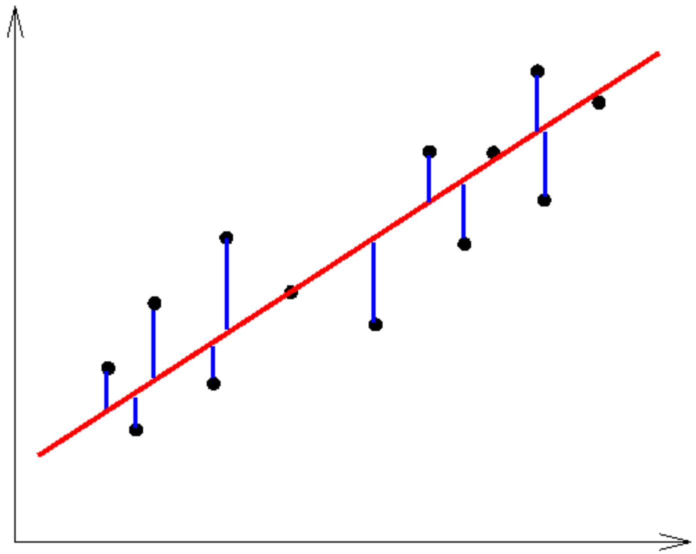
Tema 9 Método de mínimos cuadrados

1. Principio de mínimos cuadrados.
2. Conexión con el método ML
3. Modelos lineales en los parámetros
4. Formulación a partir de derivadas
5. Ejemplo. Ajuste a una línea recta
6. Modelos no lineales en los parámetros.
7. Bondad de los ajustes
8. Residuos y "*Pulls*"
9. Mínimos cuadrados con datos clasificados.
10. Resumen Propiedades.



1. Principio de mínimos cuadrados

Supongamos n medidas independientes $y_1, y_2, \ldots, y_n$ consideradas variables aleatorias y asociadas a n puntos observacionales $x_1, x_2, \ldots, x_n$ considerados fijos.

Los valores verdaderos de los observables $\mu_1, \mu_2, \ldots, \mu_n$ son desconocidos pero suponemos que pueden predecirse por un modelo teórico que lo predice para cada valor de x_i:

$$\mu_i = f\left(x_i, \theta_1, \theta_2, \ldots, \theta_k\right) = f\left(x_i, \overline{\theta}\right)$$

Función Hipótesis o modelo

Donde $\theta_1, \theta_2, \ldots, \theta_m$ con $m < n$ es un conjunto de parámetros desconocidos. De acuerdo con el ***Principio de Mínimos Cuadrados (LS)*** los mejores valores de los parámetros desconocidos son los que minimizan la cantidad:

$$\chi^2 = \sum_{i=1}^{n} w_i \left(y_i - f_i\right)^2 = \text{minimun}$$

Los parámetros que minimizan la función de χ^2, se denominan estimadores ***LS*** (*Least Squares*): $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$

- Los valores de los pesos w_i tienen en cuenta la precisión de las medidas: σ_i^2
- Bajo determinadas condiciones, el mínimo de la función encontrada χ^2 sigue una distribución de χ^2

1. Principio de mínimos cuadrados

- Si todas las medidas tienen la misma precisión o error → $\chi^2 = \sum_{i=1}^{n} (y_i - f_i)^2$

Mínimos cuadrados sin peso
(*Unweighted LS*)

- Si los errores son diferentes pero conocidos: $w_i = \dfrac{1}{\sigma_i^2}$ → $\chi^2 = \sum_{i=1}^{n} \left(\dfrac{y_i - f_i}{\sigma_i} \right)^2$

Mínimos cuadrados con peso
(*Weighted LS*)

- Si las variables observadas son de Poisson: $\sigma_i^2 \simeq f_i$ → $\chi^2 = \sum_{i=1}^{n} \left(\dfrac{y_i - f_i}{\sqrt{f_i}} \right)^2$

- Si f es una función complicada se suele utilizar la aproximación:

Mínimos cuadrados simplificado
(*Simplified LS*)

$\sigma_i^2 = y_i = n_i$ → $\chi^2 = \sum_{i=1}^{n} \left(\dfrac{y_i - f_i}{\sqrt{y_i}} \right)^2$

- Si las variables están correlacionadas: V → $\chi^2 = \sum_{i,j=1}^{n} (y_i - f_i) V_{ij}^{-1} (y_j - f_j)$

- Asumimos que las variables x_i son valores precisos sin error o despreciable frente al error de y_i
- El método no impone ningún tipo de distribución a las variables (*distribution-free*)
- Si las variables son gaussianas:
  - El mínimo $\chi^2_{\min}$ se distribuye como una distribución de chi-cuadrado.
  - Posibilidad de realizar un test del ajuste.

2. Conexión con el Método ML

Supongamos que las n medidas $y_1, y_2, \ldots, y_n$ se distribuyen gaussianamente alrededor de sus valores verdaderos aunque desconocidos μ_i y con varianza σ_i^2 conocida:

La pdf conjunta vendrá dada por:

$$g\left(y_1, y_2, \ldots, y_n, \mu_1, \mu_2, \ldots, \mu_n, \sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\frac{\left(y_i - \mu_i\right)^2}{\sigma_i^2}\right]$$

con $\mu_i = f\left(x_i, \theta_1, \theta_2, \ldots, \theta_m\right) = f\left(x_i, \bar{\theta}\right)$

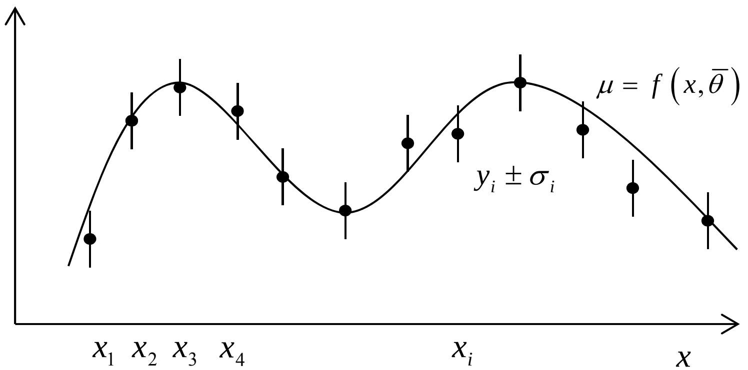
Tomando el logaritmo de la función pdf:

$$\ln L\left(\bar{\theta}\right) = \ln \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{\left(y_i - f\left(x_i, \bar{\theta}\right)\right)^2}{2\sigma_i^2}\right] = -\frac{n}{2}\ln 2\pi - \sum_{i=1}^{n} \ln \sigma_i - \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \bar{\theta}\right)\right)^2}{2\sigma_i^2}$$

Eliminando los términos aditivos que no dependen de $\bar{\theta}$

$$\ln L\left(\bar{\theta}\right) = -\frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \bar{\theta}\right)\right)^2}{\sigma_i^2} = -\frac{1}{2}\chi^2 \;\Rightarrow\; \boxed{\ln L\left(\bar{\theta}\right) = -\frac{1}{2}\chi^2}$$

Relación entre ML y LS



Maximizar $\ln L\left(\bar{\theta}\right)$ equivale a minimizar la función:

$$\boxed{\chi^2\left(\bar{\theta}\right) = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \bar{\theta}\right)\right)^2}{\sigma_i^2}}$$

