tau_x} \leq x\right)$  según esos valores de  $\tau_x$ . La función  $\mathcal{G}_x$  se obtiene a partir de la integral de la

20 Estimadores para otras distribuciones, II

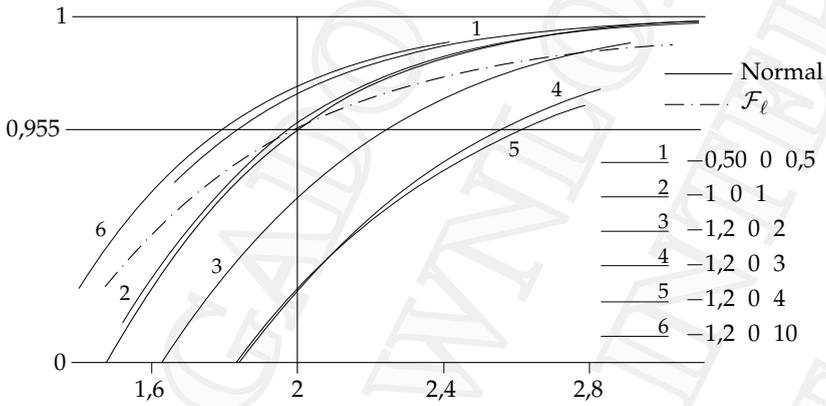


Fig. 20.1: Funciones  $G_x$  calculando  $A'_\tau$  según los pesos  $p$

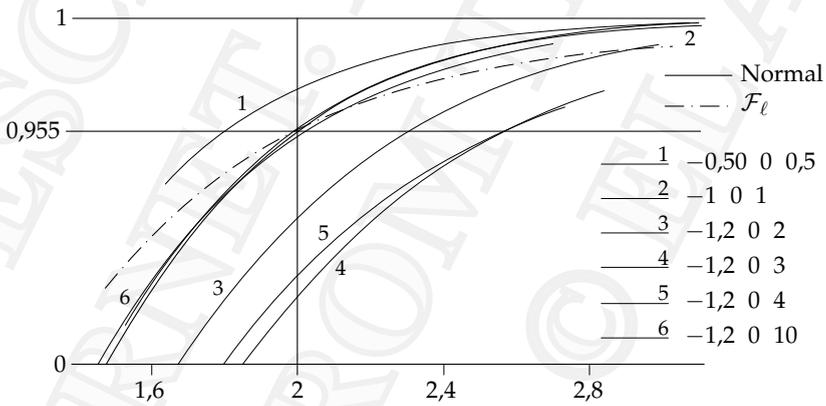


Fig. 20.2: Funciones  $G_x$  calculando  $A'_\tau$  según los pesos  $p_e$

## 20 Estimadores para otras distribuciones, II

función  $g$ . Por lo tanto si empleásemos el valor correcto de  $\tau_x$ , o sea  $\tau_x$ , tendríamos exactamente  $\mathcal{G}_x(2) = 0,955$ . Al emplear en el eje de abscisas la variable  $(x - x)/\tau_x$ , el punto  $\mathcal{G}_x^{-1}(0,955)$  será distinto de 2. Ese punto es  $2\tau_x$  pero en la escala del eje de abscisas, es decir,  $2\tau_x/\tau_x$ . Por lo tanto  $\mathcal{G}_x^{-1}(0,955)/2$  muestra el error en la estimación de  $\tau_x$ , mientras que  $\mathcal{G}_x(2)$  es la probabilidad realmente contenida en el intervalo  $[x - 2\tau_x, x + 2\tau_x]$ , que erróneamente tomaríamos como el 95,5%. En las gráficas se muestran también las distribuciones normal y  $\mathcal{F}_\ell$ .

En realidad no se han empleado exactamente  $p$  y  $p_e$  para obtener estas gráficas. Mientras que las funciones  $p$  y  $p_e$  las podemos escalar multiplicándolas por una constante, la función  $p_\tau$  es fija. Si multiplicásemos toda la función por  $k$  obtendríamos unos valores de  $\tau_x$  divididos por  $\sqrt{k}$ . Entonces aplicando directamente  $p$  o  $p_e$  se obtendrán normalmente unos valores de  $\tau_x$  demasiado grandes o demasiado pequeños, y habrá que tomar  $p_\tau = kp$  ó  $p_\tau = kp_e$ , para un valor concreto de  $k$ . En el caso de la función  $p$  es cierto que hay un factor de escala absoluto, ya que  $p = \psi'/x$ , y podría pensarse que tal vez empleando exactamente esa función los valores de  $\tau$  serían más correctos, pero esto no es así; en este caso fue necesario dividir la función  $p$  por un factor  $k = 2$ .

### Distribución a emplear para las regiones de confianza

Antes de buscar una función  $p_\tau$  que dé buenos resultados podemos resolver ya la segunda parte del problema: decidir qué distribución emplear para calcular los intervalos y regiones de confianza. En las gráficas vemos que las funciones  $\mathcal{G}$  están mucho más próximas a la normal que a  $\mathcal{F}_\ell$ . A partir de  $x = 3$  o  $2,5$ , según la configuración de las observaciones, la diferencia respecto a la normal empieza a ser significativa. Cuanto mayor sean las redundancias menor será la diferencia respecto a una normal. Para los intervalos, elipses y círculos del 95,5% de confianza no habrá casi diferencia, y resulta razonable calcularlos a partir de los elementos  $\tau_{ij}$  como si se tratase de los  $\sigma_{ij}$  de una distribución normal. Para regiones de dimensión 3 y mayores habría que aumentar los valores de los radios así obtenidos.

La aproximación a una normal en un entorno bastante amplio alrededor de  $\tau_x$  permite calcular la matriz  $T_{11}$  como  $AT_{xx}A^T$ .

**Diseño de la función  $p_\tau$**

Podemos deducir algunas características que ha de tener la función  $p_\tau$ . En primer lugar vemos que  $\tau_x$  en el caso  $-0,5 \leq 0 \leq 0,5$  es menor que en  $-1 \leq 0 \leq 1$ , por lo que para valores pequeños tiene que tener una cierta pendiente, al contrario que  $p_e$ . En segundo lugar, cuando un residuo tiende a infinito el resultado de la estimación tiende a ser igual que si ese residuo no existiese, y entonces  $p_\tau$  ha de tender a cero. Se puede emplear un último tramo proporcional a  $1/x^2$ , y entonces si se tienen dos observaciones muy separadas la precisión calculada será inversamente proporcional a la distancia entre las observaciones, lo que parece lógico. Por otra parte, resulta que una observación con residuo alto, pero no tanto como para quede eliminada completamente, por ejemplo 3 ó 4, es peor que si no existiese; es decir, peor que un error mucho mayor para el cual  $p_\tau \approx 0$ , y entonces para esos valores ha de ser  $p_\tau < 0$ . La función adoptada en el ejemplo que estamos siguiendo es la que se muestra unas páginas atrás en la gráfica comparativa de las funciones peso (fig. 19.8).

En este ejemplo sencillo  $(A_\tau^T A_\tau')^{-1}$  equivale a  $(\sum p_\tau)^{-1}$ . Los valores de  $\tau_x$  obtenidos y su diferencia relativa respecto a los valores verdaderos son los siguientes:

$v_1$	$v_2$	$v_3$	$\tau_x$	$\frac{\tau_x - \tau_x}{\tau_x}$
-0,5	0	0,5	0,49	0,041
-1	0	1	0,54	-0,027
-0,5	0,2	0,8	0,52	0,031
-0,7	-0,4	1,6	0,58	0,042
-0,8	-0,5	2,5	0,69	-0,014
-0,6	-0,3	3,6	0,75	-0,015
-0,4	0,1	4,9	0,67	-0,020
-0,3	0	6	0,61	-0,008
-0,3	0	7	0,59	-0,015
-0,2	0,1	10,1	0,58	-0,019
-1,3	-0,1	1,9	0,64	-0,049
-1,4	-0,2	2,8	0,77	-0,030

## 20 Estimadores para otras distribuciones, II

(Continuación)

$v_1$	$v_2$	$v_3$	$\tau_x$	$\frac{\tau_x - \tau_x}{\tau_x}$
-1,1	0,1	4,1	0,80	-0,020
-0,9	0,3	5,3	0,73	0,015
-0,8	0,4	6,4	0,68	0,010
-1,3	-0,1	1,9	0,64	-0,049
-1,4	-0,2	2,8	0,77	-0,030
-1,1	0,1	4,1	0,80	-0,020
-0,9	0,3	5,3	0,73	0,015
-0,8	0,4	6,4	0,68	0,010
-0,7	0,5	10,5	0,65	-0,006
-2	0	2	0,73	-0,077
-2	0	3	0,91	-0,052

Salvo en los dos últimos casos, en los que tampoco es mucho mayor, la diferencia no llega al 5%, lo que para una estimación de la dispersión es muy poco.

El valor  $p_\tau(0)$  tiene un significado particular. Supongamos  $m$  medidas iguales, y por comodidad iguales a 0. Entonces  $g(x) \propto f(x)^m$ . Por ejemplo, si  $f$  es una normal  $(0, \tau)$  entonces

$$g(x) \propto e^{-\frac{mx^2}{2\sigma^2}} = e^{-\frac{(\sqrt{m}x)^2}{2\sigma^2}} \propto f(\sqrt{m}x),$$

de donde  $\tau_x = \tau/\sqrt{m}$ , como ya sabíamos.

Si  $f(x) \propto e^{-x}$  entonces

$$g(x) \propto e^{-mx} \propto f(mx),$$

de donde  $\tau_x = \tau/m$ .

