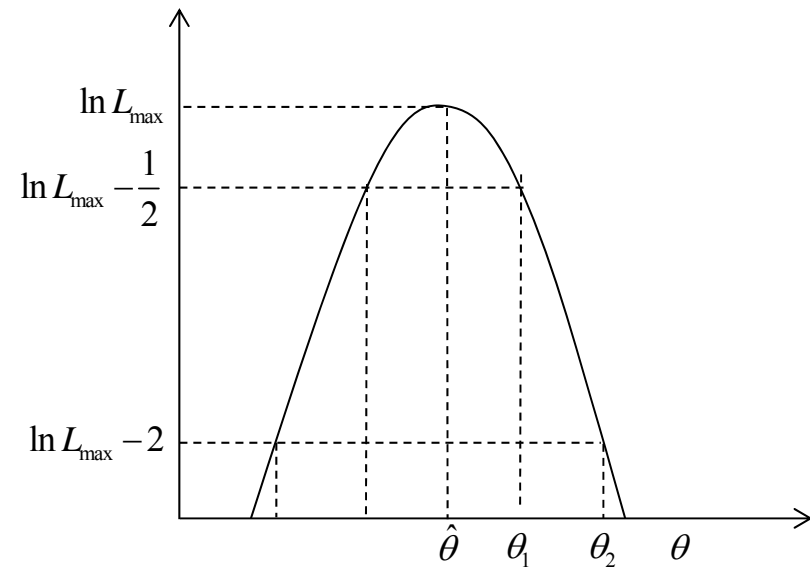


Tema 8 Método de máxima verosimilitud

1. Principio de máxima verosimilitud.
2. Ejemplos de estimadores ML
3. Cambio de parámetros
4. Propiedades asintóticas
5. Varianza de los estimadores ML.
6. Método de máxima verosimilitud extendido
7. Resumen propiedades estimadores ML



1. Principio de máxima verosimilitud

El **método de máxima verosimilitud** es un método de estimación de parámetros a partir de una muestra finita de datos.

Supongamos que tenemos una variable aleatoria x distribuida de acuerdo a una pdf $f(x, \theta)$ de la que conocemos su forma funcional pero no todos sus parámetros $\bar{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$

Supongamos que la medida se ha repetido un número n de veces obteniéndose los valores x_1, x_2, \dots, x_n

Probabilidad de que la primera medida esté entre $[x_1, x_1 + dx_1]$ será $f(x_1, \theta) dx_1$

Probabilidad de que la segunda medida esté entre $[x_2, x_2 + dx_2]$ será $f(x_2, \theta) dx_2$

.....

Probabilidad de que la última medida esté entre $[x_n, x_n + dx_n]$ será $f(x_n, \theta) dx_n$

$$\{\text{Probabilidad de que } x_i \text{ esté entre } [x_i, x_i + dx_i] \text{ para todo } i\} = \prod_{i=1}^n f(x_i, \theta) dx_i$$

Si las hipótesis son correctas, pdf y parámetros, uno espera una alta probabilidad para los datos que se han obtenido y por tanto un valor máximo para la función:

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta)$$

denominada **función de verosimilitud o likelihood**, que no es más que la función pdf conjunta para los valores observados x_i

1. Principio de máxima verosimilitud

Función de likelihood

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta)$$

En **Teoría de probabilidad**, y conocido $\bar{\theta}$, expresa la probabilidad de obtener, en un experimento idéntico al nuestro, el mismo resultado x_i que hemos observado.

En **Estadística**, es al revés, tenemos los datos, x_i , creemos saber la pdf: $f(x_i, \bar{\theta})$, y queremos estimar los parámetros $\bar{\theta}$

Se definen los **estimadores de máxima verosimilitud (ML)** $\hat{\theta}_i$ aquellos valores de θ_i que hacen que nuestros resultados sean los más probables de todos, es decir aquellos que maximizan la función de verosimilitud.

Principio de máxima verosimilitud La mejor estimación $\hat{\theta}$ de un parámetro θ es aquel valor que maximiza la función de verosimilitud:

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, \dots, k$$

A menudo se suele trabajar con el logaritmo de la función de verosimilitud:

$$\ell = \ln L \quad \longrightarrow \quad \ell = \sum_{i=1}^n \ell_i = \sum_{i=1}^n \ln f(x_i, \theta)$$

Como $L > 0$, tanto L como ℓ tienen los mismos extremos

$$\frac{\partial \ell}{\partial \theta_i} = 0, \quad i = 1, \dots, k$$

2. Ejemplo de estimadores ML

Ejemplo 1 Distribución exponencial. Estimación de la vida media. $f(x, \tau) = \frac{1}{\tau} e^{-x/\tau}$