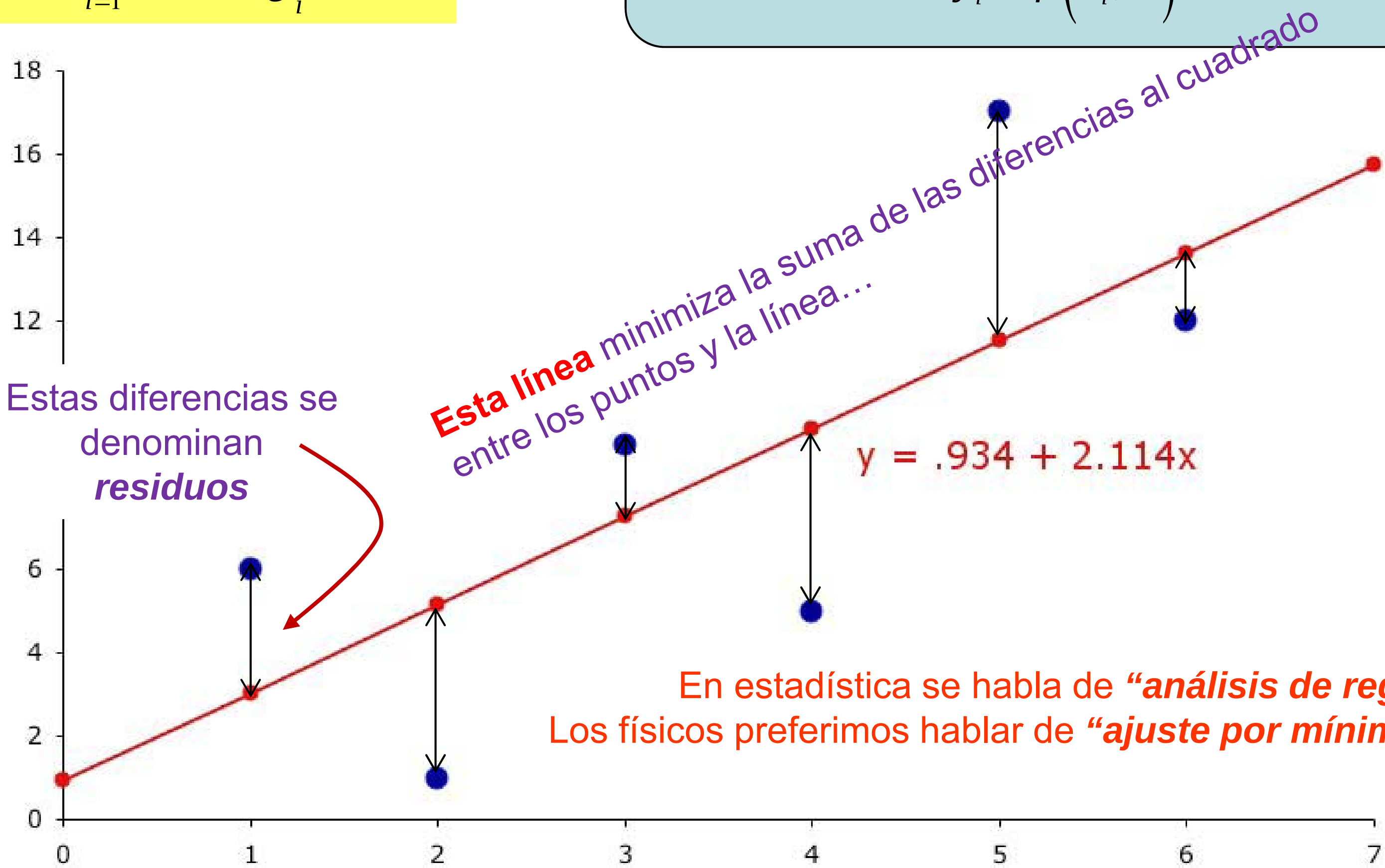
2. Conexión con el Método ML

Ejemplo: Ajuste a una línea recta

$$\chi^2\left(\overline{\theta}\right) = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \overline{\theta}\right)\right)^2}{\sigma_i^2}$$

Residuos = desviaciones de las observaciones respecto a los "valores calculados"

$$y_i - f\left(x_i, \hat{\overline{\theta}}\right)$$

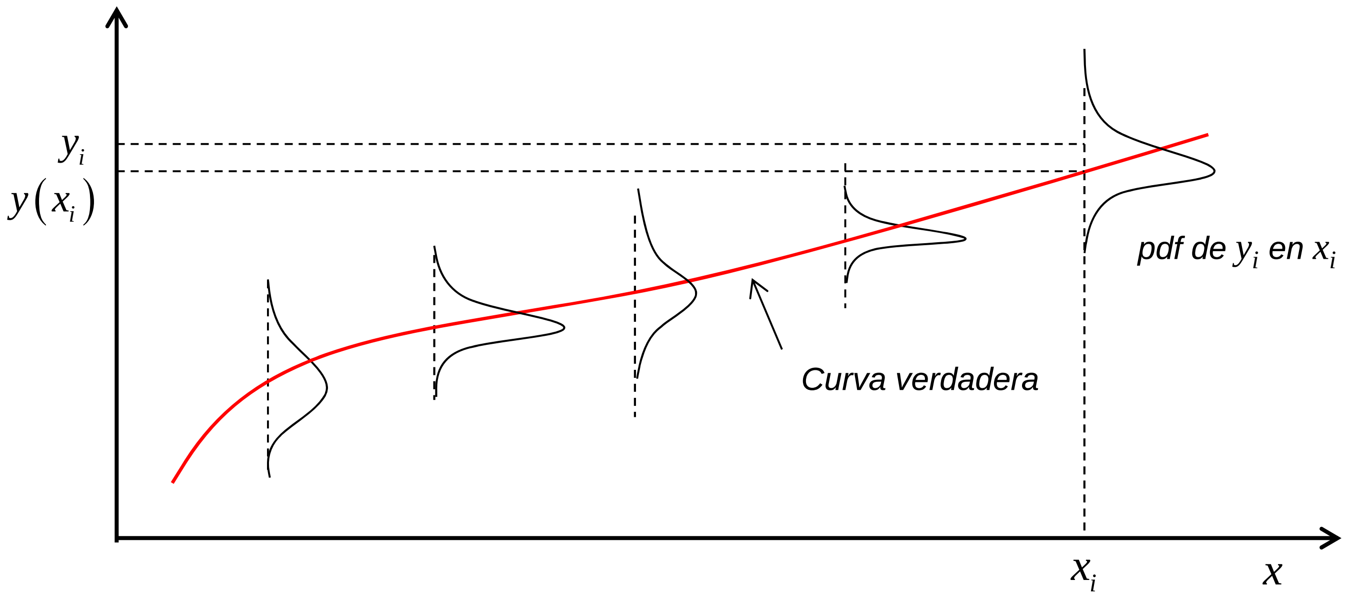


En estadística se habla de ***"análisis de regresión"***

Los físicos preferimos hablar de ***"ajuste por mínimos cuadrados"***

3. Modelo lineal en los parámetros

Las desviaciones de las medidas respecto a la curva verdadera se deben a errores en las medidas, pero conocemos las distribuciones, no necesariamente gaussianas, que siguen las medidas, y por tanto conocemos sus varianzas σ_i^2



Supongamos que la función modelo es ***lineal en los parámetros***:

$$y = f\left(x, \overline{\theta}\right) = \sum_{j=1}^{m} a_j\left(x\right)\theta_j$$

Las funciones $a_j\left(x\right)$:

- Son funciones conocidas de la variable x, no necesariamente lineales en x
- Linealmente independientes entre ellas.

3. Modelo lineal en los parámetros

Los valores de la función en cada punto x_i vendrán dados por:

$$f\left(x_i;\overline{\theta}\right)=\sum_{j=1}^{m}a_j\left(x_i\right)\theta_j=\sum_{j=1}^{m}A_{ij}\theta_j \qquad \text{con} \qquad A_{ij}=a_j\left(x_i\right)$$

Y la expresión de χ^2 ➡

$$\chi^2=\sum_{i=1}^{n}\frac{\left(y_i-f\left(x_i,\overline{\theta}\right)\right)^2}{\sigma_i^2}=\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(y_i-\sum_{j=1}^{m}A_{ij}\theta_j\right)^2$$

Derivando respecto a cada parámetro θ_k e igualando a cero:

$$\frac{\partial\chi^2}{\partial\theta_k}=\frac{\partial}{\partial\theta_k}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\left(y_i-\sum_{j=1}^{m}A_{ij}\theta_j\right)^2=\sum_{i=1}^{n}-\frac{2}{\sigma_i^2}\left(y_i-\sum_{j=1}^{m}A_{ij}\theta_j\right)A_{ik}=0 \qquad k=1,2,\ldots,m$$

Ecuaciones normales

$$\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\frac{A_{ik}A_{ij}}{\sigma_i^2}\right)\theta_j=\sum_{i=1}^{n}\frac{A_{ik}y_i}{\sigma_i^2} \qquad k=1,2,\ldots,m$$