Mediante la función  $p_\tau$  obtendremos el valor

$$\tau_x = \frac{\tau}{\sqrt{mp_\tau(0)}}.$$

Por lo tanto para la distribución normal tiene que ser  $p_\tau(0) = 1$ , lo que ya sabíamos puesto que en este caso  $p_\tau \equiv 1$ . Para la segunda

## 20 Estimadores para otras distribuciones, II

distribución de ejemplo deberíamos tener  $p_\tau(0) = 1/m$ , lo que es imposible ya que es un valor variable con  $m$ . Este hecho pone de manifiesto las limitaciones de obtener  $\tau$  mediante una función  $p_\tau$  del modo en que lo estamos haciendo. De todos modos no es tan grave como en un principio parece. El valor  $\tau_x = \tau/m$  disminuye rápidamente cuando las observaciones no son exactamente iguales, y esto a su vez es más difícil cuanto mayor es  $m$ . Además el valor tan bajo de  $\tau_x$  se debe a que la función  $e^{-x}$  tiene un pico en  $x = 0$ . Si construimos las funciones  $\mathcal{F}$  siguiendo las reglas dadas en la p. 250, en particular la regla 6, la función  $f$  no tendrá ese pico. Esto no quiere decir que el valor de  $p_\tau(0)$  para el cual se obtiene  $\tau_x$  no dependerá de  $m$ , ya que  $\mathcal{F}''(0) = 0$  es simplemente una condición local, y es posible que a excepción de un entorno del cero, y probablemente tampoco para valores grandes, la función  $f$  se parezca  $e^{-x}$ , como de hecho ocurre en el ejemplo que seguimos. Sin embargo la dependencia será menor que si fuese  $f = e^{-x}$ , y cuanto mayor es  $m$  más queda determinada  $g(x)$  por el tramo inicial de  $f$ , y el valor correcto de  $p_\tau(0)$  para cada  $m$  tenderá a un límite que se puede calcular. Entonces será suficiente con un valor de  $p_\tau(0)$  adecuado para  $m = 2$  y  $m = 3$ . A continuación mostramos para nuestro ejemplo, para algunos valores de  $m$ , los valores correctos de  $p_\tau(0, m)$  junto con el error relativo cometido en  $\tau$  por aplicar  $p_\tau(0)$ .

$m$	2	3	4	5	6	$\infty$	$p_\tau(0) = 1,618$
$p_\tau(0, m)$	1,513	1,764	1,950	2,09	2,21	3,67	
$(\tau_x - \tau_x)/\tau_x$	-0,033	0,044	0,098	0,14	0,17	0,43	

Para valores muy grandes de  $m$  el valor de  $\tau_x$  tiene poco significado, ya que es muy sensible a las derivadas de  $\mathcal{F}$  en torno al cero, y con modificaciones muy ligeras de la función se puede hacer que esos valores varíen mucho, como se muestra en el libro de ejercicios. Entonces, si alguna vez quisiésemos llevar a cabo estimaciones con un gran número de redundancias, como puede ser el caso de una calibración, hay que definir el comienzo de  $\mathcal{F}$  de manera que los valores de  $\tau_x$  según la distribución para esos casos sean correctos. Por lo tanto, en esos casos, más que obtener  $\tau_x$  a partir de la parte inicial de  $\mathcal{F}$  se obtiene la parte inicial de  $\mathcal{F}$  a partir de  $\tau_x$ .

Pero dejando aparte esas situaciones, normalmente no hay más de cuatro observaciones por parámetro, y el valor de la función  $p_\tau$

## 20 Estimadores para otras distribuciones, II

proporciona resultados más que aceptables. A la vista del cuadro se aprecia inmediatamente que en este problema sería adecuado un valor de  $p_\tau(0)$  algo mayor. La razón de tomar un valor menor estriba en que un residuo igual a cero, o muy pequeño, aparece también en otras muchas ternas con otros residuos mayores. Puede verse en el cuadro de dos páginas atrás que muchas de las estimaciones de  $\tau_x$  son subestimaciones, y ese error aumentaría si incrementásemos el valor de  $p_\tau(0)$ .

La matriz  $A'_\tau$  se obtiene dividiendo cada fila de  $A$  por  $\sqrt{p_\tau}$ . Si  $p_\tau$  es negativo habría que dividir por un número imaginario. Esta aparición de números imaginarios es transitoria, ya que en el producto  $A'^T A'_\tau$  se multiplican entre sí elementos de una misma fila, apareciendo los valores  $p_\tau$  originales. Si se quieren evitar los números imaginarios, en lugar de dividir los elementos de cada matriz del producto  $A^T A$  por  $\sqrt{p_\tau}$  puede mantenerse una matriz fija y en la otra dividir por  $p_\tau$ , lo que es equivalente.

### Casos límite para $p_\tau$

Los valores de  $p_\tau$  para residuos grandes vemos que proporcionan resultados buenos tanto en el caso en el que otras observaciones permiten obtener con precisión la solución, como las configuraciones  $-0,3 \ 0 \ \ell_3$ , como en los casos contrarios, por ejemplo  $-2 \ 0 \ 3$ . Siempre se pueden buscar casos en los que la función de pesos no dé buen resultado, pero tienen que ser como los siguientes:

$$-3 \ 3; \quad -4 \ 0 \ 4; \quad -3 \ -1,5 \ 1,5 \ 3; \quad -3,8 \ -1,2 \ 1,2 \ 3,8.$$

En alguno de estos incluso  $\sum p_\tau < 0$ . Pero son situaciones en las que hay al menos tantas observaciones con un error grande como redundancias, y no se puede detectar la observación correcta, si es que hay una. Además en todos los casos excepto en el tercero incluso la función  $p_e$  da más de una solución según el punto inicial, y para la que tiene más densidad de probabilidad de entre ellas (según  $p_e$ ), que no es la centrada, el valor  $\sum p_\tau$  es positivo y el valor de  $\tau_x$  tiene poco error, aunque esto último es más bien casualidad. El tercer ejemplo está casi en un punto en el que  $\sum p_\tau = 0$ .

Esta limitación de la función de pesos tiene poca importancia, ya que no tiene mucho sentido preocuparse por el valor de  $\tau_x$  para esas

configuraciones. De todos modos, si se quiere que toda la matriz  $T_{xx}$  cumpla las propiedades de las matrices de varianzas, cuando aparezcan valores negativos en su diagonal se pueden aumentar los valores de peso de las ecuaciones que lo requieran y volver a calcular la matriz. En lugar de buscar esas ecuaciones, que dicho sea es muy sencillo, puede ser más cómodo sustituir  $p_\tau$  por otra  $p_{\tau_2} > 0$ , pero esto estropeará los valores correctos. Una manera de obtener  $p_{\tau_2}$ , o al menos un punto de partida, es tomar el valor que en el caso de dos observaciones con residuos  $-v, v$  proporcione la estimación  $\tau_x = \tau_x$ .

**Procedimiento para la estimación de  $\tau_0$**

Falta estudiar la estimación de  $\tau_0$ . Al contrario que en el estimador mínimo cuadrático, la solución centrada así como la obtenida mediante  $p_e$  depende de  $\tau_0$  en general, por lo que ambas estimaciones no son independientes. La manera de proceder es mediante un proceso iterativo: Se calculan unos valores  $x$  y  $\tau_0$  de manera independiente partiendo de un  $\tau_0$ ; a continuación se vuelve a realizar el ajuste tomando  $\tau_0 = \tau_0$ , y así hasta que converja el proceso.

La estimación de  $\tau_0$  no se puede realizar de manera rigurosa, ya que la función  $g(\tau_0)$  es la siguiente (cf. p. 206):

$$g(\tau_0) \propto \int_E \frac{1}{\tau_0} \prod_i \frac{1}{\tau_0} f_\varepsilon\left(\frac{\ell_i}{\tau_0}\right) dE.$$

Y habría que obtener las integrales, y salvo que el número de parámetros sea muy pequeño esto no es posible. El término  $1/\tau_0$  que aparece fuera del productorio es la función  $h(\tau_0)$ , según se vio en el capítulo 16. La función  $g(\tau_0^2)$  es similar.

Obtendremos un valor a partir de los residuos, lógicamente, así que a falta de poder obtener la distribución  $g(\tau_0)$  veremos cómo se relacionan los residuos con  $\tau_0$  en sentido directo, es decir, qué distribución siguen los residuos. En consecuencia llegaremos finalmente a una estimación de  $\tau_0$  basada en un criterio directo. En el estimador mínimo cuadrático vimos que un criterio de ese tipo daría un valor  $\tau$  desviado por un factor  $\sqrt{\frac{r-2}{r}}$  de la estimación centrada, así que una vez que obtengamos  $\tau$  según el criterio directo deberemos multiplicar su valor por una constante adecuada, no necesariamente  $\sqrt{\frac{r}{r-2}}$ , para compensar la diferencia. Lo recordaremos en su momento.

## 20 Estimadores para otras distribuciones, II

Vimos que en torno a  $2\tau$  las distribuciones  $g(l)$  y  $g(x)$  están más próximas a una normal que a la función  $f$ . Sin embargo para los residuos no será así. Cuanto mayor sea el número de redundancias  $y$ , de ahí, más precisas las estimaciones de los parámetros, más se aproximarán las cantidades  $l - \ell$  a los errores reales, y por lo tanto más se aproximará la distribución de  $v$  a la  $f$  de las observaciones.

La obtención de la distribución de las cantidades  $v$  es un problema directo, y por lo tanto no muy complicado y que se puede resolver para cada problema concreto generando conjuntos de observaciones siguiendo la distribución  $f$  y registrando los valores  $v$  (normalizados) que se obtendrían. Dada la simetría de la función  $f$  no tendremos en cuenta el signo de los  $v$ , y nos fijaremos sólo en  $|v|$ . Para generar directamente observaciones  $\ell_i$  emplearemos la función  $f$  de éstas, que es la misma para todos, para un valor arbitrario de  $\tau_0$ .

