Tema 5. Estimación de máxima verosimilitud y otros métodos

A. Hermoso Carazo
Universidad de Granada

Curso 2020/2021

Tema 5. Estimación de máxima verosimilitud y otros métodos A. Hermoso Carazo

En este tema presentamos algunos métodos para construir estimadores, que son especialmente útiles cuando la determinación de los óptimos no es simple, o cuando tales estimadores no existen.

Destacamos el *método de máxima verosimilitud*, la técnica de estimación más universal debido a su simplicidad y a las buenas propiedades de los estimadores que proporciona. Fue inicialmente sugerido por Gauss, a principios del siglo XIX, y rigurosamente formalizado por Fisher en 1925.

5.1. Estimación de máxima verosimilitud

Comenzamos con un ejemplo para ilustrar la idea intuitiva del método.

Ejemplo 5.1.1: *Sea $X \to \{B(n,p);\ n \in \{2,3\},\ p \in \{1/2, 1/3\}\}$. Encontrar un estimador de (n,p) basándose en una observación de X.*

La idea es, una vez observado el valor de X, tomar como estimación el valor de (n,p) que asigna más probabilidad a dicha observación:

$x \setminus (n,p)$	$(2,1/2)$	$(2,1/3)$	$(3,1/2)$	$(3,1/3)$
0	1/4	4/9	1/8	8/27
1	1/2	4/9	3/8	12/27
2	1/4	1/9	3/8	6/27
3	0	0	1/8	1/27

- $\max_{(n,p)} P_{(n,p)}(X=0) = 4/9 = P_{(2,1/3)}(X=0) \Rightarrow T(0) = (2, 1/3)$;
- $\max_{(n,p)} P_{(n,p)}(X=1) = 1/2 = P_{(2,1/2)}(X=1) \Rightarrow T(1) = (2, 1/2)$;
- $\max_{(n,p)} P_{(n,p)}(X=2) = 3/8 = P_{(3,1/2)}(X=2) \Rightarrow T(2) = (3, 1/2)$;
- $\max_{(n,p)} P_{(n,p)}(X=3) = 1/8 = P_{(3,1/2)}(X=3) \Rightarrow T(3) = (3, 1/2)$.

Ejemplo 5.1.2: *Encontrar el estimador máximo verosímil de μ basado en una muestra aleatoria simple, $(X_1, \ldots, X_n)$, de $X \to \{\mathcal{N}(\mu, \sigma_0^2);\ \mu \in \mathbb{R}\}$.*

El espacio paramétrico en este problema es $\mathbb{R}$, el espacio muestral $\mathbb{R}^n$, y la función de densidad de la muestra para cada valor de μ:

$$f_\mu^n(x_1, \ldots, x_n) = \frac{1}{(\sigma_0^2)^{n/2}(2\pi)^{n/2}} e^{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_0^2}, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Por tanto, fijada una realización muestral, $(x_1, \ldots, x_n) \in \mathbb{R}^n$, la función de verosimilitud asociada a dicha realización es una función de μ, definida sobre el espacio paramétrico, $\mathbb{R}$, como:

$$L_{x_1,\ldots,x_n}(\mu) = \frac{1}{(\sigma_0^2)^{n/2}(2\pi)^{n/2}} e^{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_0^2}, \quad \mu \in \mathbb{R}.$$

Para maximizar esta función, tomamos logaritmos, derivamos e igualamos a cero, obteniendo la *ecuación de verosimilitud:*

- $\ln L_{x_1,\ldots,x_n}(\mu) = -\frac{n}{2}\ln\left(\sigma_0^2 2\pi\right) - \frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma_0^2}$.
- $\frac{d \ln L_{x_1,\ldots,x_n}(\mu)}{d\mu} = \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma_0^2} = 0.$

La idea intuitiva de maximizar la función masa de probabilidad para cada observación muestral se generaliza al caso continuo maximizando la de densidad y se formaliza como sigue.

Función de verosimilitud

Sea una muestra aleatoria simple, $(X_1,\ldots,X_n)$, de $X \to \{P_\theta;\ \theta \in \Theta\}$ con función de densidad o masa de probabilidad $f_\theta^n(x_1,\ldots,x_n)$. Para cada realización muestral, $(x_1,\ldots,x_n) \in \chi^n$, se define la función de verosimilitud asociada a dicha realización como:

$$L_{x_1,\ldots,x_n} : \Theta \longrightarrow \mathbb{R}^+ \cup \{0\}$$
$$\theta \longmapsto L_{x_1,\ldots,x_n}(\theta) = f_\theta^n(x_1,\ldots,x_n).$$

Estimador de máxima verosimilitud (máximo verosímil)

Un estimador de θ, $\widehat{\theta}(X_1,\ldots,X_n)$, es de máxima verosimilitud si, para cada $(x_1\ldots,x_n) \in \chi^n$, la estimación $\widehat{\theta}(x_1\ldots,x_n)$ maximiza $L_{x_1,\ldots,x_n}$:

$$\forall (x_1,\ldots,x_n) \in \chi^n, \quad L_{x_1,\ldots,x_n}\left(\widehat{\theta}(x_1,\ldots,x_n)\right) = \max_{\theta\in\Theta} L_{x_1,\ldots,x_n}(\theta).$$

- Si $L_{x_1,\ldots,x_n}$ es derivable en el interior de Θ, el máximo suele buscarse derivando $\ln L_{x_1,\ldots,x_n}$ e igualando a cero, lo que da lugar a las denominadas *ecuaciones de verosimilitud:*

$$\theta = (\theta_1,\ldots,\theta_k) \ \to \ \frac{\partial \ln L_{x_1,\ldots,x_n}(\theta)}{\partial \theta_j} = 0, \quad \forall j = 1,\ldots,k.$$

  La solución o soluciones de estas ecuaciones proporcionan los extremos relativos de $L_{x_1,\ldots,x_n}$ en el interior de Θ, en los que debe evaluarse la función para obtener el máximo. Si las ecuaciones de verosimilitud no tienen solución, $L_{x_1,\ldots,x_n}$ es monótona, y el máximo se alcanza en la frontera.

  Puede probarse que si las ecuaciones de verosimilitud tienen solución única, dicha solución es un máximo.

- Si $L_{x_1,\ldots,x_n}$ no es derivable, o las ecuaciones de verosimilitud no son fácilmente resolubles, debe acudirse a métodos numéricos para la determinación de estos estimadores.