Sistema de m ecuaciones con m incógnitas

3. Modelo lineal en los parámetros

Ecuaciones normales

$$\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\frac{A_{ik}A_{ij}}{\sigma_i^2}\right)\theta_j=\sum_{i=1}^{n}\frac{A_{ik}y_i}{\sigma_i^2}\qquad k=1,2,\ldots,m$$

$$\left.\begin{array}{l}
\displaystyle\sum_{i=1}^{n}\frac{A_{i1}A_{i1}}{\sigma_i^2}\theta_1+\sum_{i=1}^{n}\frac{A_{i1}A_{i2}}{\sigma_i^2}\theta_2+\cdots+\sum_{i=1}^{n}\frac{A_{i1}A_{im}}{\sigma_i^2}\theta_m=\sum_{i=1}^{n}\frac{A_{i1}y_i}{\sigma_i^2}\\
\displaystyle\sum_{i=1}^{n}\frac{A_{i2}A_{i1}}{\sigma_i^2}\theta_1+\sum_{i=1}^{n}\frac{A_{i2}A_{i2}}{\sigma_i^2}\theta_2+\cdots+\sum_{i=1}^{n}\frac{A_{i2}A_{im}}{\sigma_i^2}\theta_m=\sum_{i=1}^{n}\frac{A_{i2}y_i}{\sigma_i^2}\\
\qquad\vdots\qquad\qquad\qquad\vdots\qquad\qquad\qquad\vdots\qquad\qquad\qquad\vdots\\
\qquad\vdots\qquad\qquad\qquad\vdots\qquad\qquad\qquad\vdots\qquad\qquad\qquad\vdots\\
\displaystyle\sum_{i=1}^{n}\frac{A_{im}A_{i1}}{\sigma_i^2}\theta_1+\sum_{i=1}^{n}\frac{A_{im}A_{i2}}{\sigma_i^2}\theta_2+\cdots+\sum_{i=1}^{n}\frac{A_{im}A_{im}}{\sigma_i^2}\theta_m=\sum_{i=1}^{n}\frac{A_{im}y_i}{\sigma_i^2}
\end{array}\right\}$$

Sistema lineal de ecuaciones inhomogéneo de m incógnitas, solución exacta y única

En notación matricial:

El modelo lineal se puede expresar como:

$$f\left(x_i;\bar{\theta}\right)=\sum_{j=1}^{m}a_j\left(x_i\right)\theta_j=\sum_{j=1}^{m}A_{ij}\theta_j$$

$$\begin{pmatrix} f_1\\ f_2\\ \vdots\\ \\ f_n\end{pmatrix}=\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m}\\ A_{21} & A_{22} & \cdots & A_{2m}\\ \vdots & \vdots & & \vdots\\ A_{n1} & A_{n2} & \cdots & A_{nm}\end{pmatrix}\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_m\end{pmatrix}$$

$$\bar{\theta}=\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_m\end{pmatrix}\qquad A=\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m}\\ A_{21} & A_{22} & \cdots & A_{2m}\\ \vdots & \vdots & & \vdots\\ A_{n1} & A_{n2} & \cdots & A_{nm}\end{pmatrix}$$

$$\bar{f}=A\bar{\theta}$$

3. Modelo lineal en los parámetros

Ecuaciones normales

Y la cantidad a minimizar:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \overline{\theta}\right)\right)^2}{\sigma_i^2} = \sum_{i,j=1}^{n} \left(y_i - f\left(x_i, \overline{\theta}\right)\right) V_{ij}^{-1} \left(y_j - f\left(x_j, \overline{\theta}\right)\right)$$

$$\overline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad V = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

$$\chi^2 = \left(\overline{y} - A\overline{\theta}\right)^T V^{-1} \left(\overline{y} - A\overline{\theta}\right)$$

Derivando e igualando a cero:

$$\nabla \chi^2 = -2\left(A^T V^{-1} \overline{y} - A^T V^{-1} A \overline{\theta}\right) = 0 \quad \Rightarrow \quad \left(A^T V^{-1} A\right) \overline{\theta} = A^T V^{-1} \overline{y}$$

Y siempre que la matriz $\left(A^T V^{-1} A\right)$ **no sea singular**

$$\hat{\overline{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \overline{y}$$

Válida para variables no independientes introduciendo la matriz de covarianzas

3. Modelo lineal en los parámetros

Valor esperado

$$\bar{y} = y + \bar{y} - y = A\theta + \bar{y} - y = A\theta + \bar{\varepsilon}$$

$$E\left[\bar{y}\right] = E\left[A\bar{\theta} + \bar{\varepsilon}\right] = A\bar{\theta} + E\left[\bar{\varepsilon}\right] = A\bar{\theta}$$

$$\hat{\bar{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \bar{y}$$

$$E\left[\hat{\bar{\theta}}\right] = E\left[\left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \bar{y}\right] = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} E\left[\bar{y}\right] = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} A\bar{\theta} = \bar{\theta}$$

Los estimadores ***LS*** resultan ser estimadores sin sesgo, suponiendo que el modelo sea el correcto

Varianza

En nuestro caso:

$$\hat{\bar{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} \bar{y} \quad \Rightarrow \quad V\left(\hat{\bar{\theta}}\right) = SV\left(\bar{y}\right)S^T$$

$$S_{ij} = \frac{\partial \theta_i}{\partial y_j} = \left[\left(A^T V^{-1} A\right)^{-1} A^T V^{-1}\right]_{ij} \qquad S = \left[\left(A^T V^{-1} A\right)^{-1} A^T V^{-1}\right]$$

Propagación de errores:

$$V\left(\bar{y}\right) = SV\left(\bar{x}\right)S^T$$

$$V_{kl}\left(\bar{y}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial y_k}{\partial x_i}\left|\frac{\partial y_l}{\partial x_j}\right| V_{ij}\left(x\right)$$

$$S_{ij} = \frac{\partial y_i}{\partial x_j}$$

$$V\left(\hat{\bar{\theta}}\right) = SV\left(\bar{y}\right)S^T = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1} V V^{-1} A \left(A^T V^{-1} A\right)^{-1} = \left(A^T V^{-1} A\right)^{-1}$$

Teorema de Markov-Gauss.- *"Bajo ciertas condiciones, de todos los estimadores lineales y unbiased, los estimadores LS tienen la varianza mínima posible"*

$$\boxed{V\left(\hat{\bar{\theta}}\right) = \left(A^T V^{-1} A\right)^{-1}}$$

4. Formulación a partir de derivadas

Los resultados anteriores se pueden obtener también a partir de las derivadas de la función χ^2

$$\chi^2 = \left(\bar{y} - A\bar{\theta}\right)^T V^{-1} \left(\bar{y} - A\bar{\theta}\right)$$

$$V\left(\hat{\bar{\theta}}\right) = \left(A^T V^{-1} A\right)^{-1}$$

$$\left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\hat{\bar{\theta}}} = -2A^T V^{-1}\left(\bar{y} - A\bar{\theta}\right) \quad \rightarrow \quad \left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\hat{\bar{\theta}}} = 2A^T V^{-1} A = 2V^{-1}\left(\hat{\bar{\theta}}\right)$$

Podemos calcular la matriz de covarianza mediante las derivadas sin necesidad de usar las matrices A y V $\rightarrow$

$$\boxed{V^{-1}\left(\hat{\bar{\theta}}\right) = \frac{1}{2}\left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\hat{\bar{\theta}}}}$$

Incluso la solución se puede expresar en términos de las derivadas:

Como no conocemos la solución de los parámetros calculamos las derivadas en un punto aproximado $\bar{\theta} = \bar{\theta}^0$

$$\hat{\bar{\theta}} = \left(A^T V^{-1} A\right)^{-1} A^T V^{-1}\bar{y} = 2\left(\left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}^0}\right)^{-1} A^T V^{-1}\bar{y} = \left(\left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}^0}\right)^{-1}\left[\left(\left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}^0}\right)\bar{\theta}^0 - \left(\left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\bar{\theta}^0}\right)\right] =$$

$$= \bar{\theta}^0 - \left(\left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}^0}\right)^{-1}\left(\left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\bar{\theta}^0}\right) \quad \leftarrow$$

Método de **Newton-Raphson** para ecuaciones lineales.
Exacto para el caso lineal

5. Ejemplo. Ajuste a una línea recta

El caso más sencillo de dependencia lineal en los parámetros es la línea recta

$$y = f\left(x;\overline{\theta}\right) = \sum_{i=1}^{m} a_i\left(x\right)\theta_i = a_1\left(x\right)\theta_1 + a_2\left(x\right)\theta_2 = a + bx$$

Es decir:

$$\theta_1 = a \; ; \; \theta_2 = b$$

$$a_1\left(x\right) = 1 \; ; \; a_2\left(x\right) = x$$

En notación matricial:

$$A_{ij} = a_j\left(x_i\right)$$

$$\overline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \qquad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{21} \\ \vdots & \vdots \\ A_{n1} & A_{21} \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \Rightarrow \overline{f} = A\overline{\theta} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}$$