Consideremos n medidas de una distribución exponencial, queremos estimar el valor de la vida media. La función de verosimilitud vendrá dada por:

$$\ell = \ln L(\tau) = \ln \prod_{i=1}^n f(x_i; \tau) = \sum_{i=1}^n \ln f(x_i; \tau) = \sum_{i=1}^n \ln \left(\frac{1}{\tau} e^{-x_i/\tau} \right) = \sum_{i=1}^n -\ln \tau - \frac{x_i}{\tau}$$

Derivando respecto al parámetro:

$$\frac{\partial \ell}{\partial \tau} = \sum_{i=1}^n \left(-\frac{1}{\tau} + \frac{x_i}{\tau^2} \right) = \frac{1}{\tau} \sum_{i=1}^n \left(-1 + \frac{x_i}{\tau} \right) = 0 \quad \Rightarrow \quad \frac{1}{\tau} \sum_{i=1}^n x_i = n \quad \Rightarrow \quad \hat{\tau} = \frac{1}{n} \sum_{i=1}^n x_i$$

Su valor esperado:

$$\begin{aligned} E[\hat{\tau}(x_1, x_2, \dots, x_n)] &= \int \cdots \int \hat{\tau}(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n = \\ &= \int \cdots \int \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \frac{1}{\tau} e^{-x_1/\tau} \cdots \frac{1}{\tau} e^{-x_n/\tau} dx_1 dx_2 \cdots dx_n = \\ &= \frac{1}{n} \sum_{i=1}^n \left(\int x_i \frac{1}{\tau} e^{-x_i/\tau} dx_i \prod_{j \neq i} \int \frac{1}{\tau} e^{-x_j/\tau} dx_j \right) = \frac{1}{n} \sum_{i=1}^n \tau = \tau \end{aligned}$$

$\hat{\tau}$ es un estimador no sesgado (*unbiased*) de τ

$$\Rightarrow E[\hat{\tau}] = \tau$$

2. Ejemplo de estimadores ML

Ejemplo 2 Distribución normal.

Consideremos n medidas independientes x_i cada una de ellas distribuida normalmente

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \quad \longrightarrow \quad \ell = \ln L = \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \ln \sigma_i - \frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right]$$

Supongamos que todas tienen el mismo $\mu = \mu_i$ pero diferente σ_i :

$$\frac{\partial \ell}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma_i^2} = \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{\mu}{\sigma_i^2} = 0$$



$$\hat{\mu} = \frac{\sum_{i=1}^n x_i / \sigma_i^2}{\sum_{i=1}^n 1 / \sigma_i^2}$$

Media ponderada:
Cada x_i está pesado por:

$$w = 1 / \sigma_i^2$$

Su valor esperado:

$$E[\hat{\mu}] = E\left[\frac{\sum (x_i / \sigma_i^2)}{\sum (1 / \sigma_i^2)}\right] = \frac{\sum (E[x_i] / \sigma_i^2)}{\sum (1 / \sigma_i^2)} = \frac{\sum (\mu / \sigma_i^2)}{\sum (1 / \sigma_i^2)} = \frac{\mu \sum (1 / \sigma_i^2)}{\sum (1 / \sigma_i^2)} = \mu$$



$\hat{\mu}$ es un estimador no sesgado (*unbiased*) de μ

2. Ejemplo de estimadores ML

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i / \sigma_i^2}{\sum_{i=1}^n 1 / \sigma_i^2}$$

Ejemplo 2 Distribución normal. Media ponderada

Su varianza:

$$V[\hat{\mu}] = E[\hat{\mu}^2] - (E[\hat{\mu}])^2 = \frac{E\left[\left(\sum x_i / \sigma_i^2\right)^2\right]}{\left[\sum (1/\sigma_i^2)\right]^2} - \mu^2 = \frac{E\left[\left(\sum x_i x_j / \sigma_i^2 \sigma_j^2\right)\right]}{\left[\sum (1/\sigma_i^2)\right]^2} - \mu^2 =$$