Ejemplo 5.1.2: *Encontrar el estimador máximo verosímil de μ basado en una muestra aleatoria simple, $(X_1, \dots, X_n)$, de $X \to \{\mathcal{N}(\mu, \sigma_0^2);\ \mu \in \mathbb{R}\}$.*

El espacio paramétrico en este problema es $\mathbb{R}$, el espacio muestral $\mathbb{R}^n$, y la función de densidad de la muestra para cada valor de μ:

$$f_\mu^n(x_1, \dots, x_n) = \frac{1}{(\sigma_0^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma_0^2}, \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Por tanto, fijada una realización muestral, $(x_1, \dots, x_n) \in \mathbb{R}^n$, la función de verosimilitud asociada a dicha realización es una función de μ, definida sobre el espacio paramétrico, $\mathbb{R}$, como:

$$L_{x_1,\dots,x_n}(\mu) = \frac{1}{(\sigma_0^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma_0^2}, \quad \mu \in \mathbb{R}.$$

Para maximizar esta función, tomamos logaritmos, derivamos e igualamos a cero, obteniendo la *ecuación de verosimilitud:*

- $\ln L_{x_1,\dots,x_n}(\mu) = -\frac{n}{2}\ln\left(\sigma_0^2 2\pi\right) - \frac{\sum\limits_{i=1}^{n}(x_i-\mu)^2}{2\sigma_0^2}.$
- $\frac{d \ln L_{x_1,\dots,x_n}(\mu)}{d\mu} = \frac{\sum\limits_{i=1}^{n}(x_i-\mu)}{\sigma_0^2} = 0.$

Observamos que esta ecuación tiene solución única, $\widehat{\mu}(x_1, \dots, x_n) = \overline{x} \in \mathbb{R}$ y, por tanto, queda garantizado que es un máximo (se puede comprobar haciendo la derivada segunda).

Puesto que este proceso se ha realizado para una realización muestral arbitraria, $(x_1, \dots, x_n) \in \mathbb{R}$, el estimador máximo verosímil de μ (único) es:

$$\widehat{\mu}(X_1, \dots, X_n) = \overline{X}.$$

Nota: El estimador máximo verosímil coincide con el UMVUE y, además, es *eficiente*.

Ejemplo 5.1.3: *Sea $X \to \{\mathcal{N}(\mu_0, \sigma^2);\quad \sigma^2 \in \mathbb{R}^+\}$. Encontrar el estimador de máxima verosimilitud de σ^2 basado en $(X_1, \ldots, X_n)$.*

Ahora, el espacio paramétrico es $\mathbb{R}^+$, y de nuevo, el espacio muestral $\mathbb{R}^n$, y la función de densidad de la muestra:

$$f^n_{\sigma^2}(x_1, \ldots, x_n) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2}, \quad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Por tanto, la función de verosimilitud asociada a $(x_1, \ldots, x_n) \in \mathbb{R}^n$ es:

$$L_{x_1,\ldots,x_n}(\sigma^2) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu_0)^2/2\sigma^2}, \quad \sigma^2 \in \mathbb{R}^+.$$

- $\ln L_{x_1,\ldots,x_n}(\sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma^2}.$
- $\frac{d\ln L_{x_1,\ldots,x_n}(\sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma^4} = 0$ *(ecuación de verosimilitud).*

Resolviendo esta ecuación obtenemos una solución única:

$$\widehat{\sigma}^2(x_1, \ldots, x_n) = \frac{\sum\limits_{i=1}^{n}(x_i-\mu_0)^2}{n}$$

Por tanto, esta solución es un máximo, y el único estimador máximo verosímil es:

$$\widehat{\sigma}^2(X_1, \ldots, X_n) = \frac{\sum\limits_{i=1}^{n}(X_i-\mu_0)^2}{n}.$$

Nota: Como en el caso de la media, el estimador máximo verosímil de la varianza es el UMVUE y es, además, *eficiente*. Esto, en general, no ocurre, como vemos en el siguiente y otros ejemplos.

Ejemplo 5.1.4: *Sea $X \to \{\mathcal{N}(\mu, \sigma^2);\ \ \mu \in \mathbb{R},\ \ \sigma^2 \in \mathbb{R}^+\}$. Encontrar los estimadores de máxima verosimilitud de μ y σ^2 basados en una muestra aleatoria simple, $(X_1, \dots, X_n)$.*

Tenemos ahora un parámetro bidimensional, (μ, σ^2), con espacio paramétrico $\mathbb{R} \times \mathbb{R}^+$. El espacio muestral es $\mathbb{R}^n$, y la densidad de la muestra:

$$f^n_{\mu,\sigma^2}(x_1, \dots, x_n) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}, \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Por tanto, la función de verosimilitud asociada a $(x_1, \dots, x_n) \in \mathbb{R}^n$ es:

$$L_{x_1,\dots,x_n}(\mu, \sigma^2) = \frac{1}{(\sigma^2)^{n/2}(2\pi)^{n/2}} e^{-\sum\limits_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}, \quad \mu \in \mathbb{R},\ \ \sigma^2 \in \mathbb{R}.$$

Ahora, al ser el parámetro bidimensional, hay que maximizar $L_{x_1,\dots,x_n}(\mu, \sigma^2)$ en los dos argumentos y obtendremos dos ecuaciones de verosimilitud, que hay que resolver conjuntamente.

- $\ln L_{x_1,\dots,x_n}(\mu, \sigma^2) = -\dfrac{n}{2} \ln \sigma^2 - \dfrac{n}{2} \ln (2\pi) - \dfrac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}.$

- $\dfrac{\partial \ln L_{x_1,\dots,x_n}(\mu, \sigma^2)}{\partial \mu} = \dfrac{\sum\limits_{i=1}^{n}(x_i - \mu)}{\sigma^2} = 0.$
- $\dfrac{\partial \ln L_{x_1,\dots,x_n}(\mu, \sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{\sum\limits_{i=1}^{n}(x_i - \mu)^2}{2\sigma^4} = 0.$

Resolviendo la primera, obtenemos $\widehat{\mu}(x_1, \dots, x_n) = \overline{x}$, y sustituyendo este valor en la segunda, $\widehat{\sigma}^2(x_1, \dots, x_n) = \dfrac{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}{n}$.

Por tanto, los únicos estimadores máximo verosímiles de μ y σ^2, son *la media y la varianza muestral*, respectivamente:

$$\widehat{\mu}(X_1, \dots, X_n) = \overline{X}, \qquad \widehat{\sigma}^2(X_1, \dots, X_n) = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n}.$$

Nota: $\widehat{\mu}(X_1, \dots, X_n)$ es el mismo si σ^2 es conocida o no. Sin embargo, si $\mu = \mu_0$ es conocida, el estimador máximo verosímil de σ^2 depende de μ_0, y si μ es desconocida, se sustituye por $\overline{X}$ (su estimador máximo verosímil).

Notamos también que el estimador máximo verosímil de σ^2, la varianza muestral, no es insesgado, y no coincide con el UMVUE que, en este caso, es S^2.