Sabemos que cuando la redundancia es muy pequeña el valor del residuo tiende a ser mucho menor que el error, así que, siendo las  $r_i$  distintas, cada  $|v_i|$  seguirá una distribución distinta. Para no tener que estudiar una distribución para cada residuo obtendremos mejor la distribución conjunta de los  $|v_i|/\sqrt{r_i}$ . Para ello hay que resolver cómo calcular las redundancias  $r_i$ . En una estimación mínimo cuadrática no dependen de  $\varepsilon_i$ , pero en general no será así. Si un error es muy grande su residuo será también grande, por lo que tendrá un peso  $p_e$  en el ajuste pequeño. Entonces influye poco en el ajuste y su redundancia también es alta. Por otra parte, si un conjunto de  $m$  observaciones con  $r$  redundancias contiene una con un error muy alto, entonces la solución será como si esa observación no existiese, y las redundancias del resto de observaciones disminuyen, pasando a sumar  $r - 1$ . No conocemos los errores, pero conocemos los residuos, y podemos calcular los elementos  $r_i$  como la diagonal principal de la matriz

$$I - A'N^{-1}A'^T,$$

en donde la matriz  $A'$  es la obtenida mediante los pesos  $p_e$ , y  $N$  calculada a partir de esta misma  $A'$ . Así, si una observación  $\ell_i$  tiene un peso  $p_e$  muy bajo, su fila en la matriz  $A'$  tendrá valores pequeños, y entonces en la matriz  $A'N^{-1}A'^T$  el elemento  $(i, i)$  será también pequeño, y la redundancia, que es su complemento a 1, próxima a 1. El resultado de cada simulación será un conjunto de valores  $|v_i|/\sqrt{r_i}$ . Puede buscarse otra manera de calcular las redundancias parciales,

pero hay que tener en cuenta que han de sumar  $r$ .

La función obtenida será  $\mathcal{F}$  (ó  $\overline{\mathcal{F}}$ ), y la denotaremos por  $\mathcal{F}_{vr}$ . Los valores  $|v_i|/\sqrt{r_i}$  no dan directamente la función  $\mathcal{F}_{vr}$ . Vimos al principio del capítulo anterior, y no vamos a repetir aquí el argumento, que a la hora de emplear los residuos para obtener una distribución cada uno de ellos cuenta como  $r_i$ . Sobre esta manera de contar los residuos volveremos más adelante, pero es importante ver que se puede calcular  $\tau_0$  a partir de una distribución  $\mathcal{F}_{vr}$  sea cual sea la manera de contar los residuos que se empleó para obtener  $\mathcal{F}_{vr}$ . Esta función no es una estimación, sino que es un resultado exacto (en la medida en que lo sea la función  $f$  de partida): suponemos una función  $\mathcal{F}_\ell$  y a partir de ella y para el problema concreto de que se trate obtenemos  $\mathcal{F}_{vr}$ . La función  $\mathcal{F}_{vr}$  se obtendrá escalada según el  $\tau_0$  que hayamos considerado en la simulación, de manera que se mantiene constante la relación entre  $\mathcal{F}_{vr}$  y  $\mathcal{F}$ , y según sea  $\tau_0$  así será de dispersa  $\mathcal{F}_{vr}$ . El proceso de estimación de  $\tau_0$  consiste, conceptualmente, en obtener el conjunto de valores  $|v_i|/\sqrt{r_i}$  y «contarlos» según el mismo criterio que se empleó para obtener  $\mathcal{F}_{vr}$ , obteniendo una distribución que se parecerá a  $\mathcal{F}_{vr}$ . Tomaremos para el conjunto de valores obtenidos un indicador de la dispersión (por ejemplo  $\tau$ ) y veremos cómo se relaciona ese valor en  $\mathcal{F}_{vr}$  con  $\tau_0$  de la distribución  $\mathcal{F}$ .

### Medida de la dispersión para comparar $\mathcal{F}_{vr}$ y $\mathcal{U}$

Dejaremos para más adelante la elección de un criterio para el conteo de los residuos y estudiaremos la obtención de  $\tau_0$  supuesto escogido un criterio y conocida  $\mathcal{F}_{vr}$ . Lo que cuenta cada residuo en la función  $\mathcal{F}_{vr}$  lo denotaremos por  $s_i$ . El proceso para obtener  $\tau_0$  ya lo hemos descrito y sólo falta por concretar qué medida de la dispersión seleccionaremos. La distribución  $\mathcal{F}$  de los valores  $|v_i|/\sqrt{r_i}$  obtenidos, en la que cada uno de ellos cuenta  $s_i$ , la denotaremos por  $\mathcal{U}$ , y sea  $s = \sum s_i$ .  $\mathcal{U}$  es una función escalonada que en cada punto  $|v_i|/\sqrt{r_i}$  aumenta su valor en  $s_i/s$  (fig. 20.3).

Una medida evidente de la dispersión de  $\mathcal{U}$  es  $\tau_u$ . El valor  $\tau$  de la función  $\mathcal{F}_{vr}$  estará en una relación constante  $c$  con  $\tau_0$ , y una estimación de  $\tau_0$  es entonces  $\tau_u/c$ :

$$\tau_0 = \frac{\mathcal{U}^{-1}(0,995)}{2c} k_c. \quad (20.1)$$

## 20 Estimadores para otras distribuciones, II

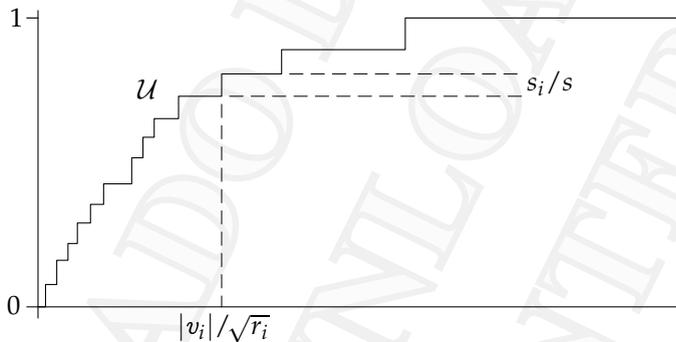


Fig. 20.3: Función  $\mathcal{U}$

El factor  $k_c$  es una constante cuya inclusión es necesaria para que la estimación sea centrada. La determinaremos más adelante.

Este estimador tiene el problema de que es necesario un conjunto grande de observaciones. Si por ejemplo  $s \approx r = 40$ , el número de observaciones mayores que  $2\tau_u$  es  $0,045 \cdot 40 = 1,8$ , de manera que  $\tau_u$  quedará determinado por el máximo número de residuos mayores para los cuales  $\sum s_i < 1,8$ , que normalmente serán 2. Estamos basando entonces el valor de  $\tau_0$  en los dos residuos mayores. Evidentemente una estimación así obtenida es muy variable, y es conveniente buscar una más estable. Hay que distinguir entre la definición de  $\tau$  como  $\frac{1}{2}\mathcal{F}^{-1}(0,995)$  y muchos otros valores que para una distribución F concreta también proporcionan el valor de  $\tau$ , como por ejemplo  $k\mathcal{F}^{-1}(q)$ , para cualquier  $0 < q < 1$  y una constante  $k$  que depende de  $q$ . En la función  $\mathcal{F}$  del ejemplo  $\mathcal{F}(\tau) \approx 0,758$ , así que  $\tau \approx \mathcal{F}^{-1}(0,758)$ , o también  $\mathcal{F}(1,4\tau) \approx 0,876$ , y entonces  $\tau \approx 0,7143\mathcal{F}^{-1}(0,876)$ . Otros modos de obtener  $\tau$  pueden ser  $kE[|x|]$ ,  $k\sigma$ , etc., cada una para un valor de  $k$ .

Todas estas expresiones que proporcionan el valor de  $\tau$  se pueden determinar para la función  $\mathcal{F}_{vr}$ . Omitiendo la constante  $k$  se obtiene un número que está en una relación constante con  $\tau_{vr}$ , que a su vez está en relación constante con  $\tau_0$ . Entonces se puede aplicar cualquiera de ellas a la función  $\mathcal{U}$  y dividir por la constante  $c$  que la relaciona con  $\tau_0$ . Una bastante buena es

$$\tau_0 = \frac{\mathcal{U}^{-1}(q_1)}{c} k_c, \quad q_1 = \mathcal{F}_{vr}(\tau_{vr}); \quad (20.2)$$

## 20 Estimadores para otras distribuciones, II

o bien

$$\tau_0 = \mathcal{U}^{-1}(q_0) k_c, \quad (20.3)$$

en donde  $q_0$  es el cuantil para el cual  $c = 1$ .

En las funciones  $\mathcal{F}$  para las que se suelen buscar estimadores distintos del mínimo cuadrático, que suelen presentar una caída de la densidad para errores grandes mucho menos drástica que la normal, no se debe calcular  $\tau_0$  a través de  $\sigma_{ur}$ , ya que depende mucho de los residuos más altos. Esta dependencia fue la causa de que sustituyésemos  $\sigma$  por  $\tau$ . Incluso es normal que  $\sigma_{vr} = \infty$ .

Se emplea a veces una media de los  $v_i^2/r_i$  ponderada según alguna función de pesos, que suele ser  $p_e$ . La ventaja de los cuantiles, especialmente cuando el número de observaciones es pequeño, es que no dependen de valores excepcionalmente grandes de algún residuo. Si el cuantil  $q$  no es muy alto la dependencia es aún menor, y también es independiente de un menor número de errores elevados de los que cabría esperar.

