3 Estimadores de Máxima Verosimilitud

3.1 Caso discreto

- Supongamos que tenemos una m.a.s. $(X_1, ..., X_n)$ de una v.a. discreta con distribución $P(x|\theta)$ conocida salvo por los parámetros θ
- Una vez realizada la muestra tenemos $(x_1, ..., x_n)$ y estamos interesados en encontrar un estimador $\hat{\theta}$ del parámetro θ que maximice la probabilidad de la muestra.
- Para una muestra dada, la función de probabilidad de la muestra es

$$P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n P(x_i|\theta) = l(\theta|(x_1, ..., x_n)) \equiv l(\theta)$$

y la denominaremos *función de verosimilitud*.

3.2 Caso continuo

- Supongamos ahora que tenemos una m.a.s $(X_1, ..., X_n)$ de una v.a. *continua* con función de densidad $f(x|\theta$ conocida salvo por los parámetros θ
- Una vez realizada la muestra tenemos $(x_1, ..., x_n)$ y estamos interesados en encontrar un estimador $\hat{\theta}$ del parámetro θ que maximice la función de densidad conjunta de la muestra.
- Para una muestra dada, la función de densidad conjunta de la muestra es

$$f(x_1, ..., x_n) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n f(x_i|\theta) = l(\theta|(x_1, ..., x_n)) \equiv l(\theta)$$

y la denominaremos *función de verosimilitud*.

3.3 Propiedades

- Así, el método de máxima verosimilitud consistirá en encontrar el valor $\hat{\theta}$ que maximiza la función de verosimilitud $l(\theta)$, que coincide con el valor $\hat{\theta}$ que maximiza $\log l(\theta)$
- El *estimador de máxima verosimilitud* $\hat{\theta}_{MV}$ se calcula igualando la primera derivada de $\log l(\theta)$ a 0 y comprobando que la segunda derivada en sus soluciones es negativa.
- El estimador de máxima verosimilitud $\hat{\theta}_{mv}$ tiene, bajo ciertas condiciones generales, las siguientes propiedades:

  · es *asintóticamente centrado*: a medida que crece el tamaño muestral el sesgo tiende a cero.

  · sigue *asintóticamente* una distribución *normal* con media θ y varianza $\frac{-1}{(\log l(\theta))''}$.

  · es un estimador *asintóticamente eficiente*: el de todos los estimadores asintóticamente centrados, el de máxima verosimilitud tiene menor varianza

  · es *invariante*: si $\hat{\theta}_m v$ es el estimador de máxima verosimilitud de θ, entonces $g(\hat{\theta}_m v)$ será el estimador de máxima verosimilitud de $h(\theta)$, para cualquier función h continua y biyectiva.

3.4 Distribuciones conocidas

- $X \sim \mathcal{B}(p)$ Bernoulli

$$l(p) = \prod_{i=1}^n P(x_i|p) = \prod_{i=1}^n \left(p^{x_i}(1-p)^(1-x_i)\right) = p^{\sum_i x_i (1-p)^{n-\sum_i x_i}}$$

$$\log l(p) = \log p^{\sum_i x_i (1-p)^{n-\sum_i x_i}} = (\textstyle\sum_i x_i) \log p + (n - \sum_i x_i) \log(1-p)$$

$$\hat{p} = \bar{x}$$

- $X \sim \mathcal{B}(k,p)$ Binomial, con k conocido

$$l(p) = \prod_{i=1}^{n} P(x_i|p) = \prod_{i=1}^{n} \left(\binom{k}{x_i} p^{x_i}(1-p)^{k-x_i}\right) = p^{\sum_i x_i}(1-p)^{nk-\sum_i x_i} \cdot \left(\prod_{i=1}^{n} \binom{k}{x_i}\right)$$

$$\log l(p) = p^{\sum_i x_i}(1-p)^{nk-\sum_i x_i} \cdot \log\left(\prod_{i=1}^{n} \binom{k}{x_i}\right) = cte + (\sum_i x_i)\log p + (nk - \sum_i x_i)\log(1-p)$$

$$\hat{p} = \frac{1}{nk}\sum_i x_i$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$ Normal

$$l(\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i|\mu,\sigma^2) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right) = \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$\log l(\mu,\sigma^2) = \log \frac{1}{(\sqrt{2\pi\sigma})^n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2\sum_i(x_i-\mu)^2}$$

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma^2} = \frac{1}{n}\sum(x_i - \bar{x})^2$$

- $X \sim \exp(\lambda)$ Exponencial

$$l(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i}\right) = \lambda^n e^{-\lambda\sum_i x_i}$$

$$\log l(\lambda) = \log \lambda^n e^{-\lambda\sum_i x_i} = n\log(\lambda) - \lambda\sum_i x_i$$

$$\hat{\lambda} = \frac{1}{\bar{x}}$$