Ejemplo 5.1.5: *Sea $X \to \{P_N;\ N \in \mathbb{N}\}$, siendo P_N la distribución uniforme discreta en $\{1, \ldots, N\}$. Encontrar, si existe, un estimador máximo verosímil de N basado en una muestra aleatoria simple, $(X_1, \ldots, X_n)$.*

El espacio paramétrico es $\mathbb{N}$ y $\chi_N = \{1, \ldots, N\}$. Entonces, el conjunto de posibles valores de X es $\bigcup_{N \in \mathbb{N}} \chi_N = \mathbb{N}$, y el espacio muestral $\mathbb{N}^n$.

La función masa de probabilidad de la muestra, para cada valor $N \in \mathbb{N}$, es:

$$P_N(X_1 = x_1, \ldots, X_n = x_n) = \begin{cases} \dfrac{1}{N^n}, & \text{máx}\, x_i \leq N \\ 0, & \text{máx}\, x_i > N, \end{cases} \qquad (x_1, \ldots, x_n) \in \mathbb{N}^n.$$

Por lo tanto, la función de verosimilitud asociada a cada posible realización muestral es una función sobre $\mathbb{N}$ definida como sigue:

$$L_{x_1,\ldots,x_n}(N) = \begin{cases} 0, & N < \text{máx}\, x_i \\ \dfrac{1}{N^n}, & N \geq \text{máx}\, x_i, \end{cases} \qquad N \in \mathbb{N}.$$

Ya que N es un parámetro discreto, esta función no es derivable, pero es inmediato ver que alcanza su valor máximo en $\text{máx}\, x_i \in \mathbb{N}$. Así, el único estimador máximo verosímil de N es $\widehat{N}(X_1, \ldots, X_n) = \text{máx}\, X_i$.

Nota: De nuevo, el estimador máximo verosímil no es insesgado.

Ejemplo 5.1.6: *Sea $X \to \{P_\theta;\ \theta \in \mathbb{R}\}$, siendo $f_\theta(x) = e^{-x+\theta}$, $x \geq \theta$. Encontrar, si existe, un estimador de máxima verosimilitud de θ basado en una muestra aleatoria simple, $(X_1, \ldots, X_n)$.*

El espacio paramétrico es $\mathbb{R}$, $\chi_\theta = [\theta, +\infty)$, y $\chi = \bigcup_{\theta \in \mathbb{R}} [\theta, +\infty) = \mathbb{R}$. Por tanto, el espacio muestral es $\mathbb{R}^n$. La función de densidad de la muestra, para cada valor del parámetro, $\theta \in \mathbb{R}$, es:

$$f_\theta^n(x_1 \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i) = \begin{cases} e^{-\sum\limits_{i=1}^n x_i + n\theta}, & \text{mín}\, x_i \geq \theta \\ 0, & \text{mín}\, x_i < \theta, \end{cases} \qquad (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Así, la función de verosimilitud asociada a cada realización muestral es:

$$L_{x_1,\ldots,x_n}(\theta) = \begin{cases} e^{-\sum\limits_{i=1}^n x_i + n\theta}, & \theta \leq \text{mín}\, x_i \\ 0, & \theta > \text{mín}\, x_i, \end{cases} \qquad \theta \in \mathbb{R}.$$

Esta función, no negativa, es creciente para $\theta \leq \text{mín}\, x_i$ y, por tanto, el máximo se alcanza en $\text{mín}\, x_i$.

Entonces, el único estimador máximo verosímil de θ es $\widehat{\theta}(X_1, \ldots, X_n) = \text{mín}\, X_i$.

Ejemplo 5.1.7: *Sea $X \to \{U[\theta - 1/2, \theta + 1/2];\ \theta \in \mathbb{R}\}$. Encontrar, si existe, un estimador máximo verosímil de θ basado en $(X_1, \ldots, X_n)$.*

$\Theta = \mathbb{R}$, $\chi_\theta = [\theta - 1/2, \theta + 1/2]$ y $\chi = \bigcup_{\theta \in \mathbb{R}} [\theta - 1/2, \theta + 1/2] = \mathbb{R}$. Por tanto, el espacio muestral es $\mathbb{R}^n$. La función de densidad de la muestra, para cada valor $\theta \in \mathbb{R}$, es:

$$f_\theta^n(x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i) = \begin{cases} 1, & \text{mín}\, x_i \geq \theta - 1/2 \ \text{ y } \ \text{máx}\, x_i \leq \theta + 1/2 \\ 0, & \text{mín}\, x_i < \theta - 1/2 \ \text{ ó } \ \text{máx}\, x_i > \theta + 1/2, \end{cases}$$

$$(x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Entonces, la función de verosimilitud asociada a cada realización muestral es:

$$L_{x_1, \ldots, x_n}(\theta) = \begin{cases} 1, & \text{máx}\, x_i - 1/2 \leq \theta \leq \text{mín}\, x_i + 1/2 \\ 0, & \theta < \text{máx}\, x_i - 1/2 \ \text{ ó } \ \theta > \text{mín}\, x_i + 1/2, \end{cases} \qquad \theta \in \mathbb{R}.$$

El valor máximo de esta función es 1, y se alcanza en cualquier número real en $\left[\text{máx}\, x_i - 1/2,\ \text{mín}\, x_i + 1/2\right]$. Por lo tanto, cualquier estadístico con valores en ese intervalo es estimador máximo verosímil de θ. Por ejemplo:

$$\alpha(\text{mín}\, X_i + 1/2) + (1 - \alpha)(\text{máx}\, X_i - 1/2), \quad \forall \alpha \in [0, 1].$$

Nota: Aquí se muestra la no unicidad de estimadores de máxima verosimilitud.

Ejemplo 5.1.8: *Dada $X \to \{B(1, p);\ p \in [1/4, 3/4]\}$, comparar el error cuadrático medio del estimador máximo verosímil de p basado en una observación con el de $T(X) = 1/2$.*

Comenzamos calculando el estimador máximo verosímil y, para ello, las estimaciones correspondientes a $X = 0$ y $X = 1$:

- $L_0(p) = P_p(X = 0) = 1 - p$, $\ p \in [1/4, 3/4] \to$ decreciente $\Rightarrow \widehat{p}(0) = 1/4$.
- $L_1(p) = P_p(X = 1) = p$, $\ p \in [1/4, 3/4] \to$ creciente $\Rightarrow \widehat{p}(1) = 3/4$.

$$\Downarrow$$

$$\widehat{p}(X) = \frac{2X + 1}{4}.$$

Calculamos ahora los errores cuadráticos medios de $\widehat{p}(X)$ y $T(X) = 1/2$:

- $E_p\left[(\widehat{p}(X) - p)^2\right] = \left(\widehat{p}(0) - p\right)^2 P_p(X = 0) + \left(\widehat{p}(1) - p\right)^2 P_p(X = 1)$
  $= (1/4 - p)^2(1 - p) + (3/4 - p)^2 p = 1/16, \quad \forall p \in [1/4, 3/4].$
- $E_p\left[(T(X) - p)^2\right] = E_p\left[(1/2 - p)^2\right] = p^2 - p + 1/4, \quad \forall p \in [1/4, 3/4].$

Ya que $p^2 - p + 1/4 \leq 1/16$, $\ \forall p \in [1/4, 3/4]$, $T(X)$ es uniformemente mejor que $\widehat{p}(X)$ si se trata de minimizar el error cuadrático medio.

