El método de mínimos cuadrados

Curso de Estadística

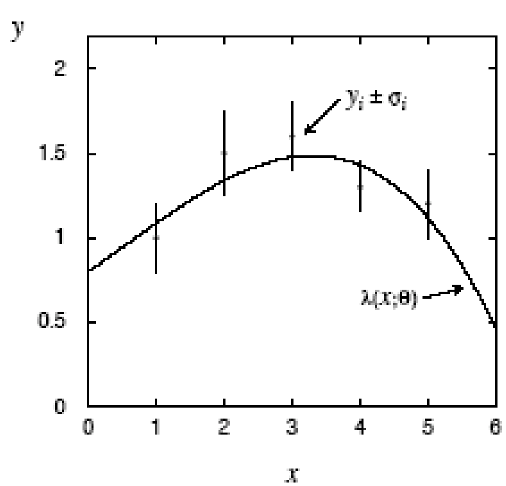
TAE, 2005

J.J. Gómez-Cadenas

Mínimos cuadrados y máxima verosimilitud

Teorema del límite central

Una medida y, puede considerarse como un variable aleatoria, distribuida gausianamente entorno a su valor verdadero λ, siempre que el error total sea la suma de un número grande de contribuciones pequeñas.



Considerar un conjunto $y_1, y_2, \ldots y_N$ de variables aleatorias independientes relacionadas con otra variable x_i que se asume conocida sin error.

Cada y_i tiene un valor medio λ_i (desconocido) y una varianza σ_i^2 (conocida)

Las N medidas de y_i pueden considerarse como la medida de un vector aleatorio N-dimensional con pdf

$$g(y_1,\ldots,y_n;\lambda_1,\ldots,\lambda_N,\sigma_1^2,\ldots,\sigma_N^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i-\lambda_i)^2}{2\sigma_i^2}\right)$$

Suponer además que el valor verdadero de las y_i es una función de la variable x que depende de un vector de parámetros desconocido en principio.

$$\lambda = \lambda(x_i;\vec{\theta}), \quad \vec{\theta} = (\theta_1,\ldots,\theta_m)$$

El objetivo del método de mínimos cuadrados es estimar el vector de parámetros θ.

Además, el método permite evaluar la bondad con la que la función λ(x,θ) ajusta los datos experimentales.

Para establecer el método tomamos logaritmos en la pdf que describe los datos:

$$\log(g) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i - \lambda_i)^2}{2\sigma_i^2}\right) \right) = A + \log L(\vec{\theta})$$

$$A = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} \right)$$

$$\log L(\vec{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\vec{\theta}))^2}{\sigma_i^2}$$

El principio de máxima verosimilitud establece que la pdf conjunta de las medidas (y por lo tanto la verosimilitud L) es máxima para los parámetros auténticos. Por lo tanto, para encontrar los parámetros maximizamos log L(θ) o bien minimizamos la cantidad:

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda(x_i;\vec{\theta}))^2}{\sigma_i^2}$$

Si las medidas no son independientes, pero pueden describirse por una pdf conjunta gausiana, con matriz de covarianza conocida, la definición anterior se generaliza a:

$$\chi^2(\vec{\theta}) = \sum_{i,j=1}^{N} (y_i - \lambda(x_i;\vec{\theta}))\left(V^{-1}\right)(y_j - \lambda(x_j;\vec{\theta}))$$

Que reduce a la expresión anterior si la matriz de covarianza es diagonal (medidas independientes)

Ajuste por mínimos cuadrados en el caso lineal

En el caso más general, un problema de ajuste se reduce a uno de minimización (del chi2). Sin embargo, cuando $\lambda(x;\theta)$ es una función lineal de los parámetros el problema puede tratarse analíticamente. Se trata del caso:

$$\lambda(x;\vec{\theta}) = \sum_{j=1}^{m} a_j(x)\theta_j$$

donde $a_j(x)$ son funciones de x.

NB: Requerimos que $\lambda(x;\theta)$ sea lineal en los parámetros, no que las funciones $a_j(x)$ sean lineales en x. Por ejemplo:

$$\lambda(x;\vec{\theta}) = e^{-x}\theta_1 + \sin(x)\theta_2 + x^2\theta_3 \quad \text{lineal en } \vec{\theta}$$

$$\lambda(x;\vec{\theta}) = x\sqrt{\theta_1} + \theta_2^2 + \theta \qquad\qquad \text{no lineal en } \vec{\theta}$$

El valor de la función $\lambda(x;\theta)$ en un punto dado xi es:

$$\lambda(x_i;\vec{\theta}) = \sum_{j=1}^{m} a_j(x_i)\theta_j = \sum_{j=1}^{m} A_{ij}\theta_j, \quad A_{ij} = a_j(x_i)$$