La variabilidad de  $\tau_0$ , que tendrá relación con la precisión de ese valor como estimación de  $\tau_0$ , la podemos obtener a partir de las simulaciones. Normalmente el valor obtenido mediante los cuantiles  $q_1$  y  $q_0$  será más estable que el obtenido mediante 0,955, siendo la diferencia menor cuanto mayor sea el número de redundancias. Sin embargo este último tiene una ventaja: Al emplear como estimación la propia definición, el resultado depende sólo de la constante  $\tau_{vr}/\tau_0$ . La función  $\mathcal{F}$  de partida rara vez será tan precisa como para que podamos considerarla perfecta. Si la distribución  $\mathcal{F}$  real de las observaciones es algo distinta seguramente la diferencia entre el valor de  $c = \tau_{vr}/\tau_0$  empleado y el real, desconocido, sea muy pequeña. Sin embargo los valores  $\mathcal{F}(\tau)$  y  $\mathcal{F}_{vr}(\tau_{vr})$  pueden tener variaciones mayores. Por ello, si la diferencia entre la variabilidad de la estimación mediante  $\mathcal{U}^{-1}(0,955)$  y las obtenidas mediante  $q_1$  o  $q_0$  no es muy grande, es preferible emplear la primera. Esto ocurrirá normalmente para problemas con un número alto de redundancias. Cuando no dispongamos de tantas redundancias será mejor la solución a través de  $q_1$ ,  $q_0$  u otro cuantil que consideremos apropiado. Finalmente, para valores muy bajos de  $r$  tal vez no quede más remedio que emplear

## 20 Estimadores para otras distribuciones, II

una media ponderada de los  $v^2$ , es decir,

$$\tau_0 = \frac{1}{c} \sqrt{\frac{\sum p_e(v_i/\sqrt{r_i}) s_i \frac{v_i^2}{r_i}}{s}} k_c.$$

### Cálculo de la constante de centrado $k_c$

Tenemos pendiente la deducción de los  $s_i$  y los valores  $k_c$ . Empezaremos por estos últimos. Se obtienen fácilmente a partir de las simulaciones. En efecto, supongamos que no aplicamos  $k_c$ . Entonces la simulación proporciona la distribución del estimador  $c\tau_0$  (es decir,  $\tau_0$  antes de dividir por  $c$ ). El estimador lo aplicamos a los valores  $|v_i|/r_i$ , para obtener (estimar) esa misma magnitud en la función  $\mathcal{F}_{vr}$ . A esta magnitud la llamaremos  $\rho$ , y entonces  $\rho = c\tau_0$ . Es posible que el estimador así aplicado no sea centrado, y el descentrado lo obtenemos a partir de la distribución del estimador obtenida mediante simulación. Supongamos para concretar las ideas que  $\mathcal{F}_{vr}(1,4\tau_{vr}) = 0,85$ , y entonces la estimación es  $\rho = \mathcal{U}^{-1}(0,85)$  con el objetivo de obtener  $\rho = 1,4\tau_{vr}$ . Realizamos las simulaciones para obtener la distribución de  $\rho$  y resulta que la esperanza de la distribución obtenida es, por ejemplo,  $E[1,48\tau_{vr}] \approx 1,057\rho$ . Entonces deberemos aplicar una corrección  $k_c \approx 1/1,057 \approx 0,946$ , y  $\rho$  deja de ser  $\mathcal{U}^{-1}(0,85)$  para ser  $\mathcal{U}^{-1}(0,85) \cdot 0,946$ . A la cantidad  $\rho/k_c$  que obtenemos de la distribución  $\mathcal{U}$  la llamaremos  $\rho'$ .

De esta manera obtenemos un estimador centrado para  $\tau_0$ , pero habrá que plantearse si se quiere un estimador centrado para  $\tau_0$  o para  $\tau_0^2$ . Dada la naturaleza de la definición de  $\tau$ , que no entraña cuadrados, puede parecer que lo correcto es una estimación centrada para  $\tau_0$ . Sin embargo la ausencia de cuadrados en su definición es algo arbitrario. El cuantil 0,955 se encuentra en la misma posición respecto a las observaciones (la variable) que ese mismo cuantil en la distribución de los cuadrados, o de cualquiera otra función creciente de la variable. Así, la definición de  $\tau$  dada es equivalente a  $\frac{1}{2}\sqrt{\mathcal{F}_{\chi^2}^{-1}(0,955)}$ . La decisión habrá de tomarse en función de la conservación de estimaciones centradas al sumar distribuciones, al igual que se hizo para la desviación típica. La pregunta es entonces: Si  $\tau_1$  y  $\tau_2$  son estimaciones centradas de las magnitudes  $\tau_1$  y  $\tau_2$  de dos distribuciones, ¿es enton-

## 20 Estimadores para otras distribuciones, II

ces  $\tau_1 + \tau_2$  una estimación centrada del parámetro  $\tau$  de la suma, o lo que se cumple es la propiedad análoga para estimaciones centradas de  $\tau_1^2$  y  $\tau_2^2$ ? Se cumpliría la primera propiedad si fuese  $\tau = \tau_1 + \tau_2$ , y la segunda si  $\tau^2 = \tau_1^2 + \tau_2^2$ . En la distribución normal  $\tau$  coincide con  $\sigma$ , por lo que se cumple lo segundo. Si las dos distribuciones que se suman son iguales, la primera propiedad equivale a  $\tau = 2\tau_1$  y la segunda a  $\tau \approx 1,414\tau_1$ . Los cocientes  $\tau/\tau_1$  para algunas distribuciones habituales son los siguientes:

Distribución	$\tau/\tau_1$	Distribución	$\tau/\tau_1$
Normal	$\sqrt{2} \approx 1,414$	Uniforme	1,65
$\frac{1}{2}\varepsilon^{- x }$	1,36	$x^2 \sim \chi_1^2$ *	$\frac{2,5}{2} = 1,25$

Parece entonces mejor emplear estimaciones centradas de  $\tau_0^2$ . La única variación que esto supone en la obtención de  $k_c$  es que se estudia la distribución de  $\rho'^2$  en lugar de la de  $\rho'$ .

El problema no se acaba ahí, ya que de esta manera obtenemos un estimador centrado desde el punto de vista directo, cuando lo deseable es un estimador centrado en sentido inverso. Hay dos posibles soluciones. La primera, la perezosa, consiste simplemente en multiplicar el valor obtenido de  $k_c$  por  $\sqrt{\frac{r}{r-2}}$ , por analogía con la estimación mínimo cuadrática. La segunda y más rigurosa supone la obtención de  $g(\rho^2)$  a partir de la  $f(\rho'^2)$  obtenida en la simulación y la función  $h(\rho^2) = 1/\rho^2$ , que en su momento justificamos. Este es un problema ya visto que tiene por solución

$$g(\rho^2) = g(t) \propto \frac{1}{t} f_t\left(\frac{\rho'^2}{t}\right) = \frac{1}{t^2} f_0\left(\frac{\rho'^2}{t}\right),$$

en donde  $f_0$  es la distribución  $f_t(\rho'^2)$  para un valor concreto de  $t$ , como por ejemplo  $t = 1$  o el correspondiente a  $\tau_0 = 1$ . De esta función  $g$  y no de  $f(\rho'^2)$  es de donde obtendremos  $k_c$ .

---

\*Tomando intervalos  $[0, x]$  en lugar de  $[\mu_x - x, \mu_x + x]$ .

## 20 Estimadores para otras distribuciones, II

### Sobre el carácter centrado de la estimación de $\tau_0$

Parece que al fin y al cabo hemos conseguido una estimación centrada de  $\tau_0$  (de su cuadrado), pero no es del todo cierto. Una estimación centrada tiene en cuenta todas las observaciones y en función de ellas se obtiene  $g(\tau_0^2)$  y a partir de ésta  $E[\tau_0^2]$ . La función  $g(\rho^2)$  la hemos calculado teniendo en cuenta nada más la función  $f(\rho'^2)$  y el valor  $\rho'$  obtenido. La estimación  $\tau_0$  es centrada para el problema: «Obtener una estimación de  $\tau_0$  sabiendo que se ha obtenido el valor  $\rho'$ », pero no para el problema: «Obtener una estimación de  $\tau_0$  sabiendo que las observaciones han sido  $l_1, l_2, \dots, l_m$ ».

Mostraremos la diferencia entre ambas estimaciones mediante un ejemplo sencillo. Sea la distribución de los errores una uniforme entre  $l - 0,5$  y  $l + 0,5$ . Realizamos tres medidas y tomamos la media aritmética,  $A$ . Vamos a ver cuál es la estimación centrada de  $l$  si lo único que conocemos es  $A$ . La distribución de  $A$  es, salvo por la división entre 3, la de la suma de tres variables uniformes iguales, cuya gráfica se muestra en la página 41. La distribución  $f(l)$  se desplaza siguiendo el valor verdadero, y entonces hace lo propio la distribución  $f(A)$ , de modo que  $g(A + \Delta l) = f_A(A - \Delta l)$ , y la distribución  $g(l)$  es la distribución  $f(A)$  volteada respecto al valor  $A$ . Dado que  $f(A)$  es simétrica respecto a  $A$  la función  $g(l)$  coincide con  $f(A)$ , y  $l = A$  es la estimación centrada de  $l$  sabiendo nada más que la media aritmética es  $A$ . Sin embargo si tenemos en cuenta las tres observaciones la función  $g(l)$  es uniforme con intervalo  $[l_{\text{máx}} - 1, l_{\text{mín}} + 1]$ , y la estimación centrada el centro de ese intervalo. Sea  $l_2$  la observación que no es ni  $l_{\text{máx}}$  ni  $l_{\text{mín}}$ . Si  $l_2 > \frac{1}{2}(l_1 + l_3)$  entonces también  $A > \frac{1}{2}(l_1 + l_3)$ , y el valor  $l = A$  es una sobreestimación, y en caso contrario una subestimación. Por cada conjunto  $(l_{11}, l_{21}, l_{31})$  de valores observados en los que  $A$  es una subestimación existe otro, con la misma probabilidad,

$$(l_{12}, l_{22}, l_{32}) = (A - (l_{11} - A), A - (l_{21} - A), A - (l_{31} - A)),$$

para el cual  $A$  es una sobreestimación con un sesgo del mismo valor absoluto, de modo que, en el conocimiento únicamente de  $A$ , ésta es una estimación centrada.

Los intervalos de confianza para  $\tau_0$  los obtenemos en  $g(\tau_0^2)$ . Si no hemos llegado a obtener esta función podemos tomar la solución

## 20 Estimadores para otras distribuciones, II

aproximada  $P(\tau_0 > h\tau_0) = P(\tau_0 < \tau_0/h)$ , por analogía con la estimación mínimo cuadrática.

### Estimación conjunta de $\tau_0$ y los parámetros

La solución del ajuste depende del valor  $\tau_0$  inicial, así que es necesario repetir el ajuste tomando el valor  $\tau_0$  calculado, lo que dará lugar a unos nuevos valores de las cantidades  $x$  y de  $\tau_0$ , y el proceso se repite hasta que converja. Puesto que el cálculo de los  $x_i$  es en sí mismo un proceso iterativo puede obtenerse  $\tau_0$  en cada iteración y emplear el valor calculado para la siguiente, de manera que se obtienen conjuntamente  $(x_1, \dots, x_n)$  y  $\tau_0$  en el mismo ajuste.