Si las medidas son independientes, la matriz de covarianzas es diagonal y es más sencillo $V_{ii}\left(y\right) = \sigma_i^2$

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{y_i - a - bx_i}{\sigma_i}\right)^2 \Rightarrow \begin{matrix} \dfrac{\partial\chi^2}{\partial a} = 0 \\ \dfrac{\partial\chi^2}{\partial b} = 0 \end{matrix} \Rightarrow \left.\begin{matrix} -2\displaystyle\sum_{i=1}^{n}\frac{y_i - a - bx_i}{\sigma_i^2} = 0 \\ -2\displaystyle\sum_{i=1}^{n}\frac{x_i\left(y_i - a - bx_i\right)}{\sigma_i^2} = 0 \end{matrix}\right\}$$

Sistema de dos ecuaciones con dos incógnitas que se resuelve fácilmente

5. Ejemplo. Ajuste a una línea recta

Sistema de ecuaciones normales

$$\left.\begin{array}{l}\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)a+\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)b=\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\\ \left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)a+\left(\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right)b=\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}\end{array}\right\}$$

$$\hat{a}=\frac{\left(\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\right)-\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}\right)}{\left(\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)-\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)^2}$$

$$\hat{b}=\frac{\left(\sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)-\left(\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)}{\left(\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right)\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right)-\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)^2}$$

La matriz de covarianza vendrá dada (suponiendo todos los errores iguales): $\sigma_i=\sigma$

$$V\left(\hat{\theta}\right)=\left(A^T V^{-1}A\right)^{-1}=\left(A^T A\right)^{-1}\sigma^2=\sigma^2\left[\begin{pmatrix}1&1&\cdots&1\\ x_1&x_2&\cdots&x_n\end{pmatrix}\begin{pmatrix}1&x_1\\1&x_2\\ \vdots&\vdots\\1&x_n\end{pmatrix}\right]^{-1}=\sigma^2\begin{pmatrix}n&\sum x_i\\ \sum x_i&\sum x_i^2\end{pmatrix}^{-1}=$$

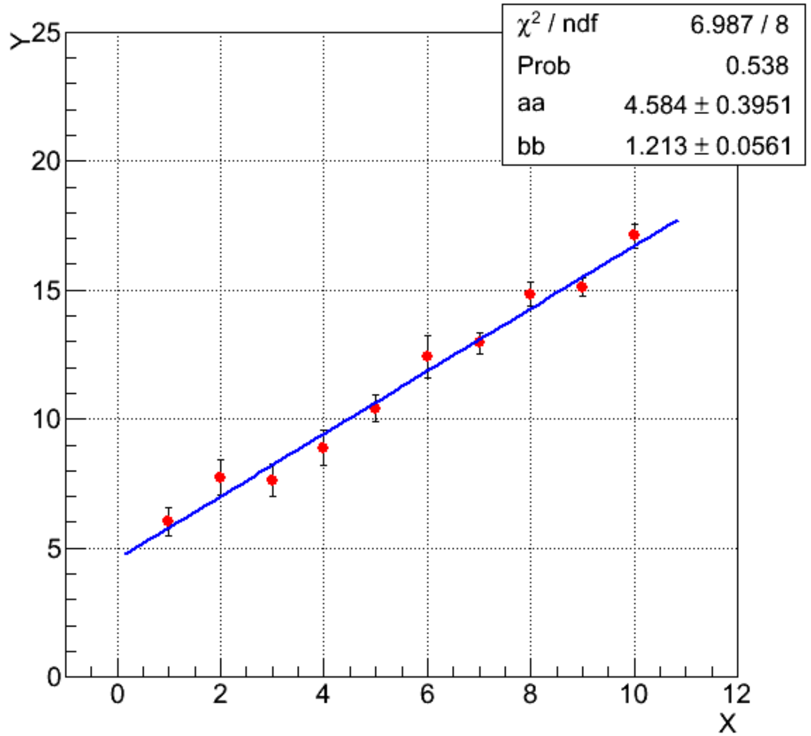
$$=\frac{\sigma^2}{n\sum x_i^2-\left(\sum x_i\right)^2}\begin{pmatrix}\sum x_i^2&-\sum x_i\\ -\sum x_i&n\end{pmatrix}=\frac{\sigma^2}{n\left(\overline{x^2}-\overline{x}^2\right)}\begin{pmatrix}\overline{x^2}&-\overline{x}\\ -\overline{x}&1\end{pmatrix}$$

$$\begin{pmatrix}V(\hat{a})&\operatorname{cov}\left(\hat{a},\hat{b}\right)\\ \operatorname{cov}\left(\hat{a},\hat{b}\right)&V\left(\hat{b}\right)\end{pmatrix}=\frac{\sigma^2}{n\left(\overline{x^2}-\overline{x}^2\right)}\begin{pmatrix}\overline{x^2}&-\overline{x}\\ -\overline{x}&1\end{pmatrix}$$

Si los errores no son iguales basta introducir los cambios:

$$\left\{\overline{x}\to\overline{x}=\frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2};\qquad \sigma^2\to\overline{\sigma^2}=\frac{\sum \sigma_i^2/\sigma_i^2}{\sum 1/\sigma_i^2}=\frac{1}{\sum 1/\sigma_i^2}\right.$$

5. Ejemplo. Ajuste a una línea recta



Valor de χ^2 en el mínimo (NB: no dividido por el los grados de libertad

```
 PARAM       CORRELATION COEFFICIENTS
             GLOBAL      1       2
            0.91298   1.000 -0.913
            0.91298  -0.913  1.000
 EXTERNAL ERROR MATRIX.    NDIM=   2    NPAR=   2    ERR DEF=1
 1.561e-001 -2.024e-002
 -2.024e-002 3.147e-003
 FCN=6.98702 FROM MIGRAD    STATUS=CONVERGED      31 CALLS         32 TOTAL
                    EDM=3.28338e-021    STRATEGY= 1      ERROR MATRIX ACCURATE
  EXT PARAMETER                     PARABOLIC         MINOS ERRORS
  NO.   NAME       VALUE            ERROR         NEGATIVE      POSITIVE
   1  aa          4.58428e+000   3.95139e-001
   2  bb          1.21304e+000   5.61001e-002
(Int_t)0
```

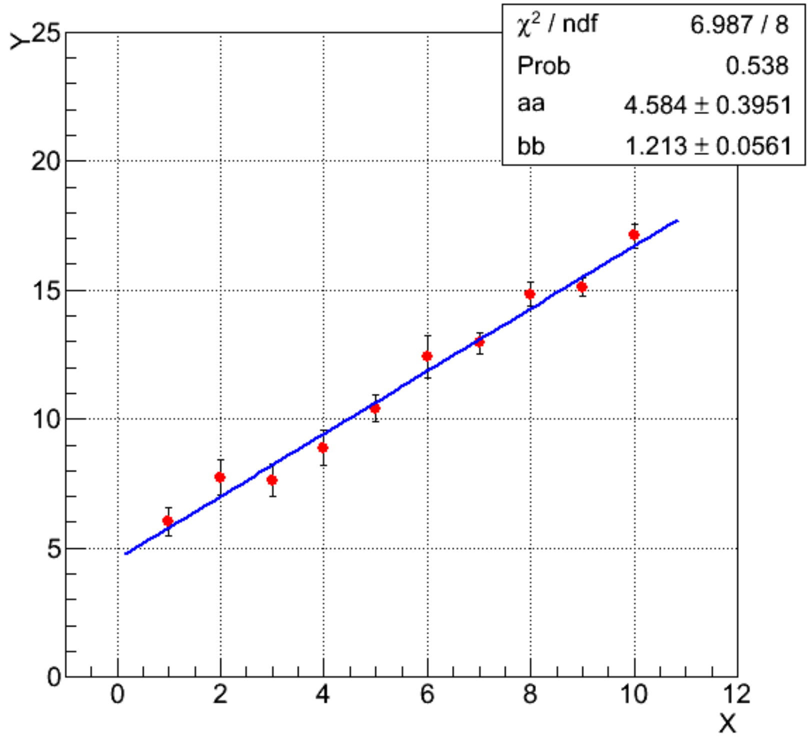
Valores de los parámetros y sus errores aproximados según MINUIT

***Status*: 'CONVERGED' puede ser '*FAILED*'**

Estimated Distance to Minimum (EDM)
debe ser orden O(10^{-6})