Los ejemplos anteriores muestran algunos inconvenientes de los estimadores de máxima verosimilitud: no tienen por qué ser únicos, ni insesgados, ni minimizar el error cuadrático medio. Sin embargo, estos estimadores tienen importantes propiedades, sobre todo para grandes muestras.

Propiedades asintóticas del estimador máximo verosímil: *Bajo condiciones bastante generales (que incluyen las condiciones de regularidad de Fréchet-Cramér-Rao), si las ecuaciones de verosimilitud tienen solución única, $\widehat{\theta}(X_1, \ldots, X_n)$, esta solución satisface:*

- *Consistencia fuerte:* $\widehat{\theta}(X_1, \ldots, X_n) \xrightarrow[n\to+\infty]{c.s.} \theta, \quad \forall\theta \in \Theta.$
- *Normalidad asintótica:* $\sqrt{n}\left[\widehat{\theta}(X_1, \ldots, X_n) - \theta\right] \xrightarrow[n\to+\infty]{L} \mathcal{N}(0, I_X^{-1}(\theta)), \quad \forall\theta\in\Theta.$

La normalidad asintótica significa que, para muestras grandes, la distribución de $\widehat{\theta}(X_1, \ldots, X_n)$ es aproximadamente normal,

$$\widehat{\theta}(X_1, \ldots, X_n) \approx \mathcal{N}\left(\theta, (nI_X(\theta))^{-1}\right).$$

Así, como $(nI_X(\theta))^{-1} = I^{-1}_{X_1,\ldots,X_n}(\theta)$ es la cota de Fréchet-Crámer-Rao para la varianza de estimadores insesgados de θ, el estimador es *asintóticamente insesgado* y *asintóticamente eficiente*.

Para muestras de tamaño arbitrario, además de su simplicidad, los estimadores de máxima verosimilitud tienen, entre otras, las siguientes propiedades.

Relación con los estadísticos suficientes: *Sea $T(X_1, \ldots, X_n)$ un estadístico suficiente para la familia de distribuciones considerada. Si θ admite un estimador máximo verosímil, $\widehat{\theta}(X_1, \ldots, X_n)$, dicho estimador es función de $T(X_1, \ldots, X_n)$.*

Demostración: Ya que $L_{x_1,\ldots,x_n}(\theta) = f_\theta^n(x_1, \ldots, x_n)$, por el *Teorema de factorización de Neyman-Fisher*:

$$\forall(x_1, \ldots, x_n) \in \chi^n, \quad L_{x_1,\ldots,x_n}(\theta) = h(x_1, \ldots, x_n)g_\theta(T(x_1, \ldots, x_n)), \quad \theta \in \Theta.$$

Entonces, como $h(x_1, \ldots, x_n)$ no depende de θ, los extremos de $L_{x_1,\ldots,x_n}$ coinciden con los de la función $g_t'(\theta) = g_\theta(t)$, con $t = T(x_1, \ldots, x_n)$; puesto que g_t' sólo depende de $(x_1, \ldots, x_n)$ a través de t, sus extremos, y en particular, el valor que maximiza $L_{x_1,\ldots,x_n}$, sólo dependerán de $t = T(x_1, \ldots, x_n)$. $\square$

Nota: La propiedad no implica que un estimador máximo verosímil sea suficiente. Por ejemplo, en el Ejemplo 5.1.7, $\min X_i + 1/2$ es estimador máximo verosímil de θ, pero no es suficiente (el suficiente es $(\min X_i, \max X_i)$).

Relación con los estimadores eficientes: *Si la familia de distribuciones es regular, y $T(X_1,\ldots,X_n)$ es un estimador eficiente para θ, entonces $T(X_1,\ldots,X_n)$ es el único estimador máximo verosímil de θ.*

Demostración: Por el *Teorema de caracterización de estimadores eficientes*:

$$\forall\theta\in\Theta,\ \exists a(\theta)\neq 0\ /\ P_\theta\left(\frac{\partial\ln f_\theta^n(X_1,\ldots,X_n)}{\partial\theta}=a(\theta)\big[T(X_1,\ldots,X_n)-\theta\big]\right)=1,$$

y, por tanto, $L_{x_1,\ldots,x_n}(\theta)=f_\theta^n(x_1,\ldots,x_n)$ es derivable y

$$\frac{\partial\ln L_{x_1,\ldots,x_n}(\theta)}{\partial\theta}=a(\theta)\big[T(x_1,\ldots,x_n)-\theta\big].$$

Así, la ecuación de verosimilitud es:

$$a(\theta)\big[T(x_1,\ldots,x_n)-\theta\big]=0,$$

y su única solución (y, por lo tanto, máximo de $L_{x_1,\ldots,x_n}$), teniendo en cuenta que $a(\theta)\neq 0$, es $\widehat{\theta}(x_1,\ldots,x_n)=T(x_1,\ldots,x_n)$, $\forall(x_1,\ldots,x_n)\in\chi^n$. Por lo tanto, $T(X_1,\ldots,X_n)$ es el único estimador máximo verosímil de θ. $\square$

Nota: El recíproco, en general, no es cierto. De hecho, como hemos visto en algunos ejemplos, un estimador máximo verosímil no tiene ni que ser insesgado.

Una de las principales propiedades de los estimadores máximo verosímiles es la de *invarianza*, que requiere introducir los siguientes conceptos.