$E[x_i x_j] = \begin{cases} E[x_i]E[x_j] = \mu_i \mu_j = \mu^2 & \text{si } i \neq j \\ E[x_i^2] = \sigma_i^2 + \mu^2 & \text{si } i = j \end{cases}$

$$= \left(\frac{1}{\sum (1/\sigma_i^2)}\right)^2 \left(\sum_i \frac{\sigma_i^2 + \mu^2}{\sigma_i^4} + \sum_{i \neq j} \frac{\mu^2}{\sigma_i^2 \sigma_j^2}\right) - \mu^2 = \left(\frac{1}{\sum (1/\sigma_i^2)}\right)^2 \left(\sum_i \frac{1}{\sigma_i^2} + \mu^2 \sum_i \frac{1}{\sigma_i^4} + \mu^2 \sum_{i \neq j} \frac{1}{\sigma_i^2 \sigma_j^2}\right) - \mu^2 =$$

$$= \frac{1}{\sum (1/\sigma_i^2)} + \mu^2 \left(\frac{\sum_i \frac{1}{\sigma_i^4} + \sum_{i \neq j} \frac{1}{\sigma_i^2 \sigma_j^2}}{\left(\sum (1/\sigma_i^2)\right)^2}\right) - \mu^2 = \frac{1}{\sum (1/\sigma_i^2)}$$

Recordemos que para un estimador no sesgado, la varianza mínima viene dada por (desigualdad de Raó-Cramér):

$$V[\hat{\mu}] = \sigma^2(\hat{\mu}) = \frac{1}{I(\mu)} = \frac{1}{-E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]} = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$$



Se trata de un estimador eficiente puesto que tiene la mínima varianza posible:

$$-E\left[\frac{\partial^2 \ell}{\partial \mu^2}\right] = -E\left[\frac{\partial}{\partial \mu} \left(\sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{\mu}{\sigma_i^2}\right)\right] = -E\left[-\sum_{i=1}^n \frac{1}{\sigma_i^2}\right] = \sum_{i=1}^n \frac{1}{\sigma_i^2}$$

2. Ejemplo de estimadores ML

Ejemplo 3 Distribución normal. Media y varianza estimadas simultáneamente

Consideremos n medidas independientes x_i cada una de ellas distribuida normalmente

Supongamos que todas tienen la misma media $\mu = \mu_i$ y la misma varianza $\sigma_i = \sigma$

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right] \quad \longrightarrow \quad \ell = \ln L = \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \ln \sigma - \frac{(x_i - \mu)^2}{2\sigma^2} \right]$$

$$\left. \frac{\partial \ell}{\partial \mu} \right|_{\hat{\mu}\hat{\sigma}} = \sum_{i=1}^n \frac{x_i - \hat{\mu}}{\hat{\sigma}^2} = 0 \quad \longrightarrow \quad \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\left. \frac{\partial \ell}{\partial \sigma} \right|_{\hat{\mu}\hat{\sigma}} = \sum_{i=1}^n \left(-\frac{1}{\hat{\sigma}} + \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^3} \right) = 0 \quad \longrightarrow \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Estimador sesgado (*biased*)
de la varianza !!!

Conclusión

- Los estimadores **ML** no son, en general, eficientes y no sesgados (*unbiased*).
- Sin embargo, se puede demostrar que, un estimador **ML** es eficiente y no sesgado si existe un estimador eficiente y no sesgado.

3. Cambio de parámetros

Distinción entre cambio de variable y cambio de parámetro en la función de likelihood:

Cambio de variable

$L(x, \theta)$ es una pdf para la variable aleatoria x . En un cambio de variable se conservan las probabilidades:

$$x \rightarrow y(x) \longrightarrow L(x; \theta) \rightarrow L'(y; \theta) \longrightarrow L(x; \theta) dx = L'(y; \theta) dy \longrightarrow L'(y; \theta) = L(x; \theta) |J|$$

Cambio de parámetro

La conservación de las probabilidades implica:

$$\theta \rightarrow g(\theta) \longrightarrow L(x; \theta) dx \rightarrow L'(x; g) dx \longrightarrow L(x; \theta) = L'(x; g)$$

L no cambia bajo transformaciones de parámetros

$$\left. \begin{array}{l} \frac{\partial L}{\partial g} = \frac{\partial L}{\partial \theta} \frac{\partial \theta}{\partial g} \\ \frac{\partial L}{\partial \theta} = 0 \end{array} \right\} \longrightarrow \frac{\partial L}{\partial g} = 0 \longrightarrow \hat{g}(\theta) = g(\hat{\theta})$$

Invariancia

Si $\hat{\theta}$ es el estimador de θ , entonces $g(\hat{\theta})$ es el estimador de $g(\theta)$

Ejemplo

Distribución exponencial $\hat{\lambda} = \frac{1}{\hat{\tau}} = \frac{n}{\sum_i x_i}$

Sin embargo, un estimador *unbiased* no tiene porque seguir siéndolo:

$$E[\hat{\lambda}] = \lambda \frac{n}{n-1} \longrightarrow b_n(\hat{\lambda}) = E[\hat{\lambda}] - \lambda = \frac{\lambda}{n-1} \rightarrow 0$$