Si se quiere estimar distintos valores de  $\tau_0$  para otros tantos conjuntos de observaciones, se puede emplear la misma estimación que para un  $\tau_0$  global pero restringida al conjunto de observaciones en cuestión. Si las observaciones son de naturaleza distinta entonces es mejor calcular una función  $\mathcal{F}_{v_r}$  para cada conjunto, y una magnitud  $\rho$  adecuada para cada uno, y aunque se elijan las mismas no serán iguales los valores  $c$  y  $k_c$ . Aún más, se debe comenzar desde el principio con funciones  $\mathcal{F}$  distintas, y en consecuencia distintas  $p_e$  y  $p_\tau$ .

### Cuánto se cuenta cada residuo

Vamos a abordar ahora el estudio de la asignación de cantidades  $s_i$ . Puesto que en principio cualquiera de ellas proporciona resultados correctos primero habrá que definir qué es lo que queremos mediante la elección de unos  $s_i$  concretos. La estimación  $\rho$  es centrada basándose únicamente en el valor  $\rho'$ , pero según acabamos de ver, en cada problema concreto puede ser más o menos sesgada. La elección de los  $s_i$  debe hacerse de modo que el sesgo de  $\rho$  en problemas concretos sea lo más pequeño posible.

Una decisión en este sentido ya se produjo al tomar en consideración los valores  $v_i/\sqrt{r_i}$  en lugar de  $v_i$ . Si hubiésemos empleado estos últimos y trabajado con la función  $\mathcal{F}_v$  podríamos igualmente haber obtenido mediante simulación la distribución  $f$  del estimador  $\rho'$ , y obtenido una estimación centrada en sentido inverso basada en ella. Sin embargo esta estimación tendría fácilmente un sesgo importante, debido a que los residuos con menor redundancia tienden a ser

## 20 Estimadores para otras distribuciones, II

menores y aquellos con una mayor, mayores, y no se distribuyen por igual en la distribución  $\mathcal{F}_v$ , teniendo ésta en su parte más baja mayor proporción de residuos con redundancia pequeña que en el total de los residuos, y sobre todo mayor presencia en los valores más altos de residuos con redundancia elevada, de manera que variaciones en pequeños conjuntos de residuos pueden causar una diferencia en la distribución  $\mathcal{U}$  respecto a la teórica  $\mathcal{F}_v$  importante en una zona concreta y que varíe significativamente es resultado de la estimación.

La elección de unos  $s_i$  u otros modifica la distribución  $\mathcal{F}_{vr}$  y los valores  $c$ . Estas constantes son iguales a  $\rho/\tau_0 = \rho_{vr}/\tau_\varepsilon$ , y son la composición (producto) de  $\rho_{vr}/\rho_\varepsilon$  y  $\rho_\varepsilon/\tau_\varepsilon$ . Estas últimas dependen sólo de la función  $\mathcal{F}_\varepsilon$ , y una vez seleccionada ésta son unas cantidades fijas. Las primeras son las dependientes de  $\mathcal{F}_{vr}$  y por tanto de los  $s_i$ , y las llamaremos  $c'$ , y  $c'_0$  a los valores que se obtienen cuando  $s_i \equiv 1$  para todos los residuos. La distribución  $\mathcal{F}_{vr}$  para esos  $s_i$  será  $\mathcal{F}_{vr0}$ .

Los valores  $c'_0$  serán próximos a 1, ya que la distribución de las cantidades  $|v_i|/\sqrt{r_i}$  es muy parecida a la de los errores. Valores de  $c'_0$  mayores que 1 indican que las cantidades  $|v_i|/\sqrt{r_i}$  son mayores que los  $|\varepsilon'|$ , en el sentido del estimador  $\rho$ , y lo contrario para  $c'_0 < 1$ . Supongamos una figura formada por un conjunto de 10 observaciones con 1 redundancia y otras 30 observaciones con 27 redundancias, independientes los conjuntos entre sí, y que los pesos  $s_i$  son iguales a 1. Está claro que el resultado depende en exceso del conjunto de 10 observaciones. Si una sola de ellas tiene un residuo algo grande, o simplemente si los errores se han acumulado en el mismo sentido, el valor de  $\rho$  será muy elevado, debido a que cuentan como 10 residuos lo que en realidad es una única estimación. Es un caso en el que, basándonos únicamente en el valor de  $\rho'$ , la estimación  $\tau_0$  será centrada, pero atendiendo al valor de todas las observaciones la estimación centrada sería menor. Al contrario ocurrirá si esos 10 residuos, aun divididos por  $\sqrt{r_i}$ , son muy pequeños. En ese caso los 10 residuos aportan una probabilidad 1/4 a la parte baja de la distribución  $\mathcal{U}$ , por lo que los cuantiles disminuyen, así como estimaciones que tengan en cuenta todos los residuos. Por ello una primera solución consiste en  $s_i = r_i$ . Así se puede interpretar el estimador mínimo cuadrático. En ese caso las cantidades  $v_i/\sqrt{r_i}$  siguen distribuciones  $N(0, \tau_0)$ , y una media de sus cuadrados con

## 20 Estimadores para otras distribuciones, II

$s_i = r_i$  es

$$\sigma_0^2 = \frac{\sum s_i \frac{v^2}{r_i}}{s} = \frac{\sum v^2}{r},$$

que es una estimación centrada en sentido directo.

En la estimación mínimo cuadrática las redundancias parciales de cada observación son fijas, pero en cualquiera otra dependen del valor de los residuos a través de  $p_e$ . Sigamos con el ejemplo del conjunto de 40 observaciones, pero suponiendo ahora que el primer subconjunto de 10 tiene 2 redundancias en lugar de 1, para poder detectar un error muy grande en una observación. Si entre esas 10 hay una observación con un residuo muy alto, entonces su redundancia parcial será prácticamente 1, y las de las otras 9 observaciones disminuyen de manera que suman 1. Análogamente ocurre si entre las otras 40 hay una con un residuo muy alto. Por lo tanto los errores muy grandes se detectan y quedan con redundancia 1, y si  $s_i = r_i$  entonces también con  $s_i = 1$ . Supongamos que los errores se dividen en pequeños o medianos y muy grandes, y que la probabilidad de estos últimos es  $p$ . La probabilidad de que entre las 40 observaciones no haya ninguna con un error muy grande es  $(1 - p)^{40} \approx 1 - 40p$ . Suponemos que  $p$  es muy pequeño y despreciamos los términos  $p^2$ , etc. La probabilidad de que haya un error grande es  $\approx 40p$ , y la de que haya más de uno es prácticamente cero. Estos errores se detectan y tienen  $s_i = 1$ , de manera que por término medio aportarán a la distribución  $\mathcal{U}$  una cantidad  $(40p)/s = \frac{40}{29}p$ , y el promedio de distribuciones  $\mathcal{U}$  es la distribución  $\mathcal{F}_{vr}$ . Por lo tanto la probabilidad de un valor muy grande en  $\mathcal{F}_{vr}$  es mayor que en  $\mathcal{F}_{vr0}$ , y se tendrá valores  $c' > c'_0$ . Si la probabilidad de que el número de errores grandes entre las  $m$  observaciones sea mayor que 1 no es despreciable, bien porque  $p$  no es tan pequeño o bien porque  $m$  es mayor, habrá en todo caso una cantidad  $m_1$  que podemos esperar por término medio de errores muy grandes; es decir, una proporción  $m_1/m$ . Esos errores contarán en la distribución  $\mathcal{F}_{vr}$  como  $m_1/r$ , que es una proporción mayor, y seguiremos teniendo  $c' > c'_0$ .

Los errores, claro está, no se dividen drásticamente en medianos y muy grandes, y la mayor proporción de residuos (divididos por  $\sqrt{r_i}$ ) de una cierta magnitud en  $\mathcal{F}_{vr}$  en relación a  $\mathcal{F}_{vr0}$  y a  $\mathcal{F}$  es mayor cuanto mayor sea la magnitud del residuo, por lo que en las estimaciones basadas en  $s_i = r_i$  los errores mayores tienen más importancia

## 20 Estimadores para otras distribuciones, II

de la que tienen en la propia distribución  $\mathcal{F}$ . No sucede lo mismo para  $\mathcal{F}_{vr0}$ , aunque  $c'_0 > 1$ , ya que en este caso el valor mayor que 1 de  $c'$  se debe a que los valores  $|v_i|/\sqrt{r_i}$  aumentan todos en general respecto a  $|\varepsilon_i|$ , y no a que los valores mayores tengan más peso.

Por ello, aunque la solución  $s_i = r_i$  es bastante buena, podemos intentar buscar otra en la que la proporción de los residuos mayores se ajuste más a la que existe en  $\mathcal{F}$  y en  $\mathcal{F}_{vr0}$ . Una solución al respecto consiste en tomar para  $s_i$  los valores de las redundancias iniciales, los que se obtienen para  $p_e \equiv 1$ . Estos valores los denotaremos por  $r_{i0}$ . Así la redundancia de cada residuo es independiente de su magnitud, y valores  $|v_i|/\sqrt{r_i}$  de cualquier magnitud tendrán por término medio una redundancia  $r/m$ . Pero esta solución es peor que mantener  $s_i = r_i$ , ya que un residuo muy grande que ha quedado con redundancia 1 tendrá un  $s_i$  distinto según  $r_{i0}$ , pudiendo incluso ser muy próximo a cero. Una observación de este tipo prácticamente no influye en la solución, y el valor  $s_i$  debería ser independiente de  $r_{i0}$ .

Podemos entonces buscar una serie de propiedades para los valores  $s_i$ . I) Cuando  $r_i \rightarrow 1$  la cantidad  $s_i$  tiende a un valor independiente del  $r_{i0}$  inicial, y II) Si todos los residuos son pequeños entonces  $s_i \approx r_{i0}$ . Estas condiciones ya se cumplen si  $s_i = r_i$ . La condición que falta es la que refleje que la proporción de valores altos en  $\mathcal{F}_{vr}$  según los  $s_i$  sea parecida a la de  $\mathcal{F}_{vr0}$ , es decir, a la de  $s_i = 1$ . Esta es: III) Si  $r_i \rightarrow 1$  entonces  $s_i \rightarrow s/m$ , en donde  $s = \sum s_i$ .