Función de verosimilitud de una función paramétrica

Sea $X\to\{P_\theta;\ \theta\in\Theta\}$, $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X, $g:\Theta\to\Lambda=g(\Theta)$ una función paramétrica arbitraria, y $\lambda=g(\theta)$. Para cada realización muestral, $(x_1,\ldots,x_n)\in\chi^n$, se define la función de verosimilitud de λ asociada a dicha realización como:

$$\begin{aligned} M_{x_1,\ldots,x_n}:\Lambda &\longrightarrow \mathbb{R}^+\cup\{0\}\\ \lambda &\longmapsto M_{x_1,\ldots,x_n}(\lambda)=\sup_{\theta\in g^{-1}(\lambda)}L_{x_1,\ldots,x_n}(\theta).\end{aligned}$$

Estimador de máxima verosimilitud de una función paramétrica

En las condiciones anteriores, $\widehat{\lambda}(X_1,\ldots,X_n)$ es estimador máximo verosímil de λ si, para cada $(x_1,\ldots,x_n)\in\chi^n$, la estimación $\widehat{\lambda}(x_1,\ldots,x_n)$ maximiza $M_{x_1,\ldots,x_n}$:

$$\forall(x_1,\ldots,x_n)\in\chi^n,\quad M_{x_1,\ldots,x_n}\big(\widehat{\lambda}(x_1,\ldots,x_n)\big)=\max_{\lambda\in\Lambda}M_{x_1,\ldots,x_n}(\lambda).$$

Teorema de invarianza de Zehna

Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X \to \{P_\theta;\ \theta\in\Theta\}$ y $g:\Theta\to\Lambda$ una función medible. Si $\widehat{\theta}(X_1,\dots,X_n)$ es estimador máximo verosímil de θ, $g(\widehat{\theta}(X_1,\dots,X_n))$ es estimador máximo verosímil de $g(\theta)$.[1]

Demostración: Sea $\lambda = g(\theta)$ y, fijada una realización muestral, $(x_1,\dots,x_n) \in \chi^n$, notemos $\widehat{\lambda} \equiv g(\widehat{\theta}(x_1,\dots,x_n))$ (así, $\widehat{\theta}(x_1,\dots,x_n)\in g^{-1}(\widehat{\lambda})$).

$$a)\ M_{x_1,\dots,x_n}(\widehat{\lambda}) = \sup_{\theta\in g^{-1}(\widehat{\lambda})} L_{x_1,\dots,x_n}(\theta) \overset{\widehat{\theta}\in g^{-1}(\widehat{\lambda})}{=} L_{x_1,\dots,x_n}(\widehat{\theta}(x_1,\dots,x_n)).$$

$$b)\ \forall\lambda\in\Lambda,\ M_{x_1,\dots,x_n}(\lambda) = \sup_{\theta\in g^{-1}(\lambda)} L_{x_1,\dots,x_n}(\theta) \le \sup_{\theta\in\Theta} L_{x_1,\dots,x_n}(\theta) = L_{x_1,\dots,x_n}(\widehat{\theta}(x_1,\dots,x_n)).$$

Entonces,

$$\forall\lambda\in\Lambda,\ M_{x_1,\dots,x_n}(\lambda) \overset{b)}{\le} L_{x_1,\dots,x_n}(\widehat{\theta}(x_1,\dots,x_n)) \overset{a)}{=} M_{x_1,\dots,x_n}(\widehat{\lambda}).$$

Esto significa que $\widehat{\lambda} = g(\widehat{\theta}(x_1,\dots,x_n))$ maximiza $M_{x_1,\dots,x_n}$, para cada $(x_1,\dots,x_n)\in\chi^n$. Por tanto, $g(\widehat{\theta}(X_1,\dots,X_n))$ es estimador máximo verosímil de $\lambda = g(\theta)$. □

[1] La medibilidad se exige para que $g(\widehat{\theta}(X_1,\dots,X_n))$ sea estimador.

Ejemplo 5.1.9: *La duración de cierto tipo de bombillas es una variable aleatoria, X, con distribución exponencial de media θ, desconocida. Después de observar la longitud de vida de n de estas bombillas seleccionadas de forma independiente, estimar, por máxima verosimilitud, la probabilidad de que la duración de una bombilla sea superior a 500 horas.*

Comenzamos calculando la función a estimar, $g(\theta) = P_\theta(X > 500)$:

$$f_\theta(x) = \frac{1}{\theta}e^{-x/\theta},\ x>0 \to g(\theta) = \int_{500}^{+\infty}\frac{1}{\theta}e^{-x/\theta}dx = \left[-e^{-x/\theta}\right]_{500}^{+\infty} = e^{-500/\theta}.$$

Ahora, basta obtener el estimador máximo verosímil de θ y aplicar el *Teorema de Invarianza*.

$$(x_1,\dots,x_n)\in(\mathbb{R}^+)^n \to L_{x_1,\dots,x_n}(\theta) = f_\theta^n(x_1,\dots,x_n) = \frac{e^{-\sum_{i=1}^n x_i/\theta}}{\theta^n},\ \ \theta\in\mathbb{R}^+.$$

- $\ln L_{x_1,\dots,x_n}(\theta) = -n\ln\theta - \dfrac{\sum_{i=1}^n x_i}{\theta}.$
- $\dfrac{\partial \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = -\dfrac{n}{\theta} + \dfrac{\sum_{i=1}^n x_i}{\theta^2} = 0 \Rightarrow \widehat{\theta}(x_1,\dots,x_n) = \dfrac{\sum_{i=1}^n x_i}{n} = \overline{x}\in\mathbb{R}^+.$

Entonces, $\overline{X}$ es el estimador máximo verosímil de θ y el de $e^{-500/\theta}$ es $e^{-500/\overline{X}}$.

Ejemplo 5.1.10: *Tomadas n observaciones independientes de una variable $X \to \{\mathcal{P}(\lambda);\ \lambda > 0\}$, sólo se sabe que ha aparecido k veces el valor 0. A partir de esta información, obtener la estimación más verosímil de λ.*

Se han tomado n observaciones independientes de X (muestra aleatoria simple), pero no se dispone de sus valores concretos, de forma que la estimación máximo verosímil basada en dichos valores, $\widehat{\lambda}(x_1, \ldots, x_n)$, no soluciona el problema.

Lo información disponible es el *número de valores muestrales iguales a 0* y, por tanto, la estimación de λ debe basarse exclusivamente en esta información. Esto es, disponemos de una observación de la siguiente variable:

Y : *número de valores iguales a 0 en n observaciones independientes de X*,

y buscamos la estimación más verosímil de λ basada en esta observación. Concretamente, como $Y = k$, buscamos $\widehat{\lambda}(k)$.

Para ello, debemos maximizar la función de verosimilitud asociada a la observación k, $L_k(\lambda) = P_\lambda(Y = k)$, de manera que lo primero es calcular la distribución de Y.

Ya que Y contabiliza el número de éxitos (número de valores iguales a cero) en n pruebas de Bernoulli independientes (n observaciones de X independientes) y la probabilidad de éxito es $P_\lambda(X = 0)$, es obvio que, $Y \to B\big(n, P_\lambda(X = 0)\big)$.

$$Y \to B(n,p), \quad p = P_\lambda(X = 0) = e^{-\lambda}.$$

Por tanto, basta estimar p a partir de la observación $Y = k$ y aplicar el *Teorema de Invarianza* para estimar $\lambda = -\ln p$.