Invariancia entre estimadores incompatible con ausencia de bias

En los estimadores ML sesgados, el sesgo se anula en el límite asintótico

$n \rightarrow \infty$

4. Propiedades asintóticas

En el límite asintótico, para un gran número de medidas independientes, los estimadores ML son eficientes y *unbiased*. Para verlo realizamos el siguiente desarrollo en serie de Taylor de la función $\frac{\partial \ln L}{\partial \theta}$ en torno a $\hat{\theta}$:

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln L}{\partial \theta} \Big|_{\hat{\theta}} + \frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\hat{\theta}} (\theta - \hat{\theta}) + \dots$$

El primer término se anula por el Principio de máxima verosimilitud, y el resto de términos se desprecian

$$\frac{\partial \ln L}{\partial \theta} \approx \frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\hat{\theta}} (\theta - \hat{\theta}) \approx \frac{\partial^2}{\partial \theta^2} \sum_i \ln f(x; \theta) \Big|_{\hat{\theta}} (\theta - \hat{\theta}) = \sum_i \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \Big|_{\hat{\theta}} (\theta - \hat{\theta})$$

Sustituyendo la suma por el valor medio: $\frac{\partial \ln L}{\partial \theta} \approx n \frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \Big|_{\hat{\theta}} (\theta - \hat{\theta})$

Como la media muestral tiende a su valor esperado en el límite de grandes números:

$$\frac{\partial \ln L}{\partial \theta} \approx n E \left[\frac{\partial^2}{\partial \theta^2} \ln f(x; \theta) \Big|_{\hat{\theta}} \right] (\theta - \hat{\theta}) = E \left[\frac{\partial^2}{\partial \theta^2} \sum_i \ln f(x; \theta) \Big|_{\hat{\theta}} \right] (\theta - \hat{\theta}) = E \left[\frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\hat{\theta}} \right] (\theta - \hat{\theta})$$

Como la información viene dada por: $I(\theta) = -E \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right]$ \rightarrow $\frac{\partial \ln L}{\partial \theta} \approx -I(\hat{\theta})(\theta - \hat{\theta})$

Consecuencias

1. Asintóticamente

$$I(\theta) = I(\hat{\theta})$$



$$I(\theta) = -E \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right] \xrightarrow{\text{Derivando de nuevo}} -E \left[E \left[\frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\hat{\theta}} \right] \right] = E \left[I(\hat{\theta}) \right] = I(\hat{\theta})$$

4. Propiedades asintóticas

Consecuencias

$$I(\theta) = I(\hat{\theta})$$

$$\frac{\partial \ln L}{\partial \theta} \simeq -I(\hat{\theta})(\theta - \hat{\theta})$$

2. Relación lineal entre la función $\frac{\partial \ln L}{\partial \theta}$ y $\hat{\theta}$  $\hat{\theta}$ es un estimador eficiente y no sesgado

Recordemos que el límite de mínima varianza (desigualdad de Rao-Cramér) se obtuvo a partir de:

$$\rho^2 = \frac{\text{cov}\left[\hat{\theta}, \frac{\partial \ln L}{\partial \theta}\right]^2}{V\left[\frac{\partial \ln L}{\partial \theta}\right] V[\hat{\theta}]} \leq 1$$

La igualdad $\rho^2 = 1$ se da cuando la relación entre las variables es lineal

Un estimador $\hat{\theta}$ puede ser eficiente si y solo si es función lineal de la función $\frac{\partial \ln L}{\partial \theta}$

$$E\left[\frac{\partial \ln L}{\partial \theta}\right] = 0 \quad \text{---} \quad E\left[-I(\hat{\theta})(\theta - \hat{\theta})\right] = -I(\hat{\theta})E[\theta - \hat{\theta}] = 0 \quad \text{---} \quad E[\hat{\theta}] = \theta \quad \text{Estimador unbiased}$$

Los estimadores ML son, bajo ciertas condiciones, asintóticamente eficientes y unbiased:

3. La función de verosimilitud tiende asintóticamente a una gaussiana. Integrando:

$$\frac{\partial \ln L}{\partial \theta} \simeq -I(\hat{\theta})(\theta - \hat{\theta}) \quad \text{---} \quad \ln L \simeq -\frac{I(\hat{\theta})}{2}(\hat{\theta} - \theta)^2 + \ln k$$

$$k = L(\hat{\theta}) = L_{\max}$$

La función de verosimilitud es proporcional a una gaussiana de media $\hat{\theta}$ y varianza $1/I(\hat{\theta})$

$$L \simeq L_{\max} \exp\left[-\frac{1}{2}I(\hat{\theta})(\hat{\theta} - \theta)^2\right] \propto N(\theta, \hat{\theta}, I^{-1}(\hat{\theta}))$$

5. Varianza de los estimadores ML

Estimador eficiente y unbiased

Varianza de un estimador eficiente y *unbiased*

$$V[\hat{\theta}] = 1/I(\hat{\theta})$$



Desigualdad de Rao-Cramér

$$\sigma^2(\hat{\theta}) = V[\hat{\theta}] \geq \frac{\left(1 + \frac{\partial}{\partial \theta} b_n(\hat{\theta})\right)^2}{I_{\bar{x}}(\theta)}$$

En el límite asintótico los estimadores ML son tanto eficientes como unbiased.