En este caso, la expresión general:

$$\chi^2(\vec{\theta}) = \sum_{i,j=1}^{N} (y_i - \lambda(x_i;\vec{\theta}))\left(V^{-1}\right)(y_j - \lambda(x_j;\vec{\theta}))$$

reduce a (en notación matricial):

$$\chi^2(\vec{\theta}) = (\vec{y} - \vec{\lambda})^T V^{-1}(\vec{y} - \vec{\lambda}) = (\vec{y} - A\vec{\theta})^T V^{-1}(\vec{y} - A\vec{\theta})$$

Para encontrar los parámetros minimizamos el chi2

$$\nabla\chi^2(\vec{\theta}) = -2(A^T V^{-1}\vec{y} - A^T V^{-1} A\vec{\theta}) = 0$$

Si $A^T V^{-1} A$ no es singular podemos resolver para los parámetros

$$\vec{\theta} = (A^T V^{-1} A)^{-1} A^T V^{-1}\vec{y} = B\vec{y}$$

Es decir los parámetros θ son funciones lineales de las medidas y.

Para encontrar la matrix de covarianza de los parámetros propagamos errores

$$\vec{\theta} = (A^T V^{-1} A)^{-1} A^T V^{-1} \vec{y} = B\vec{y}$$

$$U = BVB^T = (A^T V^{-1} A)^{-1}$$

Si λ(x;θ) es lineal en θ el chi2 es cuadrático en θ. Expandiendo en Taylor entorno a los parámetros (en el mínimo la derivada se anula):

$$\chi^2(\vec{\theta}) = \chi^2(\hat{\theta}) + \frac{1}{2}\sum_{i,j=1}^{m}\left[\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right]_{\theta=\vec{\theta}} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)$$

$$= \chi^2(\hat{\theta}) + \sum_{i,j=1}^{m} U_{ij}^{-1}$$

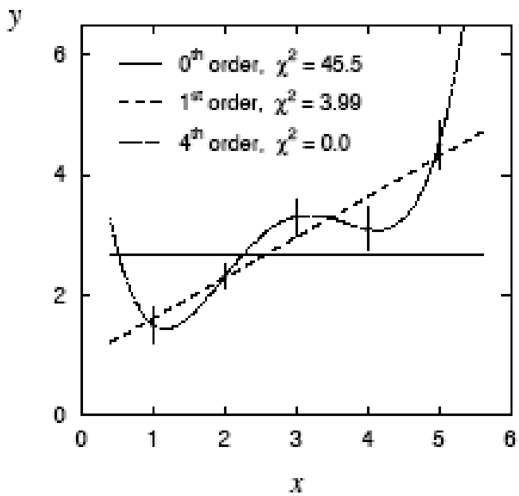
Por lo tanto : $\chi^2(\vec{\theta}) = \chi^2(\hat{\theta}) + 1 = \chi^2_{\min} + 1 \Rightarrow \hat{\theta} \rightarrow \hat{\theta} \pm \hat{\sigma}$

Corresponde a los contornos en el espacio de parámetros cuyas tangentes se separan una desviación estándar de los parámetros estimados en el mínimo

Ejemplo: Ajuste a un polinomio

$$\lambda(x;\theta_0,\ldots,\theta_m) = \sum_{j=0}^{m} \theta_j\, x^j$$

$$a_j(x) = x^j$$



0th order (1 parameter)

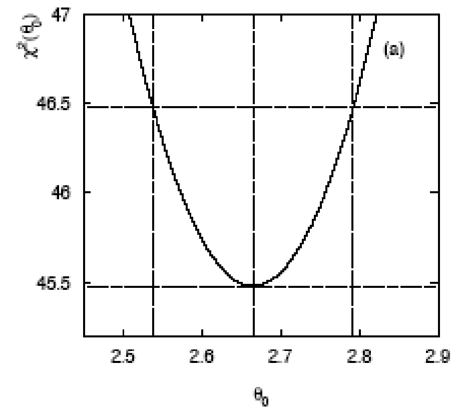
1st order (2 parameters)

4th order (5 parameters)

1-parameter fit (i.e. horizontal line):

$$\hat{\theta}_0 = 2.66 \pm 0.13$$

$$\chi^2_{\min} = 45.5$$



$\sigma_{\hat{\theta}_0}$ from $\chi^2(\hat{\theta}_0 \pm \sigma_{\hat{\theta}_0}) = \chi^2_{\min} + 1.$

2-parameter case (line with nonzero slope):