- $L_k(p) = P_p(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 < p < 1.$
- $\ln L_k(p) = \ln \binom{n}{k} + k \ln p + (n-k)\ln(1-p).$
- $\dfrac{\partial \ln L_k(p)}{\partial p} = \dfrac{k}{p} - \dfrac{n-k}{1-p} = 0 \ \Rightarrow\ \widehat{p}(k) = \dfrac{k}{n} \in [0,1].$

Por tanto, aplicando la invarianza, la estimación más verosímil de $\lambda = -\ln p$ es $\widehat{\lambda}(k) = -\ln(k/n) = \ln(n/k)$.

Notemos que $\widehat{\lambda}(0) = +\infty$ y $\widehat{\lambda}(n) = 0$ no son consecuentes con la distribución de Poisson, por lo que, en tales casos, convendría tomar más datos para hacer la estimación, o replantearse si, realmente, X tiene dicha distribución.

El teorema de invarianza de Zehna es especialmente útil en situaciones de este tipo, en las que no se conocen los valores muestrales concretos, sino sólo alguna función de dichos valores.

5.2 Otros métodos de estimación

5.2.1. Método de los momentos

Fue propuesto por K. Pearson en 1894. Se usa para estimar funciones paramétricas que pueden expresarse en términos de ciertos momentos de la distribución; los estimadores se obtienen por sustitución directa de los momentos poblacionales por los muestrales.

Descripción: *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable $X \to \{P_\theta;\ \theta \in \Theta\}$, y sean*

$$m_{\theta,j} = E_\theta\left[X^j\right], \quad j \in \mathbb{N}, \qquad A_j = \frac{\sum\limits_{i=1}^{n} X_i^j}{n}, \quad j \in \mathbb{N}.$$

El método consiste en estimar cada función del tipo $h(m_{\theta,1}, \ldots, m_{\theta,k})$, con h medible, por $h(A_1, \ldots, A_k)$.

Nota: La medibilidad de h se exige para que $h(A_1, \ldots, A_k)$ sea un estadístico.

Tema 5. Estimación de máxima verosimilitud y otros métodos A. Hermoso Carazo
5.2 Otros métodos de estimación
5.2.1. Método de los momentos

Nota: Si una función paramétrica puede expresarse de distintas formas en términos de los momentos, se suele usar la expresión que involucre los momentos de menor orden. Por ejemplo, si $X \to \{\mathcal{P}(\lambda);\ \lambda \in \mathbb{R}^+\}$:

$$\lambda = E_\lambda[X] = m_1, \quad \text{y} \quad \lambda = Var_\lambda[X] = m_2 - m_1^2.$$

Así, el parámetro λ puede expresarse de dos formas en términos de los momentos poblacionales, y lo usual es estimarlo por la primera expresión, $\widehat{\lambda} = A_1 = \overline{X}$.

El método es puramente intuitivo, carece de justificación teórica y en ocasiones puede conducir a estimaciones absurdas. Sin embargo, según la función estimada, puede proporcionar buenos estimadores. Por ejemplo:

- *Los estimadores de los momentos no centrados, o de funciones lineales de ellos son insesgados.* (Como vimos en el Tema 1, $E_\theta[A_j] = m_{\theta,j}$).
- *Los estimadores de los momentos no centrados, o de funciones continuas de ellos son fuertemente consistentes:*

$$A_j^{(n)} = \frac{\sum\limits_{i=1}^{n} X_i^j}{n} \xrightarrow[n\to\infty]{c.s} m_{\theta,j}, \quad \forall \theta \in \Theta.$$

Ejemplo 5.2.2: *En seis medidas del punto de ebullición de un compuesto, los errores, en grados centígrados, fueron 0.33, 0.12, 0.94, 0.2, 0.15 y 0.13 . Si se supone que estos datos son observaciones independientes de una variable aleatoria X con función de densidad*

$$f_\theta(x) = \frac{2(\theta - x)}{\theta^2}, \quad 0 < x < \theta,$$

dar una estimación del parámetro θ mediante el método de los momentos.

Calculamos los momentos de esta distribución hasta poder despejar θ:

$$m_1 = E_\theta[X] = \frac{2}{\theta^2}\int_0^\theta (\theta x - x^2)dx = \frac{\theta}{3} \Rightarrow \theta = 3m_1 \Rightarrow \widehat{\theta}(X_1, \ldots, X_n) = 3\overline{X}.$$

En nuestro caso, $\overline{x} = 0.3116$. Por tanto, la estimación de θ a partir de los datos es 0.9348, lo que contradice a la observación 0.94.

Como ya hemos indicado, en ocasiones el método puede llevar a conclusiones erróneas.

5.2.2. Método de mínimos cuadrados

Este método, inicialmente propuesto por Gauss (1809), se usa en un contexto diferente al que hemos trabajado hasta el momento, donde el parámetro a estimar define la distribución de la variable.

El método de mínimos cuadrados se usa para estimar parámetros que definen una determinada magnitud, a partir de observaciones de la misma afectadas por errores de medida aditivos, aleatorios y no observables.

Sea $\varphi(t, \theta)$ una magnitud cuyos valores dependen de una serie de condiciones experimentales, t, y de un parámetro desconocido, θ, que pretende estimarse a partir de observaciones de esta magnitud. En principio, la determinación de θ puede hacerse midiendo $\varphi(t, \theta)$ bajo distintas condiciones experimentales, $t_1, \ldots, t_n$, y despejando el valor de θ. Sin embargo, en la práctica, las observaciones estarán sujetas a errores de medida aleatorios y, por lo tanto, serán variables aleatorias:

$$\begin{aligned} X_1 &= \varphi(t_1, \theta) + \varepsilon_1 \\ &\vdots \\ X_n &= \varphi(t_n, \theta) + \varepsilon_n. \end{aligned}$$

En situaciones de este tipo, para estimar θ a partir de $X_1, \ldots, X_n$, suele usarse el *método de mínimos cuadrados*, que consiste en tomar como estimador de θ el que minimice la suma de los cuadrados de los errores cometidos al aproximar los verdaderos valores, $\varphi(t_i, \theta)$, por los observados, X_i:

$$\textit{Minimizar} \quad \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(X_i - \varphi(t_i, \theta)\right)^2.$$

Las propiedades de los estimadores de mínimos cuadrados dependen de la situación particular considerada. En el Tema 8 se aplicará para estimar los parámetros de un modelo lineal.