Bajo ciertas condiciones se cumple que:

$$I_{\bar{x}}(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] = -E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]$$

Mas que calcular el valor esperado, para el límite asintótico se puede simplemente calcular la segunda derivada con los datos y los valores de los parámetros obtenidos:

$$\widehat{V}^{-1}[\hat{\theta}] = -\frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\hat{\theta}}$$



En el caso de varios parámetros:

$$\widehat{V}_{jk}^{-1}[\hat{\theta}] = I_{jk}(\bar{\theta}) = -E\left[\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k}\right]$$



$$\widehat{V}_{jk}^{-1}[\hat{\theta}] = I_{jk}(\bar{\theta}) = -\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \Big|_{\hat{\theta}}$$



Los errores estadísticos disminuyen con $1/\sqrt{n}$

$$\widehat{V}_{jk}^{-1}[\hat{\theta}] = -\sum_{i=1}^n \frac{\partial^2 \ln f(x_i; \theta)}{\partial \theta_j \partial \theta_k} \Big|_{\hat{\theta}} = n \frac{\partial^2 \ln f(x_i; \theta)}{\partial \theta_j \partial \theta_k} \Big|_{\hat{\theta}}$$



$$V \propto 1/n \Rightarrow \sigma \propto 1/\sqrt{n}$$

5. Varianza de los estimadores ML

Ejemplo **Distribución normal.**

Consideremos n medidas independientes x_i distribuidas normalmente con la misma media $\mu = \mu_i$ y diferente σ_i

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma_i}\right)^2\right] \quad \longrightarrow \quad \ln L = \sum_{i=1}^n \left[-\frac{1}{2} \ln(2\pi) - \ln \sigma_i - \frac{(x_i - \mu)^2}{2\sigma_i^2} \right]$$

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma_i^2} = \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \sum_{i=1}^n \frac{\mu}{\sigma_i^2}$$

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\sum_{i=1}^n \frac{1}{\sigma_i^2} \quad \longrightarrow \quad V^{-1}[\hat{\mu}] = -\frac{\partial^2 \ln L}{\partial \mu^2} \Big|_{\hat{\mu}} = \sum_{i=1}^n \frac{1}{\sigma_i^2} \quad \longrightarrow \quad V[\hat{\mu}] = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$$

5. Varianza de los estimadores ML

Método gráfico

Desarrollando la función $\ln L(\theta)$ en serie de Taylor en torno al estimador $\hat{\theta}$

$$\ln L(\theta) = \ln L(\hat{\theta}) + \left[\frac{\partial \ln L}{\partial \theta} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!} \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \dots$$

Como:

$$\ln L(\hat{\theta}) = \ln L_{\max}$$

$$\left[\frac{\partial \ln L}{\partial \theta} \right]_{\theta=\hat{\theta}} = 0$$



$$\ln L(\theta) = \ln L_{\max} + \frac{1}{2!} \left[\frac{\partial^2 \ln L}{\partial \theta^2} \right] (\theta - \hat{\theta})^2$$

En el límite asintótico



$$\widehat{\sigma^2}[\hat{\theta}] = - \frac{1}{\left(\frac{\partial^2 \ln L}{\partial \theta^2} \right)_{\hat{\theta}}}$$



$$\ln L(\theta) = \ln L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\widehat{\sigma_{\hat{\theta}}^2}}$$