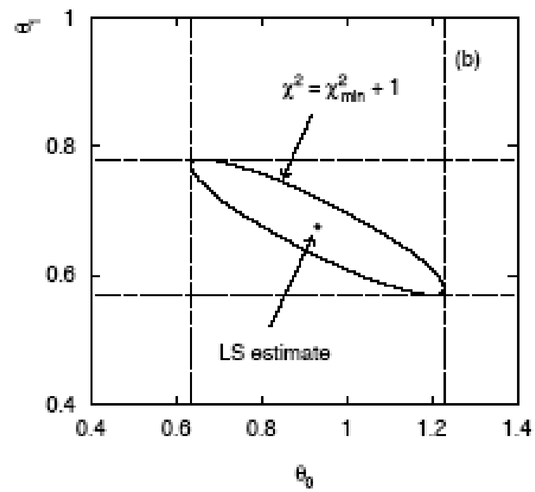
$$\hat{\theta}_0 = 0.93 \pm 0.30,$$

$$\hat{\theta}_1 = 0.68 \pm 0.10$$

$$\widehat{\mathrm{cov}}[\hat{\theta}_0, \hat{\theta}_1] = -0.028$$

$$r = -0.90$$

$$\chi^2 = 3.99$$



Calidad de un ajuste

Si en nuestro problema:

Los datos yi, i=1,2,...N son gausianos

la hipótesis $\lambda(x;\theta)$ es lineal en los parámetros θ_i, i=1,2,...,m

La pdf que describe la hipótesis (el modelo) $\lambda(x;\theta)$) es correcta:

Entonces el χ^2_{min} sigue una distribución Chi2 con nd = N-m grados de libertad

El valor-P o nivel de confianza es, por definición:

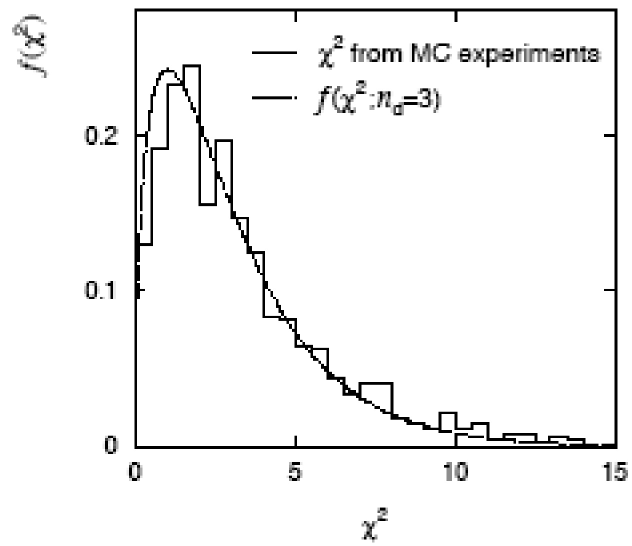
$$P = \int_{\chi^2}^{+\infty} f(z;n_d)dz$$

Donde f(z;nd) es la distribución Chi2 con nd grados de libertad.

Ajuste al polinomio con 2 parámetros

$$\chi^2_{\min} = 3.99,\ N - m = 3 \rightarrow P = 0.263$$

Simulación MC del ajuste a 2 parámetros. 26.3 % de las veces el ajustes tendrá un c2min más alto.



Ajuste al polinomio con 1 parámetro

$$\chi^2_{\min} = 45.5,\ N - m = 4 \rightarrow P = 3.1 \times 10^{-9}$$

Calidad del ajuste vs errores pequeños

El hecho de que un ajuste arroje errores pequeños no implica que el ajuste sea bueno (ni al contrario)

Curvatura del χ^2 cerca del mínimo --> tamaño del error (estadístico)

Valor del c2min --> calidad del ajuste

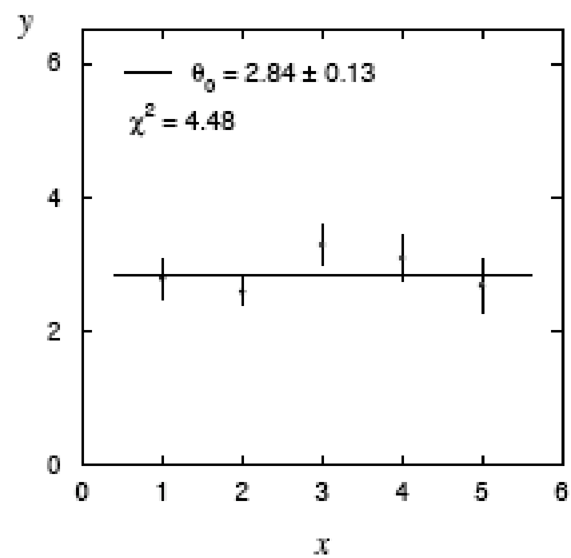
Si en el ajuste polinómico del ejemplo movemos los puntos para alinearlos más e concordancia con la hipótesis de una recta con pendiente nula (manteniendo el tamaño de los errores)

$\hat{\theta}_0 = 2.84 \pm 0.13$

$\chi^2_{\min} = 4.48$

Variance same as before,

now $\chi^2_{\min}$ ‘good’.