Ejemplo 5.2.3: *Para estimar la aceleración de la gravedad, θ, en una ciudad, se deja caer un objeto desde un altura determinada durante ciertos intervalos de tiempo, $t_1, \ldots, t_n$, y se mide el espacio recorrido. Si X_i representa la medida correspondiente al tiempo t_i, con un error de medida ε_i, dar la estimación mínimo cuadrática de θ.*

El espacio recorrido, en términos de la aceleración de la gravedad y el tiempo, se expresa como $e = \theta t^2/2$. Por lo tanto, las medidas X_i son:

$$X_i = \frac{\theta t_i^2}{2} + \varepsilon_i, \quad i = 1, \ldots, n.$$

Para estimar θ por mínimos cuadrados, tenemos que minimizar la suma de cuadrados de los errores:

$$L(\theta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(X_i - \frac{\theta t_i^2}{2}\right)^2.$$

Entonces, derivando e igualando a cero obtenemos:

- $$\frac{dL(\theta)}{d\theta} = -\sum_{i=1}^{n} t_i^2 \left(X_i - \frac{\theta t_i^2}{2}\right) = 0 \Rightarrow \widehat{\theta}(X_1, \ldots, X_n) = \frac{2\sum\limits_{i=1}^{n} X_i t_i^2}{\sum\limits_{i=1}^{n} t_i^4}.$$

Ya que la derivada segunda es positiva, esta solución proporciona el estimador de mínimos cuadrados.

Ejercicio 1:

$(x_1, \dots, x_n) \in \mathbb{R}^n \Rightarrow f_\theta(x_1, \dots, x_n) = \{ 0, \text{ si } \min x_i < \theta \quad (x_1, \dots, x_n) \in \mathbb{R}^n$

Por tanto, fijada una realización muestral, $(x_1, \dots, x_n) \in \mathbb{R}^n$, la función de verosimilitud asociada a dicha realización es una función de θ, definida sobre el espacio paramétrico $\mathbb{R}$, como:

$$L_{x_1, \dots, x_n}(\theta) = \begin{cases} e^{n\theta - \sum_{i=1}^{n} x_i}, & \text{si } \theta \leq \min x_i \\ 0, & \text{si } \theta > \min x_i \end{cases} \qquad \theta \in \mathbb{R}$$

- Esta función, no negativa, es creciente para $\theta \leq \min x_i$ y por tanto, el máximo se alcanza en $\min x_i$.

Entonces, el único estimador máximo verosímil de θ es $\hat{\theta}(x_1, \dots, x_n) = \min X_i$

y de e^θ es $e^{\min x_i}$.

Veamos si son insesgados:

Al ser $T(X_1, \dots, X_n) = T$ un estadístico suficiente y completo y sabemos que el UMVUE es único, como los estimadores que hemos calculado no coinciden con los UMVUES del ejercicio 7, podemos concluir que no son insesgados.

otra forma de razonarlo es con la definición

$$E(\hat{\theta}) = E(\min X_i) \underset{\downarrow}{=} E(t) = \int^{\infty} t f(t)\, dt = \int^{\infty} t\, n e^{n(\theta - t)} dt = \frac{n\theta + 1}{n} = \theta + \frac{1}{n} \neq \theta \Rightarrow$$ Por tanto, $\min x_i$ no es insesgado

meet.google.com está compartiendo tu pantalla. Dejar de compartir Ocultar

$F_{\min}(x) = 1 - (\dots) = 1 - e \dots, \forall x \geq \theta$

$F_\theta(x) = \int^{x} e^{\theta - t} dt = 1 - e^{\theta - x}, \forall x \geq \theta$

Problema 2

Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X con distribución exponencial. Basándose en los resultados del problema 9 de la relación 4, encontrar los estimadores máximo verosímiles de la media y de la varianza de X.

$$X \to \{exp(\lambda);\ \lambda > 0\} \longrightarrow \begin{cases} E_\lambda[X] = \dfrac{1}{\lambda} \\ Var_\lambda[X] = \dfrac{1}{\lambda^2}. \end{cases}$$

$\forall a \neq 0,\ b \in \mathbb{R},\ a\overline{X} + b$ *es el estimador eficiente de* $aE_\lambda[X] + b = \frac{a}{\lambda} + b.$

- *Estimador máximo verosímil de $E_\lambda[X]$:*

  Ya que $\overline{X}$ es eficiente para $E_\lambda[X]$, es el único estimador máximo verosímil.

- *Estimador máximo verosímil de $Var_\lambda[X] = 1/\lambda^2$:*

$\overline{X}$ estimador máximo verosímil de $1/\lambda \Rightarrow \overline{X}^2$ estimador máximo verosímil de $1/\lambda^2$.

$\Uparrow$

Teorema de Invarianza

Problema 4

Sea $(X_1, \cdots, X_n)$ una muestra de $X \to \{B(k_0, p);\ p \in (0,1)\}$. Estimar, por máxima verosimilitud y por el método de los momentos, el parámetro p y la varianza de X.

Aplicación: Se lanza 10 veces un dado cargado y se cuenta el número de veces que sale un 4. Este experimento se realiza 100 veces de forma independiente, obteniéndose los siguientes resultados:

n^o de 4	0	1	2
frecuencia	84	15	1

Estimar, a partir de estos datos, la probabilidad de salir un cuatro.

$p \in (0,1), \quad \chi^n = \{0, 1, \ldots, k_0\}^n$

$$P_p(X_1 = x_1, \ldots, X_n = x_n) = \binom{k_0}{x_1} \cdots \binom{k_0}{x_n} p^{\sum_{i=1}^n x_i} (1-p)^{nk_0 - \sum_{i=1}^n x_i}, \quad (x_1, \ldots, x_n) \in \chi^n.$$

Relación 5.Estimación de máxima verosimilitud y otros métodos A. Hermoso Carazo
└ Problema 4

∗ *Estimación máximo verosímil de p y de $Var_p[X] = k_0 p(1-p)$:*

$$(x_1, \ldots, x_n) \in \chi^n \to L_{x_1,\ldots,x_n}(p) = \binom{k_0}{x_1} \cdots \binom{k_0}{x_n} p^{\sum_{i=1}^n x_i} (1-p)^{nk_0 - \sum_{i=1}^n x_i}, \quad p \in (0,1)$$

- $\ln L_{x_1,\ldots,x_n}(p) = \ln\left[\binom{k_0}{x_1} \cdots \binom{k_0}{x_n}\right] + \left(\sum_{i=1}^n x_i\right) \ln p + \left(nk_0 - \sum_{i=1}^n x_i\right) \ln(1-p).$
- $\dfrac{\partial \ln L_{x_1,\ldots,x_n}(p)}{\partial p} = \dfrac{\sum_{i=1}^n x_i}{p} - \dfrac{nk_0 - \sum_{i=1}^n x_i}{(1-p)} = 0 \Rightarrow \widehat{p}(x_1, \ldots, x_n) = \dfrac{\overline{x}}{k_0} \in [0,1].$

$$\widehat{p}(X_1, \ldots, X_n) = \frac{\overline{X}}{k_0} \overset{T.\ Invarianza}{\Rightarrow} \widehat{Var_p[X]}(X_1, \ldots, X_n) = \overline{X}\left(1 - \frac{\overline{X}}{k_0}\right).$$

∗ *Estimación de p y de $Var_p[X] = k_0 p(1-p)$ por el método de los momentos:*

- $m_1 = k_0 p \Rightarrow p = \dfrac{m_1}{k_0} \Rightarrow \widehat{p}_m(X_1, \ldots, X_n) = \dfrac{\overline{X}}{k_0}.$
- $Var_p[X] = m_1\left(1 - \dfrac{m_1}{k_0}\right) \Rightarrow \widehat{Var_{pm}}[X](X_1, \ldots, X_n) = \overline{X}\left(1 - \dfrac{\overline{X}}{k_0}\right).$

Nota: Se podía pensar en estimar $Var_p[X] = m_2 - m_1^2$ por $A_2 - A_1^2$, pero debe usarse la expresión que involucre el menor número de momentos.