Para cumplir estas condiciones puede emplearse una función  $s_i(r_i)$  de dos tramos, lineal en cada uno de ellos, definida mediante su valor en tres puntos:

$$s_i(0) = 0, \quad s_i(r_{i0}) = r_{i0}, \quad s_i(1) = \frac{\sum s_i}{m}.$$

Si  $r_i > r_{i0}$  entonces  $s_i$  está definido en función de  $\sum s_i$ , valor en el cual interviene el propio  $s_i$ . Se puede plantear un sistema de ecuaciones, pero la manera más fácil de resolverlo es de modo iterativo, ya que una vez calculados los valores  $s_i$  a partir de un  $(\sum s_i)_0$  el valor de  $\sum s_i$  no variará mucho. Se empieza con  $(\sum s_i)_0 = r$ .

Esta solución tiene la paradoja de que para ciertas observaciones cuanto mayor sea el residuo menor es  $s_i$ . La diferencia en la solución respecto a la obtenida con  $s_i = r_i$  será pequeña.

**Obtención de la distribución de los errores**

Llegados a este punto podemos obtener como último resultado del ajuste lo que supone el punto de partida de la estimación: la distribución de los errores. Supongamos para evitar constantes innecesarias que  $\mathcal{F}$  se ha tomado de manera que en ella  $\tau = 1$ . Los valores que se tomarán para ello son

$$|\tilde{\epsilon}_i| = \mathcal{F}^{-1} \left( \mathcal{F}_{vr} \left( \frac{|v_i|/\tau_0}{\sqrt{r_i}} \right) \right), \quad \text{cantidad}(\tilde{\epsilon}_i) = s_i, \quad (20.4)$$

que ya no necesita más explicación. Entonces el proceso de desarrollo de un estimador para una distribución consta de los siguientes pasos.

1. A partir de un conjunto de valores  $\tilde{\epsilon}$  ajustar a ellos una distribución  $\overline{\mathcal{F}}$ .
2. Derivar de  $\overline{\mathcal{F}}$  la función  $p$  y demás información interesante, como  $\mathcal{F}(\tau)$ .
3. Seleccionar un conjunto de casos representativos sencillos para obtener las funciones de peso siguientes.
4. Tomando como base la función  $p$ , encontrar una función  $p_e$  para que el estimador sea lo más centrado posible.
5. Construir una función  $p_\tau$ , y si es necesario decidir qué hacer si algún valor de precisión resulta negativo.
6. Decidir un criterio para los  $s_j$ .
7. Seleccionar uno o varios casos reales para realizar simulaciones, y en función de ellas elegir un estimador  $\rho$  e incluir las constantes necesarias para que sea una estimación centrada de  $\tau_0$ .
8. Obtener un conjunto grande de valores  $\tilde{\epsilon}$ , y volver a empezar.

**Agrupación de observaciones ligadas**

Cuando se aplica el estimador hay muchas veces pares o conjuntos de observaciones relacionados de manera que si una de ellas tiene un error muy grande entonces las otras también lo tendrán; por ejemplo cuando se miden coordenadas de un punto sobre una imagen.

## 20 Estimadores para otras distribuciones, II

Si una de las coordenadas tiene un error muy grande indica normalmente una equivocación en el punto a medir, y la otra sólo podrá ser correcta de casualidad. En esos casos debe aplicarse el mismo valor  $p_e$  a ambas observaciones, basándose en la que tiene residuo mayor. Lo que en realidad ocurre es que desde el punto de vista operativo se trata de una única observación, aunque de ella se obtengan dos valores. En este caso el valor en función del cual se obtiene  $p_e$  ha de ser la distancia al punto ajustado:  $v_{x,y} = \sqrt{v_x^2 + v_y^2}$ . En otras situaciones podrá ser  $v_{1,2} = \max\{v_1, v_2\}$  o lo que lógicamente se determine. Cuando estos pares de observaciones se combinan con otras individuales del mismo tipo, estas parejas están recibiendo por término medio menos peso del que deberían, ya que un peso obtenido a partir de una cantidad  $v_{1,2} \geq \max\{v_1, v_2\}$  es menor que el obtenido a partir del propio residuo, incluso aunque los valores  $v_1$  y  $v_2$  sean pequeños. Entonces habrá que compensarlo de alguna manera, bien en el valor  $p_e$  obtenido o bien en las cantidades  $v_{1,2}$ , tomando en lugar de ellas  $\mathcal{F}(\mathcal{F}_{\varepsilon_{1,2}}^{-1}(v_{1,2}))$ , en donde  $\mathcal{F}_{\varepsilon_{1,2}}$  se obtiene a partir de  $\mathcal{F}$ . La mayoría de las veces las observaciones que van emparejadas no se presentarán además de manera aislada, y lo que se hace es tomar directamente como  $\mathcal{F}$  la distribución de los  $\varepsilon_{1,2}$ , y desarrollar las funciones de peso y todo lo demás a partir de estas magnitudes y esta función.

### Aplicación previa de otra función de pesos para obtener valores iniciales

Vimos que el proceso iterativo en la resolución de un ajuste mediante una función  $p_e$  o mediante  $\psi''$  requiere un punto inicial muy próximo a la solución, que se suele obtener mediante una estimación previa mínimo cuadrática. Según cómo sea la disposición relativa de las observaciones, puede haber ciertas observaciones que si tienen un error muy grande el punto inicial así obtenido no sea suficiente para alcanzar una solución correcta. Sucede muchas veces cuando la redundancia de la observación es muy baja, lo que a su vez puede ser debido a que al tener mucho error ha quedado «aislada». Para solucionar este problema se efectúan a partir de la solución mínimo cuadrática unas iteraciones con una función  $p_e$  distinta, más parecida a las de estimadores de máxima probabilidad. Puede ser una función con una caída muy brusca en torno a 2,5–3, o bien se toman los valores  $|v_i|/\sqrt{r_i}$

## 20 *Estimadores para otras distribuciones, II*

en lugar de  $|v_i|$ , o ambas cosas. Así se consigue eliminar observaciones que de otro modo permanecerían con residuo pequeño en todas las iteraciones. Estas iteraciones iniciales no deben entenderse como una modificación del estimador, sino simplemente como un modo de obtener unos valores iniciales suficientemente buenos. Cuando se realiza un proceso iterativo y se alcanza la convergencia las propiedades del punto de convergencia quedan determinadas únicamente por la función  $p_e$  que se empleó en las últimas iteraciones; todo lo anterior es simplemente valores aproximados.



## Sobre el método de los mínimos cuadrados

A. M. Legendre (1805)

EN la mayoría de los problemas en los que se trata de obtener a partir de medidas dadas por la observación los resultados más exactos que éstas puedan ofrecer, casi siempre terminamos llegando a un sistema de ecuaciones de la forma

$$E = a + bx + cy + fz + \&c.$$

en las cuales  $a, b, c, f, \&c.$  son coeficientes conocidos, que varían de una ecuación a otra, y  $x, y, z, \&c.$  son las incógnitas que hay que determinar con la condición de que el valor de  $E$  se reduzca, para cada ecuación, a una cantidad nula o muy pequeña.

Si tenemos el mismo número de ecuaciones que de incógnitas  $x, y, z, \&c.$ , no hay ninguna dificultad en la determinación de las incógnitas, y podemos hacer que los errores  $E$  sean absolutamente nulos. Pero la mayoría de las veces el número de ecuaciones es superior al de incógnitas, y es imposible eliminar todos los errores.

En esta circunstancia, que es la de la mayoría de los problemas físicos y astronómicos, en los que se busca determinar ciertos elementos importantes, interviene necesariamente la arbitrariedad en la distribución de los errores, y no podemos esperar que todas las hipótesis conduzcan exactamente a los mismos resultados; pero sobre todo hay que proceder de manera que los errores extremos, sin atender a su signo, queden encerrados en límites tan estrechos como sea posible.

De todos los principios que podemos proponer para este objetivo, pienso que no hay ninguno más general, más exacto, ni de una aplicación más fácil que aquél que hemos empleado en las investigaciones precedentes, y que consiste en hacer mínima la suma de los cuadrados de los errores. De esta forma se establece entre los errores una suerte de equilibrio que, impidiendo prevalecer a los extremos,

## Sobre el método de los mínimos cuadrados

es muy adecuado para revelar el estado del sistema que más se aproxima a la verdad.

La suma de los cuadrados de los errores  $E^2 + E'^2 + E''^2 + \&c.$ , siendo

$$\begin{aligned} & (a + bx + cy + fz + \&c.)^2 \\ & + (a' + b'x + c'y + f'z + \&c.)^2 \\ & + (a'' + b''x + c''y + f''z + \&c.)^2 \\ & + \&c. ; \end{aligned}$$

si buscamos su mínimo, haciendo variar  $x$  sola, tendremos la ecuación

$$0 = \int ab + x \int b^2 + y \int bc + z \int bf + \&c. ,$$

en la cual por  $\int ab$  entendemos la suma de los productos semejantes  $ab + a'b' + a''b'' + \&c.$  ; por  $\int b^2$  la suma de los cuadrados de los coeficientes de  $x$ , a saber  $b^2 + b'^2 + b''^2 + \&c.$ , y así sucesivamente.

De manera semejante, el mínimo respecto a  $y$  dará

$$0 = \int ac + x \int bc + y \int c^2 + z \int cf + \&c. ,$$

y el mínimo respecto a  $z$ ,

$$0 = \int af + x \int bf + y \int cf + z \int f^2 + \&c. ,$$

y vemos que los mismos coeficientes  $\int bc, \int bf, \&c.$  son comunes a dos ecuaciones, lo que contribuye a facilitar el cálculo.

En general, para formar la ecuación de mínimo respecto a una de las incógnitas, hay que multiplicar todos los términos de cada ecuación planteada por el coeficiente de la incógnita en esa ecuación, tomado con su signo, y efectuar la suma de todos estos productos.

Obtendremos de esta manera tantas ecuaciones de mínimo como incógnitas hay, y habrá que resolver estas ecuaciones por los métodos ordinarios. Pero podremos bien reducir todos los cálculos, tanto de las multiplicaciones como de la resolución, no incluyendo en cada operación mayor número de cifras enteras o decimales que el que pueda exigir el grado de aproximación de que es susceptible el problema.