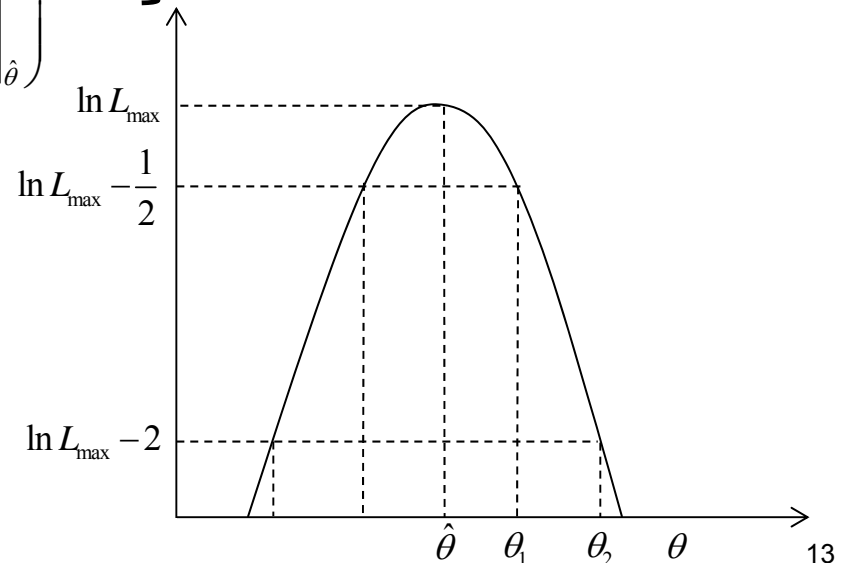
En particular:

$$\ln L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \ln L_{\max} - \frac{1}{2} \quad \sigma_{\hat{\theta}} = |\hat{\theta} - \theta_1|$$

$$\ln L(\hat{\theta} \pm 2\hat{\sigma}_{\hat{\theta}}) = \ln L_{\max} - 2 \quad 2\sigma_{\hat{\theta}} = |\hat{\theta} - \theta_2|$$

En general:

$$\ln L(\hat{\theta} \pm n\hat{\sigma}_{\hat{\theta}}) = \ln L_{\max} - \frac{1}{2} n^2$$



5. Varianza de los estimadores ML

EJEMPLO

Distribución exponencial.

$$f(x, \tau) = \frac{1}{\tau} e^{-x/\tau}$$

Muestra de 50 sucesos Monte Carlo generados con una distribución exponencial con $\tau = 1.0$

$$\ln L(\hat{\tau} \pm \hat{\sigma}_{\hat{\tau}}) = \ln L_{\max} - \frac{1}{2} \ln L$$