La varianza del estimador (su error estadístico) nos dice:

Si el experimento se repite muchas veces cual es la dispersión entorno al valor estimado q.

No nos dice si la hipótesis es correcta.

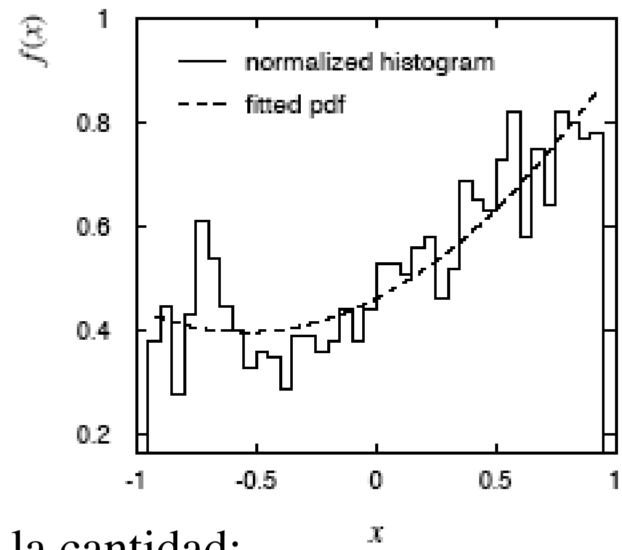
El valor-P (nivel de confianza, probablidad del c2) nos dice:

Si la hipótesis es correcta y el experimento se repite muchas veces, que fracción de los sucesos arrojará igual o peor acuerdo entre los datos y la hipótesis, de acuerdo con el c2min.

Un valor pequeño de P implica que la hipótesis es falsa o bien que hay errores sistemáticos que no se han tomado en cuenta.

Mínimos cuadrados con datos binados

Considerar un histograma con N bines y n entradas al que queremos ajustar un cierto modelo (es decir una hipotética pdf f(x; q)



Ajuste por mínimos cuadrados: Minimiza la cantidad:

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{\lambda_i(\vec{\theta})} = \sum_{i=1}^{N} \frac{(y_i - np_i(\vec{\theta}))^2}{np_i(\vec{\theta})}$$

Alternativamente (Mínimos cuadrados modificado)

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{y_i} = \sum_{i=1}^{N} \frac{(y_i - np_i(\vec{\theta}))^2}{y_i}$$

MCM se usa muy a menudo (es más cómodo) pero el problema es que el c2min resultante no tiene porqué estar distribuido c2 (podemos perder la capacidad de decidir sobre la calidad del ajuste

Combinación de medidas por mínimos cuadrados

Suponer que una cantidad de valores desconocido l ha sido medida N veces (en N experimentos diferentes), resultando en yi, si, i=1,2,...,N independientes.

Puesto que l es el mismo para todos los experimentos, l(x) = cte y por tanto:

$$\chi^2(\lambda) = \sum_{i=1}^{N} \frac{(y_i - \lambda)^2}{\sigma_i^2}$$

$$\frac{\partial \chi^2}{\partial \lambda} = -2\sum_{i=1}^{N} \frac{(y_i - \lambda)}{\sigma_i^2} = 0 \rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{N} y_i / \sigma_i^2}{\sum_{j=1}^{N} 1/\sigma_j^2}$$

Que no es sino la media pesada de las medidas.

La varianza se obtiene a partir de la segunda derivada:

$$V[\hat{\lambda}] = \frac{1}{\sum_{j=1}^{N} 1/\sigma_j^2}$$

Cuando las medidas yi no son todas independientes, pero la matriz de covarianza V se conoce, el procedimiento se generaliza fácilmente. Partiendo de:

$$\chi^2(\lambda) = \sum_{i,j=1}^{N} (y_i - \lambda) V_{ij}^{-1} (y_j - \lambda)$$

Y repitiendo el procedimiento de cancelar la derivada obtenemos:

$$\hat{\lambda} = \sum_{i=1}^{N} w_i y_i, \quad w_i = \frac{\sum_{j=1}^{N} V_{ij}^{-1}}{\sum_{k,l=1}^{N} V_{kl}^{-1}}, \quad \sum_{i=1}^{N} w_i = 1$$

La varianza se obtiene análogamente:

$$V[\hat{\lambda}] = \sum_{i,j=1}^{N} w_i V_{ij} w_j \rightarrow V[\hat{\lambda}] = \vec{w}^T V \vec{w}$$