Si por un azar singular, fuese posible satisfacer todas las ecuaciones haciendo nulos todos los errores, obtendremos igualmente este resultado mediante las ecuaciones de mínimo ; ya que si después de haber encontrado los valores  $x, y, z, \&c.$  que hacen nulos  $E, E', E'', \&c.$ ,

## Sobre el método de los mínimos cuadrados

hacemos variar  $x, y, z, \&c.$  en  $\delta x, \delta y, \delta z, \&c.$ , es evidente que  $E^2$  que era cero se transformará por esta variación en  $(a\delta x + b\delta y + c\delta z, \&c.)^2$ . Lo mismo sucederá para  $E'^2, E''^2, \&c.$  De donde vemos que la suma de los cuadrados de los errores tendrá por variación una cantidad de segundo orden respecto a  $\delta x, \delta y, \&c.$ ; lo que concuerda con la naturaleza del mínimo.

Si después de haber determinado todas las incógnitas  $x, y, z, \&c.$ , sustituimos sus valores en las ecuaciones planteadas, conoceremos los distintos errores  $E, E', E'', \&c.$  a los cuales este sistema da lugar, y que no pueden ser reducidos sin aumentar la suma de sus cuadrados. Si entre estos errores se encuentran algunos que juzgamos demasiado grandes para ser admisibles, entonces rechazaremos las ecuaciones que han producido esos errores, como procedentes de experimentos muy defectuosos, y determinaremos las incógnitas por medio de las ecuaciones restantes, que entonces darán errores mucho menores. Y hay que observar que no estamos obligados entonces a volver a empezar todos los cálculos; ya que, puesto que las ecuaciones de mínimo se forman por la adición de los productos efectuados en cada una de las ecuaciones planteadas, será suficiente con descartar de la adición los productos dados por las ecuaciones que dieron lugar a errores muy considerables.

La regla por la cual tomamos la media entre los resultados de diferentes observaciones, no es más que una consecuencia muy simple de nuestro método general, al que llamaremos *Método de los mínimos cuadrados*.

En efecto, si los experimentos han dado distintos valores  $a', a'', a''', \&c.$  para una cierta cantidad  $x$ , la suma de los cuadrados de los errores será  $(a' - x)^2 + (a'' - x)^2 + (a''' - x)^2 + \&c.$ , e igualando esta suma a un mínimo, tenemos

$$0 = (a' - x) + (a'' - x) + (a''' - x) + \&c.;$$

de donde resulta  $x = \frac{a' + a'' + a''' + \&c.}{n}$ , siendo  $n$  el número de observaciones.

Paralelamente, si para determinar la posición de un punto en el espacio, hemos encontrado, en un primer experimento, las coordenadas  $a', b', c'$ ; en un segundo, las coordenadas  $a'', b'', c''$  y así sucesivamente; sean  $x, y, z$ , las verdaderas coordenadas de ese punto; entonces el error del primer experimento será la distancia del punto  $(a', b', c')$  al

## Sobre el método de los mínimos cuadrados

punto  $(x, y, z)$ ; el cuadrado de esta distancia es

$$(a' - x')^2 + (b' - y')^2 + (c' - z')^2,$$

y la suma de los cuadrados semejantes siendo igual a un mínimo, obtenemos tres ecuaciones que dan  $x = \frac{\int a}{n}$ ,  $y = \frac{\int b}{n}$ ,  $z = \frac{\int c}{n}$ , siendo  $n$  el número de puntos dados por los experimentos. Estas fórmulas son las mismas que aquéllas por las cuales encontramos el centro de gravedad común de varias masas iguales, situadas en los puntos dados; de donde vemos que el centro de gravedad de un cuerpo cualquiera participa de esta propiedad general.

*Si dividimos la masa de un cuerpo en moléculas iguales y suficientemente pequeñas para que sean consideradas como puntos, la suma de los cuadrados de las distancias de las moléculas al centro de gravedad será un mínimo.*

Vemos entonces que el método de los mínimos cuadrados permite conocer, de alguna manera, el centro alrededor del cual se organizan todos los resultados proporcionados por los experimentos, de manera que tienden a separarse de él lo menos posible. La aplicación que vamos a hacer de este método a la medida del meridiano, permitirá mostrar con toda claridad su simplicidad y fecundidad.

---

Sigue una aplicación a la medida de grados de meridiano.

# Índice de símbolos

## 1 Linealización

- $\approx$  Relación aproximada (normalmente una función linealizada o resultados derivados), 3
- $\Delta x$  Incremento de la variable respecto al valor aproximado, 3
- $\frac{\partial f}{\partial x}$  Derivada parcial de una función respecto de una de sus variables, 5

## 2 Errores aleatorios

- $\varepsilon$  Error, 7
- $l$  Valor real de la magnitud observada, 7
- $\ell$  Valor observado, 7
- $f()$  Función de densidad, 8
- $P()$  Función probabilidad (función de conjunto), 9
- $[a, b]$  Intervalo cerrado  $a \leq x \leq b$ , 9
- $F()$  Función de distribución, 10
- $(\ell, m)$  Valores observados de dos magnitudes, 14

## 3 Estimación

- $l$  Valor ajustado de la magnitud observada, 22
- $\varepsilon$  Valor ajustado del error, 22
- $v$  Residuo, 22
- $x$  Valor ajustado en general, 26
- $f_x(\ell)$  Función de densidad de la observación para un valor real  $x$ , 27
- $g(l)$  Función de densidad del valor real  $l$  según los valores observados, 27
- $\propto$  Proporcional a, 30
- $m$  Número de valores observados, 31
- $n$  Número de magnitudes independientes que determinan el problema, 31
- $f_{l_1, \dots, l_m}(l_1, \dots, l_m)$  Función de densidad de las observaciones para unos valores reales, 31
- $G(l)$  Función de distribución, integral de  $g(l)$ , 33
- $h()$  Función de densidad a priori de los valores reales, 36
- $h$  Semiamplitud del intervalo de confianza, 37

## Índice de símbolos

$\mathcal{F}()$  Función de densidad conjunta de valores reales y valores observados, 39

### 4 Algunas distribuciones

$\sigma$  Desviación típica, 41

$\mu$  Media, 42

$N(\mu, \sigma)$  Distribución normal de media  $\mu$  y desviación típica  $\sigma$ , 42

$\sim$  Sigue una distribución, 42

$\bar{\ell}$  Media de los valores  $\ell$ , 42

$\chi^2$  Distribución  $\chi^2$  (función de densidad), 43

$X^2$  Distribución  $\chi^2$  (función de distribución), 43

### 5 Estimadores

S Factor de escala o similar, 48

$E[]$  Esperanza matemática (media), 49

$(x_1, \dots, x_n)$  Parámetros independientes que determinan el problema, 51

$f_0()$  Distribución de las observaciones cuando  $l = 0$ , 52

$\psi(\varepsilon)$   $f_0(\varepsilon) = ke^{-\psi(\varepsilon)}$ , 53

### 7 El estimador mínimo cuadrático

$x_0$  Valor aproximado del parámetro, 62

$l_0$  Valor aproximado de la magnitud observada, 63

L  $\ell - l_0$ , 63

V Vector de residuos, 64

A Matriz de coeficientes del sistema linealizado, 64

$X_\Delta$  Vector de incrementos  $\Delta x$  de los valores ajustados respecto a los valores aproximados, 64

L Vector de valores L, 64

X Vector de valores ajustados, 64

$\mathcal{L}$  Vector de valores observados, 64

$L_0$  Vector de valores aproximados, 64

N Matriz de las ecuaciones normales, 66

### 8 Cuadrilátero

$\sigma_\alpha$  Desviación típica de las medidas angulares, 69

$\sigma_\delta$  Desviación típica de las medidas de distancia, 69

$\ell'$  Observaciones con desviación típica normalizada, 70

$\sigma_\alpha, \sigma_\beta, \dots, \sigma_\lambda$  Desviaciones típicas de distintos conjuntos de observaciones, 70

$\sigma_0$	Desviación típica de referencia, 71
$v'$	Residuo de las observaciones $\ell'$ , 71
$a', L'$	Coefficientes y término independiente de las ecuación linealizada de $v'$ , 71
$p$	Peso, 71
$\Sigma_{\ell\ell}$	Matriz de covarianzas de las variables $\ell$ , 72
$\sigma_{ij}$	Covarianza, 72
$r$	Redundancias de un problema, 74
$B$	Matriz de las ecuaciones de condición, 76
$\Delta$	Vector de incrementos $\Delta l$ , 76
$d$	Vector de valores independientes de las ecuaciones de condición, 76
$K$	Vector de multiplicadores de Lagrange, 77

## 9 Ángulos y distancias

$S$	Distancia horizontal entre dos puntos, 81
$\Sigma$	Desorientación, 82
$\theta$	Azimut, 82
$L$	Lectura angular horizontal, 82

## 10 Regiones de confianza

$\Sigma_{xx}, \sigma_{x_i}^2, \sigma_{x_i x_j}$	Matriz de covarianzas y sus elementos, 91
$M_x$	Vector de medias, 92
$Q$	Matriz ortogonal (de diagonalización), 93
$\Lambda$	Matriz diagonal, 93
$Y$	Vector de variables independientes, 93
$\Sigma_{xx}$	Matriz de covarianzas de los valores reales $x$ , 97
$k, p(\%)$	Probabilidad en tanto por ciento de los intervalos o regiones $ k\sigma $ , 97
$\text{Tr}()$	Traza de una matriz, 98
$Z$	Punto de uno de los ejes de la elipse o elipsoide, 99
$\theta$	Ángulos que forma un eje de la elipse con el eje $x$ , 99
$L_r$	Vector de valores reales $l$ , 101
$\Sigma_{ll}$	Matriz de covarianzas de los valores reales $l$ , 101
$E$	Vector de errores $\varepsilon$ , 102
$\Sigma_{xx}$	Matriz de covarianzas de los valores ajustados $x$ , 102
$\Delta_x$	Vector de diferencias de unos valores $x$ cualesquiera respecto a los valores reales $x$ , 103
$X_r$	Vector de valores reales de los parámetros, 103
$L_{aj}$	Vector de valores ajustados de las observaciones, 103
$\Sigma_{ll}$	Matriz de covarianzas de los valores ajustados $l$ , 103