$$\hat{\tau} = 1.013 \quad \Delta \hat{\tau}_- = |\hat{\tau} - \tau_-| = 0.13$$

$$\tau_- = 0.884 \quad \Delta \hat{\tau}_+ = |\hat{\tau} - \tau_+| = 0.16$$

$$\tau_+ = 1.172 \quad \Delta \hat{\tau} = \frac{|\tau_+ - \tau_-|}{2} = 0.145$$

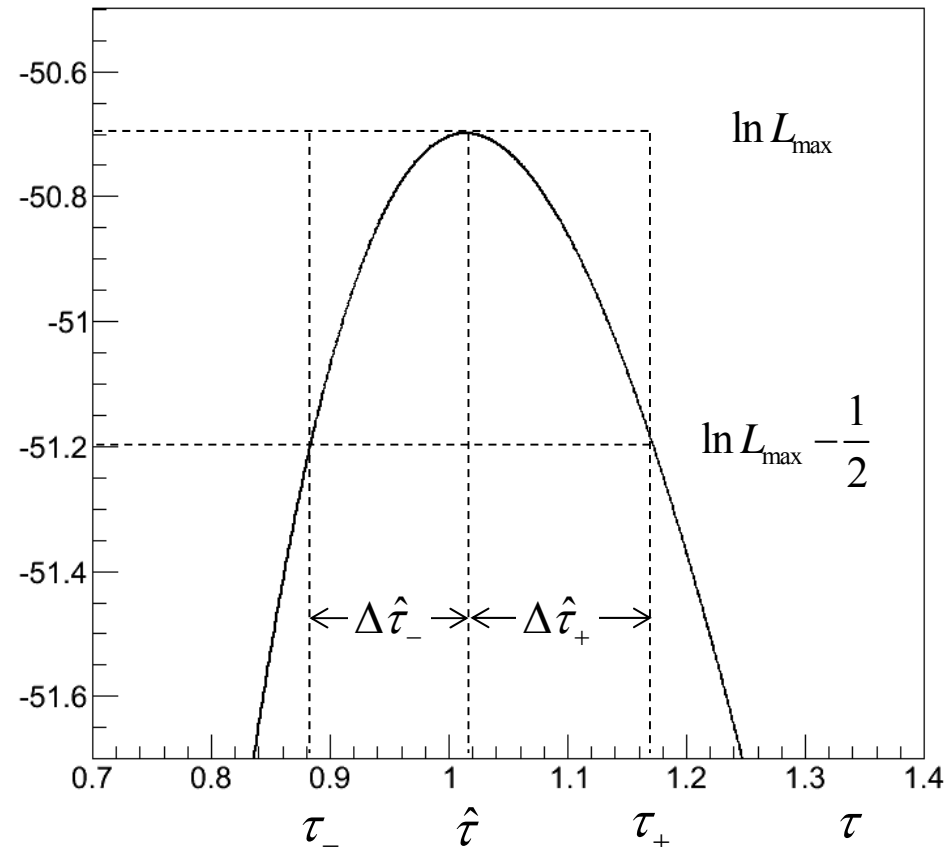
Mediante las segundas derivadas

$$\ell = \ln L(\tau) = \sum_{i=1}^n -\ln \tau - \frac{x_i}{\tau}$$

$$V[\hat{\tau}] \geq \frac{1}{-E\left[\frac{\partial^2 \ln L}{\partial \tau^2}\right]} = \frac{1}{-E\left[-\frac{n}{\tau^2} \left(1 - \frac{2\bar{x}}{\tau}\right)\right]}$$

$$= \frac{1}{-\frac{n}{\tau^2} \left(1 - \frac{2E[\bar{\tau}]}{\tau}\right)} = \frac{\tau^2}{n}$$

$$\Delta \hat{\tau} = \frac{\hat{\tau}}{\sqrt{n}} = 0.143$$



*¡Cuidado, no es un error gaussiano!
(solo en el límite asintótico)*

6. Método ML extendido

Consiste en incorporar el número de sucesos total como un parámetro más a estimar. (Propuesto por Fermi)

El resultado del experimento consiste en: n, x_1, x_2, \dots, x_n

Función de verosimilitud extendida

La función de verosimilitud sería entonces:

(Extended Maximum Likelihood, EML)

$$L(\bar{x}; \theta) = \prod_{i=1}^n f(x_i; \theta)$$



$$L_E(\bar{x}; \theta, \nu) = \frac{e^{-\nu} \nu^n}{n!} \prod_{i=1}^n f(x_i; \theta)$$

Introducimos un nuevo parámetro: ν el número medio de sucesos

Tomando logaritmos y eliminando los términos meramente aditivos:

Probabilidad de obtener n sucesos si lo esperado es ν

$$\ln L_E(\bar{x}; \theta, \nu) = \ln \frac{e^{-\nu} \nu^n}{n!} \prod_{i=1}^n f(x_i; \theta) = -\ln(n!) + n \ln \nu(\theta) - \nu(\theta) + \sum_{i=1}^n \ln f(x_i; \theta)$$

Eliminamos $n!$ pues no depende de los parámetros

$$\ln L_E(\bar{x}; \theta, \nu) = -\nu(\theta) + \sum_{i=1}^n \ln(\nu(\theta) f(x_i; \theta))$$

El método estándar **ML** utiliza la forma funcional de los datos para estimar los parámetros. La normalización (*número de sucesos*) está fijada de antemano.

El método extendido **EML** considera también la **normalización** como parámetro libre

6. Método ML extendido

Función de verosimilitud extendida

$$\ln L_E(\bar{x}; \theta, \nu) = -\nu(\theta) + \sum_{i=1}^n \ln(\nu(\theta) f(x_i; \theta))$$

Pueden darse dos situaciones:

A) El número medio de sucesos ν depende de θ : $\nu = \nu(\theta)$

Errores más pequeños para $\hat{\theta}$

B) El número medio de sucesos ν es independiente de θ :

$$\frac{\partial \ln L_E(\bar{x}; \theta, \nu)}{\partial \nu} = 0 \quad \longrightarrow \quad \hat{\nu} = n \quad \frac{\partial \ln L_E(\bar{x}; \theta, \nu)}{\partial \theta} = 0 \quad \longrightarrow \quad \hat{\theta} = \hat{\theta}_{ML}$$

Se obtienen los mismos parámetros que con el método **ML**, sin embargo los errores en los parámetros estarán sobrestimados pues el método asume que hay fluctuaciones en ν

Aún así, a veces es preferible utilizar el método de **EML** (ML extendido):

- Funciones de distribución difíciles de normalizar, integración numérica lenta.
- Interesante para funciones que son suma de varias componentes.