Error Matrix Quality
Debe ser 'ACCURATE', pero puede ser 'APPROXIMATE' si hay problemas

5. Ejemplo. Ajuste a una línea recta



Matriz de correlación ρ_{ij}

(Calculada como)

$$\rho_{ij} = V_{ij} / \sigma_i \sigma_j$$

```
 PARAMETER  CORRELATION COEFFICIENTS
       NO.  GLOBAL      1       2
        1  0.91298   1.000 -0.913
        2  0.91298  -0.913  1.000
 EXTERNAL ERROR MATRIX.    NDIM=   2    NPAR=  2    ERR DE
 1.561e-001 -2.024e-002
 -2.024e-002 3.147e-003
 FCN=6.98702 FROM MIGRAD    STATUS=CONVERGED      31 CALL
                    EDM=3.28338e-021    STRATEGY= 1
  EXT PARAMETER                       PARABOLIC          MINOS
  NO.   NAME        VALUE               ERROR        NEGATIVE
   1  aa          4.58428e+000   3.95139e-001
   2  bb          1.21304e+000   5.61001e-002
(Int_t)0
```

Matriz de covarianza

(calculada como)

$$V_{ij} = \frac{1}{2}\left(\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right)^{-1}$$

Errores simétricos
Symmetric errors calculados a partir de 2nd derivada de –ln(L) or χ^2

6. Modelos no lineales en los parámetros

Si la dependencia de la función modelo $f(x)$ con los parámetros no es lineal no hay solución exacta.

Minimización mediante un proceso iterativo hasta alcanzar la mejor aproximación.

De acuerdo con el **Principio de Mínimos Cuadrados**, la mejor estimación de los parámetros viene dada por los valores que minimizan la cantidad:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i, \bar{\theta}\right)\right)^2}{\sigma_i^2} \quad \Rightarrow \quad \chi^2 = \left(\bar{y} - \bar{f}\right)^T V^{-1} \left(\bar{y} - \bar{f}\right)$$

Partimos de una solución aproximada: $\bar{\theta}^0 = \left\{\theta_1^0, \theta_2^0, \ldots, \theta_m^0\right\}$

Derivando respecto a los parámetros y evaluando en dicha solución aproximada:

$$g_k\left(\bar{\theta}^0\right) = \left.\frac{\partial \chi^2}{\partial \theta_k}\right|_{\bar{\theta}=\bar{\theta}^0} = \sum_{i=1}^{n} \left(-\frac{2}{\sigma_i^2}\right)\left(y_i - f_i\left(x_i; \bar{\theta}^0\right)\right) \left.\frac{\partial f_i}{\partial \theta_k}\right|_{\bar{\theta}=\bar{\theta}^0}$$

Si $\bar{\theta}^0$ fuera el mínimo se anularían, como no lo es, buscamos un incremento tal que:

$$\bar{g}\left(\bar{\theta}^0 + \Delta\bar{\theta}^0\right) = 0$$

Desarrollamos en serie de potencias y nos quedamos a primer orden:

$$g_k\left(\bar{\theta}^0 + \Delta\bar{\theta}^0\right) = g_k\left(\bar{\theta}^0\right) + \left.\frac{\partial g_k}{\partial \theta_1}\right|_{\bar{\theta}^0} \Delta\theta_1^0 + \left.\frac{\partial g_k}{\partial \theta_2}\right|_{\bar{\theta}^0} \Delta\theta_2^0 + \cdots + \left.\frac{\partial g_k}{\partial \theta_m}\right|_{\bar{\theta}^0} \Delta\theta_m^0 = 0 \qquad k = 1, 2, \cdots, m$$

Sistema de *m* ecuaciones con *m* incógnitas que son los desplazamientos: $\Delta\theta_k^0$

6. Modelos no lineales en los parámetros

Los coeficientes del sistema de ecuaciones son las derivadas segundas de la función a minimizar

$$G_{lk} \equiv \frac{\partial g_k}{\partial \theta_l} = \frac{\partial^2 \chi^2}{\partial \theta_l \partial \theta_k} = \sum_{i=1}^{n} \left(-\frac{2}{\sigma_i^2} \right) \left[-\frac{\partial f_i}{\partial \theta_l} \frac{\partial f_i}{\partial \theta_k} + \left(y_i - f_i \right) \frac{\partial^2 f_i}{\partial \theta_l \partial \theta_k} \right]$$

Cuyos valores se calculan numérica o analíticamente y todas las cantidades se calculan en $\overline{\theta} = \overline{\theta}^0$

En notación vectorial, el sistema de ecuaciones se escribe como:

$$G \Delta \overline{\theta}^0 = -\overline{g}\left(\overline{\theta}^0\right) \quad \Rightarrow \quad \boxed{\Delta \overline{\theta}^0 = -G^{-1}\left(\overline{\theta}^0\right) \overline{g}\left(\overline{\theta}^0\right)}$$

La nueva solución $\overline{\theta}^1 = \overline{\theta}^0 + \Delta \overline{\theta}^0$ es el punto de partida de la siguiente iteración

En el caso lineal:

$$f\left(x_i; \overline{\theta}\right) = \sum_{j=1}^{m} a_j\left(x_i\right) \theta_j \quad \Rightarrow \quad \frac{\partial f}{\partial \theta_k} = a_k\left(x\right) \quad \Rightarrow \quad \frac{\partial^2 f}{\partial \theta_l \partial \theta_k} = 0$$

Se recuperan las ecuaciones normales:

6. Modelos no lineales en los parámetros

Desarrollando en serie de potencias en torno a $\bar{\theta}^0$

$$\chi^2(\bar{\theta}) = \chi^2(\bar{\theta}^0) + \left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\bar{\theta}^0} (\bar{\theta} - \bar{\theta}^0) + \frac{1}{2}(\bar{\theta} - \bar{\theta}^0) \left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}^0} (\bar{\theta} - \bar{\theta}^0) + \cdots$$

En caso de dependencia lineal en los parámetros, los términos de tercer orden y superiores se anulan y la función resulta ser una parábola

Repitiendo el desarrollo pero en torno al mínimo $\bar{\theta}_m$ y despreciando los términos de orden 3 en adelante:

$$\chi^2(\bar{\theta}) = \chi^2(\bar{\theta}_m) + (\bar{\theta} - \bar{\theta}_m)^T \left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\bar{\theta}_m} + \frac{1}{2}(\bar{\theta} - \bar{\theta}_m)^T \left.\frac{\partial^2 \chi^2}{\partial \bar{\theta}^2}\right|_{\bar{\theta}=\bar{\theta}_m} (\bar{\theta} - \bar{\theta}_m)$$

Teniendo en cuenta que:

$$\left.\begin{aligned} &\chi^2(\bar{\theta}_m) = \chi^2_{\min} \\ &\left.\frac{\partial \chi^2}{\partial \bar{\theta}}\right|_{\bar{\theta}=\bar{\theta}_m} = 0 \\ &(V^{-1})_{ij} = \frac{1}{2}\left[\frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j}\right]_{\bar{\theta}=\bar{\theta}_m} \end{aligned}\right\} \Rightarrow \boxed{\chi^2(\bar{\theta}) = \chi^2_{\min} + (\bar{\theta} - \bar{\theta}_m)^T V^{-1}(\hat{\bar{\theta}})(\bar{\theta} - \bar{\theta}_m)}$$

Ecuación base para el cálculo de intervalos de confianza mediante el método ***LS*** (similar al método ***ML***):

6. Modelos no lineales en los parámetros

$$\chi^2(\bar{\theta}) = \chi^2_{\min} + (\bar{\theta} - \bar{\theta}_m)^T V^{-1}(\hat{\bar{\theta}})(\bar{\theta} - \bar{\theta}_m)$$

En caso de un solo parámetro:

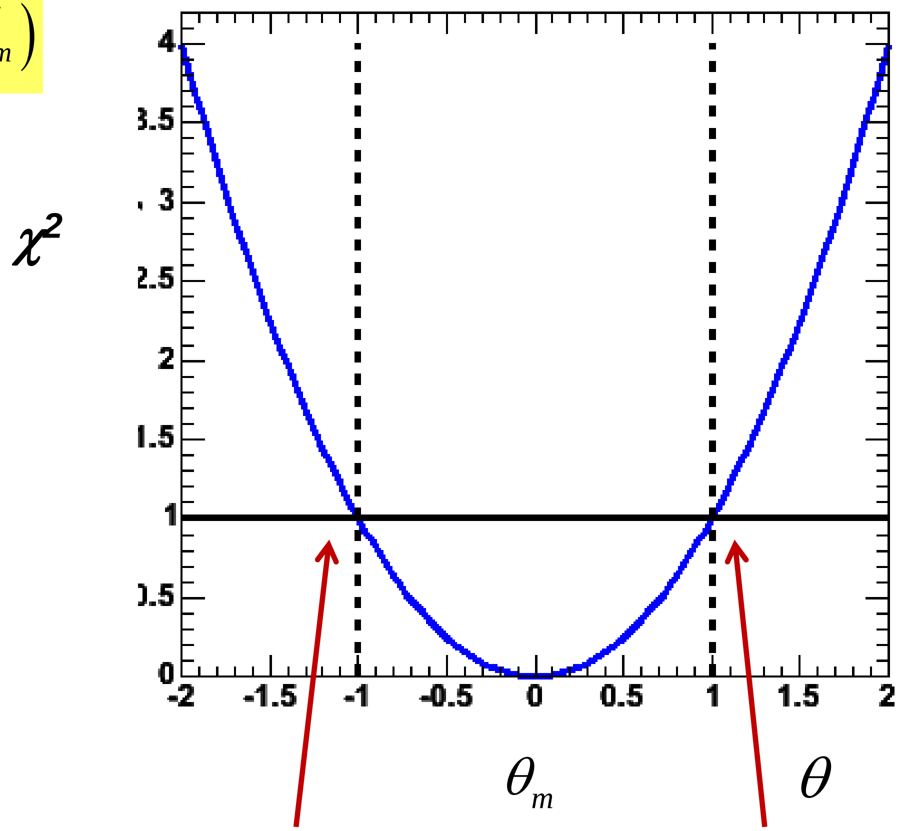
$$\chi^2(\theta) = \chi^2_{\min} + \frac{(\theta_m - \theta)^2}{\sigma^2}$$

$$\chi^2(\theta_m \pm \sigma) = \chi^2_{\min} + 1$$

$$\chi^2(\theta_m \pm 2\sigma) = \chi^2_{\min} + 4$$

$$\vdots \qquad \vdots$$

$$\chi^2(\theta_m \pm a\sigma) = \chi^2_{\min} + a^2$$



El error en el parámetro viene dado por la variación de χ^2 en una unidad

7. Bondad de los ajustes

Si las n medidas $y_1, y_2, \ldots, y_n$ se distribuyen normalmente, el valor de χ^2 obtenido se puede utilizar como un test de la hipótesis modelo:

$$\left(\frac{y_i - f\left(x_i;\theta\right)}{\sigma_i}\right) \Rightarrow$$

Representa una medida de la desviación entre la medida y_i y el modelo $f\left(x_i;\theta\right)$

χ^2 es una medida del acuerdo entre la hipótesis, **Teoría**, y los datos observados, **Experimento.**

Se puede demostrar que si:

1. Las valores de $y_1, y_2, \ldots, y_n$ son variables aleatorias distribuidas normalmente con varianzas conocidas σ_i^2 (o distribuidas multi-normalmente con matriz de covarianza V)
2. La hipótesis $f\left(x;\theta_1,\theta_2,\ldots,\theta_m\right)$ es lineal en los parámetros θ_i
3. La hipótesis es correcta.

$$\Downarrow$$

El mínimo de la función de χ^2 se distribuye de acuerdo a una distribución de $f\left(\chi^2;\nu\right)$ con $\nu = n - m$ grados de libertad

$$\chi^2\left(\bar{\theta}\right) = \sum_{i=1}^{n} \frac{\left(y_i - f\left(x_i,\bar{\theta}\right)\right)^2}{\sigma_i^2} \qquad \chi^2\left(\bar{\theta}\right) = \sum_{i,j=1}^{n} \left(y_i - f\left(x_i,\bar{\theta}\right)\right)\left(V^{-1}\right)_{ij}\left(y_j - f\left(x_j,\bar{\theta}\right)\right)$$

7. Bondad de los ajustes

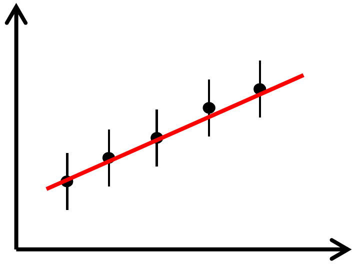
Como el valor esperado de la distribución de $f\left(\chi^2;\nu\right)$ es el número de grados de libertad se suele utilizar la variable χ^2 dividida por el número de grados de libertad χ^2/ν como una medida de la bondad del ajuste:

$$E\left[\chi^2\right]/\nu = 1$$

Se pueden dar los siguientes casos:

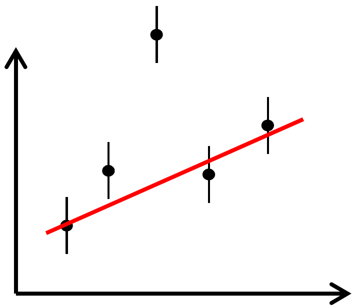
1. Si $\chi^2/\nu \simeq 1$, el resultado es el esperado. Todo es correcto.
2. Si $\chi^2/\nu \ll 1$, el ajuste es mejor de lo esperado para esos errores.

   Probablemente los errores estén *sobre-estimados* o correlacionados.

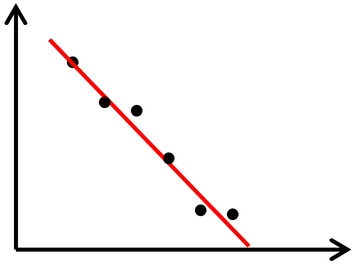


Errores sobre-estimados

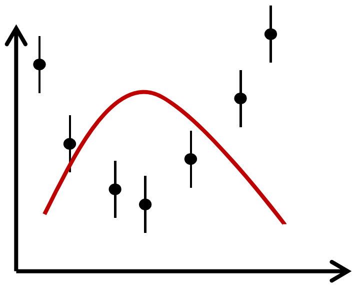
1. Si $\chi^2/\nu \gg 1$, hay razones para dudar de que la hipótesis sea correcta.
   - Medidas erróneas
   - Errores subestimados
   - Modelo teórico incorrecto
   - Mala suerte



Medidas erróneas



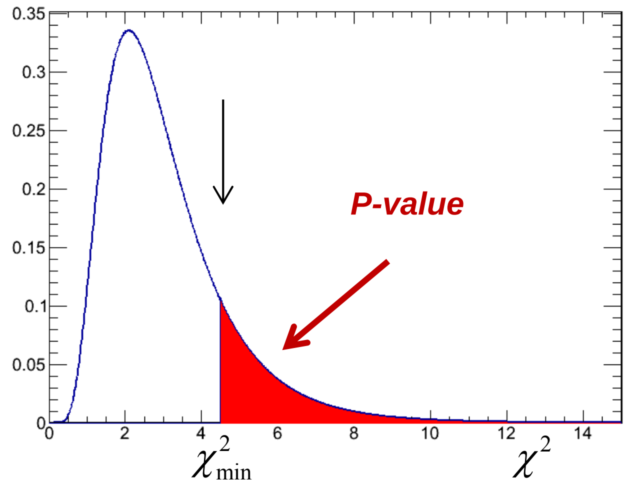
Errores sub-estimados



Modelo incorrecto

7. Bondad de los ajustes

Se utiliza también como medidor de la bondad del ajuste el **P-value**, o la probabilidad de que la hipótesis de lugar a un valor de $\chi^2_{\min}$ igual o mayor que el obtenido



$$P\left(\chi^2_{\min}\right) = \int_{\chi^2_{\min}}^{\infty} f\left(\chi^2, \nu\right) d\chi^2$$

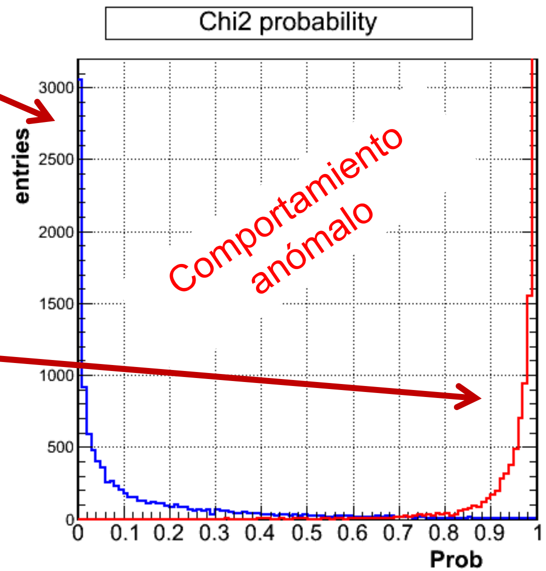
ROOT function
TMath::Prob(chi2,ndf)