## Índice de símbolos

- $\Sigma_{\varepsilon\varepsilon}$  Matriz de covarianzas de los errores, 105  
 $E[\ ]_{\varepsilon_1}$  Esperanza de una variable para un valor concreto de  $\varepsilon_1$ , 107  
 $r_i$  Redundancia parcial, 110  
 $\rho$  Coeficiente de correlación, 113

### 11 Altimetría y otros

- H Desnivel, 117  
D Distancia entre dos puntos, 117  
V Lectura angular cenital, 117  
R Radio de la Tierra, 117  
 $x, y$  Error de centrado en una estación, 120  
 $z$  Error constante de la mira, 121  
 $\omega, \varphi$  Componentes del error de verticalidad de una estación, 122  
 $L^*, V^*$  Lecturas horizontal y cenital que corresponderían a  $\omega = \varphi = 0$ , 122

### 12 Transformación de semejanza plana

- $P_1, \dots, P_k$  Conjunto de puntos, 125  
 $P_i$  Conjunto de coordenadas de los puntos en el sistema  $i$ , 126  
R Matriz de rotación, 126  
T traslación, 126  
 $\lambda$  Factor de escala, 126  
 $\alpha$  Ángulo de giro, 127  
 $T_1, T_2$  Traslaciones en los sistemas 1 y 2, 128  
 $G = (x_g, y_g)$  Centro de gravedad, 128  
 $1''$  Coordenadas transformadas de los puntos 1 antes de aplicar  $T_2$ , 128  
 $1', 2'$  Coordenadas  $P_1 - T_1, P_2 - T_2$ , 129  
 $\delta$  Diferencia (residuo) entre las coordenadas en los sistemas 2 y 1 de un punto una vez transformadas a un sistema común, 129  
 $\delta_r, \delta_t$  Componentes radial y tangencial de  $\delta$ , 129  
 $\sigma_\delta^2$  Varianza de la diferencia  $\delta$ , 132  
 $\theta$  Ángulo de una semirrecta respecto a una dirección origen de ángulos, 132  
 $G_1, G_2, \lambda, \alpha$  Solución completa de la semejanza bidimensional, 138  
 $\sigma_T^2, \sigma_\alpha^2, \sigma_\lambda^2$  Varianzas de los parámetros de transformación, 142  
 $\Sigma_{X_2 \times X_2}$  Matriz de covarianzas de un punto transformado, 143

### 13 Otras transformaciones de coordenadas

- $R_0, R_\Delta$  Matriz de rotación tridimensional aproximada y matriz incremento, 145  
 $\omega, \varphi, \kappa$  Parámetros de una matriz de rotación próxima a I, 146

- $1'''$  Coordenadas  $\frac{r_2}{r_1}P_1'$ , 147  
 $R_1, R_2$  Matrices de rotación particulares en los sistemas 1 y 2, 149  
 $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  Matriz de afinidad, 151  
 $B$  Matriz de componente afín, 153  
 $\alpha, \beta$  Componentes de afinidad, 153  
 $\lambda_X$  Factor de escala local, 154  
 $d$  Dimensión, 154  
 $A_X$  Matriz de la transformación linealizada en el punto  $X$ , 154  
 $\Sigma_{\delta\delta}$  Matriz de covarianzas de la diferencia  $\delta$ , 155

**14 Cambio de los parámetros fijos, I: Puntual**

- $A$  Matriz de cambio de definición del sistema de coordenadas, 158  
 $X', Y'$  Coordenadas numéricamente iguales que  $X, Y$  pero cambiando la definición del sistema de coordenadas, 160

**15 Cambio de los parámetros fijos, II: Global**

- $M$  Matriz de cambio de definición del sistema de coordenadas,  $= I - A$ , luego  $= A - I$ , 176  
 $k_i$  Derivada de  $F$  (v. infra) respecto de  $x_i$ , 187  
 $\varepsilon$  Magnitud de la corrección global, 187  
 $F(x_i, \dots, x_n) = C$  Condición definitoria del sistema de coordenadas, 188  
 $\alpha_j$  Variación en cada coordenada causada por la corrección  $\varepsilon$ , 196

**16 Estimación de la varianza**

- $h_0$  Distribución a priori de  $\sigma_0$ , 204  
 $t$  Equivale a  $\sigma_0^2$ , 205  
 $g(\sigma_0^2)$  Densidad de probabilidad estimada para  $\sigma_0^2$ , 206  
 $z$  Variable auxiliar  $\Sigma v^2 / (2t)$  (luego  $\Sigma v^2 / t$ ), 208  
 $\sigma_0^2$  Valor ajustado de  $\sigma_0^2$ , 208  
 $c$   $\sigma_0^2 / \sigma_0^2$ , 211  
 $P_d(k)$  Probabilidad para región  $d$ -dimensional de semiejes  $k\sigma_i$ , 211  
 $k(95,5\%)$  Valor de  $k$  en función de  $r$  para el cual se tiene una probabilidad del 95,5%, 212  
 $\sigma_{0\alpha}, \sigma_{0\beta}, \dots, \sigma_{0\lambda}$  Valores  $\sigma_0$  para distintos conjuntos de observaciones, 216

**18 Interpretación geométrica**

- $\Omega$  Subconjunto de valores reales posibles, 233  
 $\bar{\Omega}$  Subespacio ortogonal a  $\Omega$ , 235  
 $\varepsilon_i''$  Componentes del vector  $E$  en un sistema ortonormal en el cual una de las componentes está contenida en  $\Omega$ , 237

## Índice de símbolos

- $\Omega - \varepsilon_1''$  Subespacio formado por los vectores de  $\Omega$  perpendiculares a  $\varepsilon_1'' \in \Omega$ , 237  
 $\alpha$  Complementario del ángulo que un eje  $l$  forma con  $\Omega$ , 238  
 $\bar{\varepsilon}_{1\Omega}$  Proyección de la componente  $\varepsilon_1$  sobre  $\Omega$ , 238  
 $P_\Omega$  Matriz de proyección sobre  $\Omega$ , 238  
 $\bar{a}$  Vector que expresa la libertad del sistema, 241

### 19 Estimadores para otras distribuciones, I

- $\bar{\varepsilon}$  Estimador que sigue la misma distribución que los errores originales, 247  
 $\tau$  Una magnitud proporcional a la dispersión de una distribución, 248  
 $\mathcal{F}()$  Función de distribución de  $|\varepsilon|$ , 249  
 $\bar{\mathcal{F}}()$   $1 - \mathcal{F}$ , 249  
 $q(x)$  Polinomio, 250  
( $-a_i$ ) Fila  $i$  de la matriz  $A$ , 253  
 $A'$  Matriz  $A$  (no.) multiplicada cada fila por una cierta constante, 254  
 $\Psi'$  Vector de valores  $\psi'(L_i)$ , 254  
 $\bar{\mathcal{F}}, \psi'$  Muestra de funciones  $\bar{\mathcal{F}}$  y  $\psi'$ , 256  
 $p()$  Función de peso, 256  
 $\bar{\mathcal{F}}(x)$  Función  $\bar{\mathcal{F}}$  de ejemplo, 261  
 $p_e()$  Función de pesos del estimador, 264  
 $F^{-1}()$  Función inversa de  $F$  (para generar valores siguiendo la distribución), 267

### 20 Estimadores para otras distribuciones, II

- $T_{xx}$  Matriz  $(\tau_{ij})$  de los valores reales  $x$ , 271  
 $A'_\tau$  Matriz  $A$  dividida cada fila por  $\sqrt{p_\tau}$ , 272  
 $p_\tau()$  Función de peso para la estimación de los valores  $\tau_x$ , 272  
 $\mathcal{G}_x()$  Función de distribución de  $|x - \bar{x}|/\tau_x$ , 272  
 $T_{ll}$  Matriz  $(\tau_{ij})$  de los valores reales  $l$ , 274  
 $\tau_0$  Parámetro a estimar, 279  
 $\mathcal{F}_{vr}()$  Función de densidad de  $|v_i|/\sqrt{r_i}$ , 281  
 $s_i$  Lo que cuenta cada residuo en la función  $\mathcal{F}_{vr}$ , 281  
 $\mathcal{U}()$  Distribución  $\mathcal{F}$  de los valores  $|v_i|/\sqrt{r_i}$  obtenidos, 281  
 $c$   $\tau_{vr}/\tau_0$  (luego  $\rho_{vr}/\tau_0$ ), 281  
 $k_c$  Constante incluida para que la estimación  $\tau_0$  sea centrada, 281  
 $q$  Cuantil, 282  
 $q_1$  Cuantil  $\mathcal{F}_{vr}(\tau_{vr})$ , 282  
 $q_0$  Cuantil tal que  $c = 1$ , 283  
 $\rho$  Una magnitud proporcional a la dispersión de una distribución, 284

## Índice de símbolos

$\rho'$	Valor de $\rho$ de la distribución $\mathcal{U}$ , 284
$\tau_0^2$	$\tau_0^2$ , estimación centrada, 284
$c'$	$\rho_{vr} / \rho_\varepsilon$ , 288
$c'_0$	Valores $c'$ cuando $s_i \equiv 1$ , 288
$\mathcal{F}_{vr0}$	Distribución $\mathcal{F}_{vr}$ para $s_i \equiv 1$ , 288
$r_{i0}$	Redundancias obtenidas para $p_e \equiv 1$ , 290
$\mathcal{F}_{vr}^{-1}$	Función inversa de $\mathcal{F}_{vr}$ , para obtener las estimaciones $\tilde{\varepsilon}_i$ , 291
$v_{1,2}$	Residuo conjunto de dos observaciones relacionadas. (anál. $\varepsilon_{1,2}$ ), 292
$\mathcal{F}_{\varepsilon_{1,2}}$	Función $\mathcal{F}$ de las cantidades $\varepsilon_{1,2}$ , 292