6. Método ML extendido

Ejemplo

Supongamos que la pdf de una variable aleatoria x consiste en la superposición de varias componentes:

$$f(x; b) = \sum_{j=1}^m b_j f_j(x) \quad \text{con} \quad \sum_{j=1}^m b_j = 1 \quad j = 1, \dots, m$$

Objetivo.- Estimar las contribuciones relativas de cada componente b_j

Mediante el método **ML** podemos usar : $b_m = 1 - \sum_{j=1}^{m-1} b_j$

Mediante el método **EML** (función de likelihood extendida):

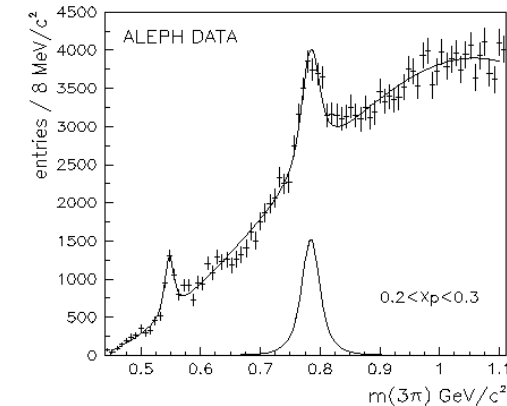
Tomando logaritmos, eliminando términos independientes de b y considerando el número total de sucesos n observados distribuido según Poisson con media ν

$$\ln L(\nu, b) = \ln \frac{e^{-\nu} \nu^n}{n!} \prod_{i=1}^n f(x_i; b) = -\nu(\theta) + \sum_{i=1}^n \ln \left(\sum_{j=1}^m \nu b_j f_j(x_i) \right)$$

Definiendo el número esperado de sucesos de tipo i como $\mu_i = \nu b_i$ la función de likelihood puede expresarse como una función de los m parámetros $\bar{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$

$$\ln L(\bar{\mu}) = -\sum_{j=1}^m \mu_j + \sum_{i=1}^n \ln \left(\sum_{j=1}^m \mu_j f_j(x_i) \right)$$

No hay ya ligaduras entre los parámetros μ_j . El número total de sucesos n es la suma de variables de Poisson independientes con media μ_j



6. Método ML extendido

Ejemplo

Consideremos dos tipos de sucesos, señal y fondo, cada uno distribuido según su propia pdf , $f_s(x)$, $f_b(x)$, de la variable aleatoria x :

$$f(x, \mu_s, \mu_b) = \frac{\mu_s}{\mu_s + \mu_b} f_s(x) + \frac{\mu_b}{\mu_s + \mu_b} f_b(x)$$

Queremos estimar μ_s y μ_b

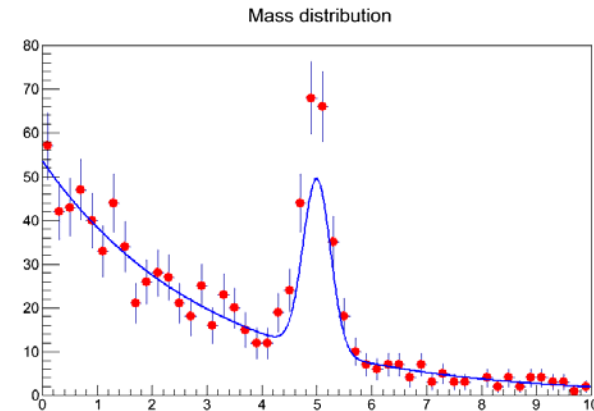
El número de sucesos señal n_s se distribuye como una poissoniana de media μ_s

El número de sucesos fondo n_b se distribuye como una poissoniana de media μ_b

$$P(n, \mu_s, \mu_b) = \frac{(\mu_s + \mu_b)^n}{n!} \exp[-(\mu_s + \mu_b)]$$

La función de likelihood extendida:

$$\ln L(\mu_s, \mu_b) = -(\mu_s + \mu_b) + \sum_{i=1}^n \ln [(\mu_s + \mu_b) f(x_i, \mu_s, \mu_b)]$$



7. Resumen propiedades

En general los estimadores ML son:

- **Consistentes**, se acercan al valor verdadero cuando $n \rightarrow \infty$ (límite asintótico)
- **Invariantes** ante transformaciones de parámetros $\hat{g}(\theta) = g(\hat{\theta})$
- **Eficientes** en el límite asintótico. Varianza mínima
- En el límite asintótico, $n \rightarrow \infty$, la función de verosimilitud tiende a un **comportamiento gaussiano**
- Para un número finito n de medidas, los estimadores ML son en general sesgados (**biased**),
- Cuando son sesgados, el sesgo tiende a cero con el número de medidas

$$\widehat{V}^{-1}[\hat{\theta}] = -\frac{\partial^2 \ln L}{\partial \theta^2} \bigg|_{\hat{\theta}}$$

El método de máxima verosimilitud:

- No proporciona el valor más probable del parámetro sino el parámetro para el cual los datos obtenidos tienen la máxima probabilidad.
- En general se necesitan métodos numéricos para encontrar el máximo de la función de verosimilitud.
- El método en sí **no** proporciona ningún **test de calidad** respecto a las hipótesis, el método funciona aun si la hipótesis es errónea.