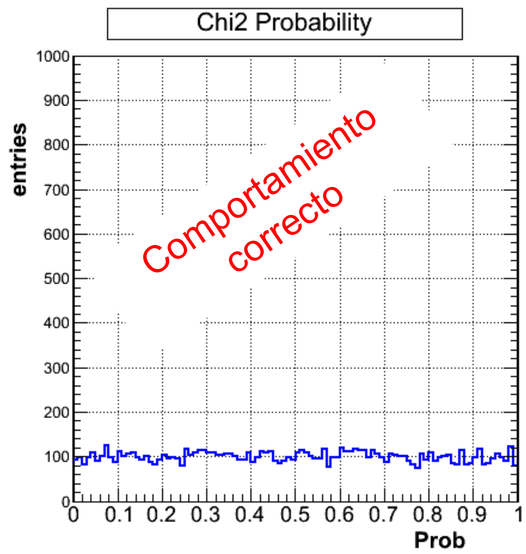
Integral acumulativa es uniforme → $P\left(\chi^2_{\min}\right)$ Se distribuye uniformemente en [0,1]

Si el modelo y las fluctuaciones (errores) son correctas la distribución Prob($\chi^2_{\min}$,ndf) debe ser plana:

Desviaciones del comportamiento plano indican que algo va mal:

- Si $\chi^2/\nu \gg 1 \rightarrow P\left(\chi^2_{\min}\right)\downarrow$
  - *Sucesos de fondo*
  - *Mal ajuste*
  - *Errores muy pequeños*
- *Si* $\chi^2/\nu \ll 1 \rightarrow P\left(\chi^2_{\min}\right)\uparrow$
  - Errores demasiado grandes





7. Bondad de los ajustes

Como el valor esperado de χ^2 es igual al número de grados de libertad, el cociente se considera también una medida del acuerdo entre datos e hipótesis pero…

$$\chi^2/n_d \simeq 1$$

¡¡***cuidado***!!

No conviene fiarse solo de:

$$\chi^2/n_d$$

Comprobar siempre el P-value

Ejemplo

$\chi^2 = 15 \;\; ; \;\; n_d = 10$

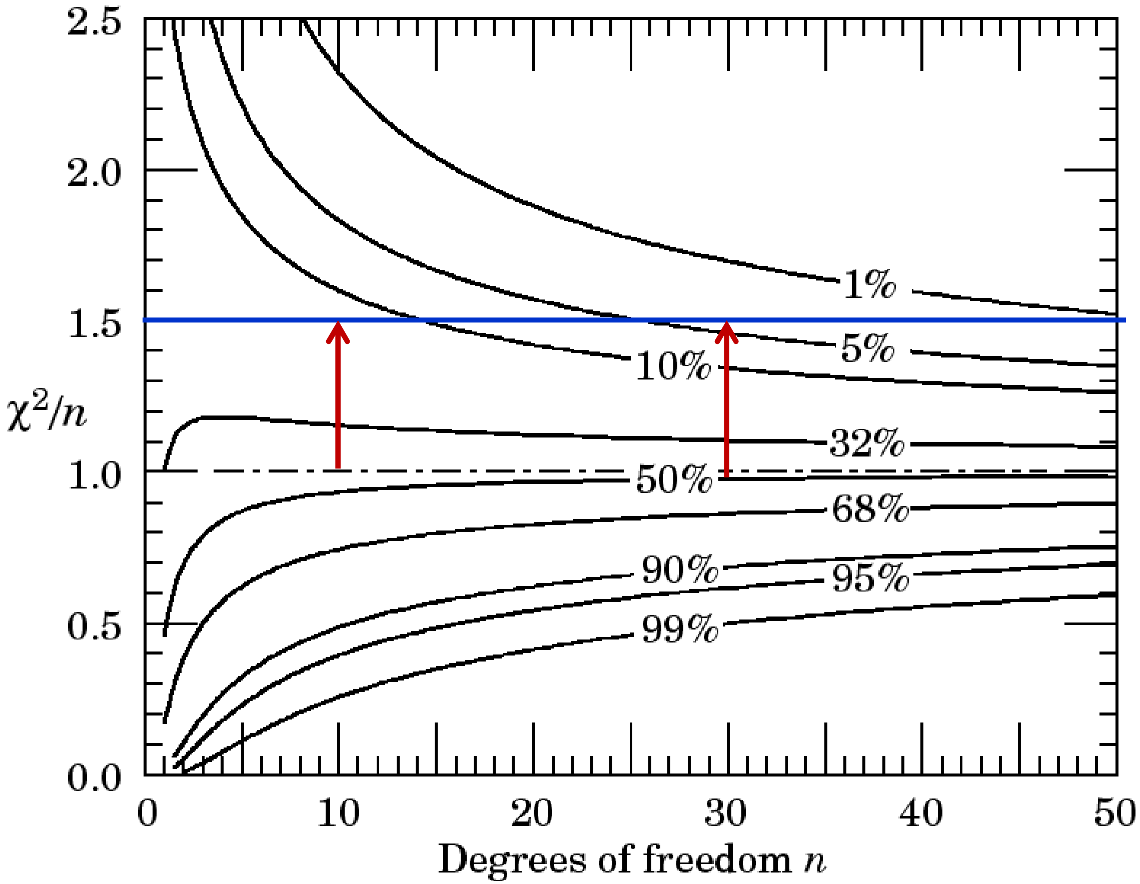
$\rightarrow P = 0.13$

Ajuste aceptable

$\chi^2 = 45 \;\; ; \;\; n_d = 30$

$\rightarrow P = 0.04$!!

Ajuste empieza a ser cuestionable



8. Residuos y Pulls

Una forma sencilla de comprobar que el ajuste es bueno es mirar la distribución de residuos y pulls

Residuos se definen como las diferencias o desviaciones entre las medidas originales y los valores ajustados:

$$\varepsilon_i = y_i - \mu_i$$

Si las precisiones de los puntos son diferentes, mejor dividir por el error

Pulls (Stretch functions) se definen como los residuos divididos por la precisión

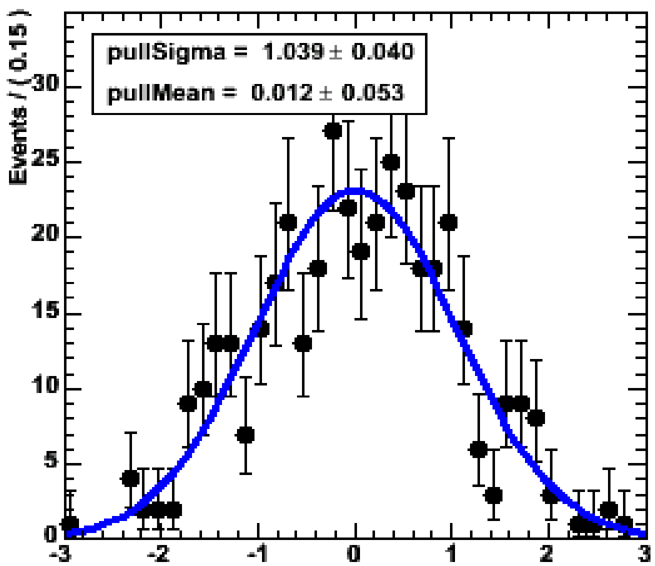
$$z_i = \frac{\varepsilon_i}{\sigma(\varepsilon_i)}$$

Si las medidas son independientes y el problema considerado es suficientemente lineal, el comportamiento esperado es el de una distribución normal estándar $N(0,1)$

Propiedades de los “pulls”

- El valor medio debe ser cero si no hay sesgo (*bias*)
- La anchura debe ser 1 si el error es correcto

Ejemplo sin sesgo y con error correcto dentro de la precisión estadística



9. Mínimos cuadrados con datos clasificados

En ocasiones la función modelo $f(x;\theta)$ es una distribución o pdf. Supongamos n observaciones de la variable aleatoria x a partir de las cuales construimos un histograma con N bines:

- Las medidas son ahora el contenido de los bines: $y_i = n_i \quad i = 1, 2, \ldots, N$
- La función $f(x;\overline{\theta})$ una pdf hipotética para la cual queremos calcular sus parámetros: $\overline{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$

El número de entradas predichas en cada *bin* vendrá dado por:

$$f_i(\overline{\theta}) = n \int_{x_i^{\min}}^{x_i^{\max}} f(x;\overline{\theta})\,dx = np_i(\overline{\theta})$$

$p_i(\overline{\theta})$ Probabilidad de obtener una entrada en el bin i-ésimo

$x_i^{\min}, x_i^{\max}$ Límites de *bin* *i*-ésimo

Y los parámetros se obtienen minimizando la cantidad:

$$\chi^2(\overline{\theta}) = \sum_{i=1}^{N} \frac{\left(y_i - f(x_i,\overline{\theta})\right)^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{\left(n_i - np_i(\overline{\theta})\right)^2}{\sigma_i^2}$$

Donde:

- n_i son los contenidos observados de cada bin
- σ_i^2 son las varianzas del número de entradas de cada bin

Dos planteamientos equivalentes:

- Los valores observados se distribuyen multi-nomialmente ligados por: ➡ $\sum_{i=1}^{N} n_i = \sum_{i=1}^{N} np_i = n$
- Los valores observados se distribuyen según *N* variables de Poisson independientes

9. Mínimos cuadrados con datos clasificados

Valores distribuidos multi-nomialmente.

En este caso la matriz de covarianza viene dada por:

$$V = \begin{pmatrix} np_1(1-p_1) & -np_1 p_2 & \cdots & -np_1 p_N \\ -np_2 p_1 & np_2(1-p_2) & \cdots & -np_2 p_N \\ \vdots & \vdots & & \vdots \\ -np_N p_1 & -np_N p_2 & \cdots & np_N(1-p_N) \end{pmatrix}$$

Matriz singular debido a $\sum_{i=1}^{N} n_i = \sum_{i=1}^{N} np_i = n$

Método LS no es aplicable $|V| = 0$

Solución:

Eliminamos una fila y columna correspondiente a una de las clases (la última), para que la nueva matriz V^* no sea singular:

Se puede comprobar fácilmente que la matriz inversa de $V^{\prime *}$ es:

$$V^{*-1} = \frac{1}{n}\begin{pmatrix} p_1^{-1} + p_N^{-1} & p_N^{-1} & \cdots & p_N^{-1} \\ p_N^{-1} & p_2^{-1} + p_N^{-1} & \cdots & p_N^{-1} \\ \vdots & \vdots & & \vdots \\ p_N^{-1} & p_N^{-1} & \cdots & p_{N-1}^{-1} + p_N^{-1} \end{pmatrix}$$

Minimizamos entonces la cantidad:

$$\chi^2(\bar{\theta}) = (\bar{y} - n\bar{p})^T V^{*-1} (\bar{y} - n\bar{p}) = \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} (n_i - np_i) V_{ij}^{*-1} (n_j - np_j)$$

9. Mínimos cuadrados con datos clasificados

Tras unas cuantas operaciones:

$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N-1}\sum_{j=1}^{N-1}(n_i - np_i)V_{ij}^{*-1}(n_j - np_j) = \frac{1}{n}\left\{\sum_{i=1}^{N-1}\frac{(n_i - np_i)^2}{p_i} + \frac{1}{p_N}\sum_{i=1}^{N-1}\sum_{j=1}^{N-1}(n_i - np_i)(n_j - np_j)\right\} =$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{N-1}\frac{(n_i - np_i)^2}{p_i} + \frac{1}{p_N}\left[\sum_{i=1}^{N-1}(n_i - np_i)\right]^2\right\} = \frac{1}{n}\left\{\sum_{i=1}^{N-1}\frac{(n_i - np_i)^2}{p_i} + \frac{1}{p_N}\left[(n - n_N) - (n - np_N)\right]^2\right\} =$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{N-1}\frac{(n_i - np_i)^2}{p_i} + \frac{1}{p_N}(n_N - np_N)^2\right\} =$$

$$= \sum_{i=1}^{N}\frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{N}\left(\frac{n_i - np_i}{\sqrt{np_i}}\right)^2$$

$$\sum_{i=1}^{N} n_i = \sum_{i=1}^{N-1} n_i + n_N = n$$

$$\sum_{i=1}^{N} np_i = \sum_{i=1}^{N-1} np_i + np_N = n$$

Equivalente a n variables poissonianas independientes

Ambos planteamientos son equivalentes

Finalmente los estimadores **LS** se obtienen minimizando la cantidad

$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N}\frac{\left(n_i - np_i(\bar{\theta})\right)^2}{np_i(\bar{\theta})}$$

9. Mínimos cuadrados con datos clasificados

$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N} \frac{\left(n_i - np_i(\bar{\theta})\right)^2}{np_i(\bar{\theta})} = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{f_i}$$

Igualando las derivadas a cero:

$$\frac{\partial \chi^2}{\partial \theta_k} = -2\sum_{i=1}^{N}\left(\frac{n_i - f_i}{f_i} + \frac{1}{2}\left(\frac{n_i - f_i}{f_i}\right)^2\right)\frac{\partial f_i}{\partial \theta_k} = 0 \qquad k = 1,2,\cdots,m$$

Difícil solución analítica, se suele resolver numéricamente

Para valores de n grandes, el segundo termino se puede despreciar

$$\frac{\partial \chi^2}{\partial \theta_k} = -2\sum_{i=1}^{N}\left(\frac{n_i - f_i}{f_i}\right)\frac{\partial f_i}{\partial \theta_k} = 0 \qquad k = 1,2,\cdots,m$$

Mínimos cuadrados simplificado

Aproximar la varianza del número de entradas por el número de entradas

$$\chi^2(\bar{\theta}) = \sum_{i=1}^{N} \frac{(n_i - f_i)^2}{f_i} = \sum_{i=1}^{N} \frac{(n_i - np_i)^2}{n_i}$$

Más sencillo de manejar pero:

- Errores mal calculados
- Parámetros sensibles a fluctuaciones estadísticas

Para valores de n grandes, ambas formulaciones coinciden

9. Mínimos cuadrados con datos clasificados

En el límite asintótico de grandes números y siempre que $E[n_i] = np_i$ los valores de χ^2 se distribuyen como una distribución de:

$\chi^2(N-1)$ ➡ Se pierde un grado de libertad debido a la condición de normalización $\sum_{i=1}^{N} n_i = \sum_{i=1}^{N} np_i = n$

Una vez minimizados, los valores de valores de $\chi^2_{\min}$ se distribuyen como $\chi^2(N-1-m)$ siempre y cuando el número de sucesos sea suficientemente grande como para suponer que:

$\left(\frac{n_i - f_i}{\sqrt{f_i}}\right)$ y $\left(\frac{n_i - f_i}{\sqrt{n_i}}\right)$ se comportan gaussianamente

¡Cuidado! Hay que tener en cuenta la anchura del *bin* w_i

$$f_i = nw_i f(x_i;\theta)$$

Criterios para la elección del tamaño del bin

- Suele haber dos criterios:
  - *a)* *Bines* igualmente espaciados
  - *b)* *Bines* de igual probabilidad
- No se puede elegir el tamaño del *bin* para hacer que $\chi^2_{\min}$ sea lo más pequeño posible
- El número de entradas de cada *bin* ha de ser suficientemente grande para que las variables $(n_i - f_i)/\sqrt{f_i}$ y $(n_i - f_i)/\sqrt{n_i}$ sean gaussianas.
- En general se acostumbra a requerir al menos un mínimo de 5 entradas en cada *bin*

10. Resumen propiedades

Comentarios sobre el método LS:

- Si la función f es lineal en los parámetros, hay solución analítica.
- En el caso lineal, los estimadores LS son *unbiased* y tienen varianza mínima.
- El valor de $\chi^2_{\min}$ proporciona información sobre la calidad del ajuste.

Ventajas

- Fácil de usar (implementar), incluso para muestras de datos muy grandes.
- Proporciona información sobre la calidad del ajuste (*Goodness-of-fit*).
- Método general para comparar dos distribuciones, datos vs. Teoría, etc.

Desventajas

- En caso de *binning* siempre hay una pérdida de información.
- Hay que ser cuidadosos con *bines* de pocas entradas.
  - Al menos >= 5 entradas, *bines* no vacíos.
  - De lo contrario, errores no gaussianos, distribución no es de χ^2
- En caso de correlaciones hay que utilizar la matriz de covarianzas.