Aplicación: Se lanza 10 veces un dado cargado y se cuenta el número de veces que sale un 4. Este experimento se realiza 100 veces de forma independiente, obteniéndose los siguientes resultados:

n^o de 4	0	1	2
frecuencia	84	15	1

Estimar, a partir de estos datos, la probabilidad de salir un cuatro.

Se dispone de una muestra aleatoria simple, $(X_1, \ldots, X_{100})$, de la variable:

X: *Número de "4" en diez lanzamientos independientes* $\longrightarrow B(10, p)$

$$\widehat{p}(x_1, \ldots, x_n) = \frac{\overline{x}}{10}.$$

$$\overline{x} = \frac{(0 \times 84) + (1 \times 15) + (2 \times 1) + (3 \times 0)}{100} = 0.17.$$

$$\Downarrow$$

$$\widehat{p}(x_1, \ldots, x_{100}) = \frac{\overline{x}}{10} = 0.017.$$

Problema 5

Se lanza un dado hasta que salga un 4 y se anota el número de lanzamientos necesarios; este experimento se efectúa veinte veces de forma independiente. A partir de los resultados obtenidos, estimar la probabilidad de sacar un 4 por máxima verosimilitud.

$(X_1, \ldots, X_{20})$ m.a.s. de X : *Número de lanzamientos hasta que sale un 4.*

$p =$ *probabilidad de salir un 4 en el dado* $\longrightarrow p \in (0, 1]$ (si $p = 0$, nunca acabaría el experimento).

$* \; p < 1 \longrightarrow P_p(X = x) = p(1-p)^{x-1}, \quad x \in \mathbb{N} \quad (X - 1 \to G(p)).$

$* \; p = 1 \longrightarrow P_{p=1}(X = 1) = 1 \to$ responde a la forma anterior, conviniendo que $0^0 = 1$.

$$P_p(X_1 = x_1, \ldots, X_{20} = x_{20}) = p^{20}(1-p)^{\sum\limits_{i=1}^{20} x_i - 20}, \quad x_1, \ldots, x_{20} \in \mathbb{N}, \quad p \in (0, 1].$$

$$* \; L_{x_1,\ldots,x_{20}}(p) = p^{20}(1-p)^{\sum_{i=1}^{20} x_i - 20}, \quad 0 < p \leq 1.$$

$$* \; \ln L_{x_1,\ldots,x_{20}}(p) = 20 \ln p + \Big(\sum_{i=1}^{20} x_i - 20\Big) \ln(1-p).$$

$$* \; \frac{\partial \ln L_{x_1,\ldots,x_{20}}(p)}{\partial p} = \frac{20}{p} - \frac{\sum_{i=1}^{20} x_i - 20}{1-p} = 0 \;\Rightarrow\; \widehat{p}(x_1,\ldots,x_{20}) = 1/\overline{x} \in (0,1].$$

$$\Downarrow$$

$$\widehat{p}(X_1,\ldots,X_{20}) = \frac{1}{\overline{X}}.$$

■

6.- $n=20$
Pudo arrancar su tractor en el intento $\{1,3,5,1,2,1,3,7,2,4,4,8,1,3,6,5,2,1,6,2\}$. Siguiendo constante p la probabilidad de arrancar en cada intento, ver la estimación más verosímil de la prob. de que el tractor arranque en el 2º intento.

Es una distribución geométrica como en el caso anterior $\Rightarrow$ el estimador máximo verosímil de p es $\frac{1}{\overline{X}}$. La probabilidad de arrancar en el segundo intento es función de p y vale:

$$g(p) = P_p(X=2) = p(1-p)$$

$$\Rightarrow \widehat{g(p)}(X_1 \ldots X_n) = g(\hat{p}(X_1 \ldots X_n)) = \frac{1}{\overline{X}}\left(1 - \frac{1}{\overline{X}}\right)$$

Es una distribución geométrica como en el caso anterior $\Rightarrow$ el estimador máximo verosímil de p es $\frac{1}{\overline{X}}$. La probabilidad de acertar en el segundo intento es función de p y vale:

$$g(p) = P_p(X=2) = p(1-p)$$

$$\Rightarrow \widehat{g(p)}(X_1 \dots X_n) = g(\hat{p}(X_1 \dots X_n)) = \frac{1}{\overline{X}}\left(1-\frac{1}{\overline{X}}\right)$$

$$\widehat{g(p)}(x_1 \dots x_{20}) = \frac{1}{\bar{x}}\left(1-\frac{1}{\bar{x}}\right) \qquad \bar{x} = 3'35$$

$$\Rightarrow \widehat{g(p)}(x_1 \dots x_{20}) = 0'2094$$

Problema 7

Una variable aleatoria discreta toma los valores 0, 1 y 2 con las siguientes probabilidades

$$P_p(X=0) = p^2, \quad P_p(X=1) = 2p(1-p), \quad P_p(X=2) = (1-p)^2,$$

siendo p un parámetro desconocido. En una muestra aleatoria simple de tamaño 100, se ha presentado 22 veces el 0, 53 veces el 1 y 25 veces el 2. Calcular la función de verosimilitud asociada a dicha muestra y dar la estimación más verosímil de p.

La información disponible es el número de veces que aparece cada valor de X en la muestra aleatoria simple de tamaño 100. Por tanto, la estimación debe estar basada en esta información; esto es, en los valores de las variables:

N_i : *Número de observaciones muestrales con el valor i*, $i = 0, 1, 2$.

Cada observación de la muestra, independientemente de las demás, será 0, 1 ó 2. Por tanto, dos cualesquiera de las variables N_0, N_1, N_2 forman un vector aleatorio con distribución multinomial y se tiene:

$$P_p\left(N_0 = 22, N_1 = 53, N_2 = 25\right) = \frac{100!}{22!\ 53!\ 25!}\left(p^2\right)^{22}\left(2p(1-p)\right)^{53}\left((1-p)^2\right)^{25}.$$

Notemos también que el espacio paramétrico es $(0,1)$ $(p \neq 0,1)$ ya que los tres valores de la variable se dan en la muestra observada. Así, la verosimilitud asociada a los valores observados, $n_0 = 22,\ n_1 = 53,\ n_2 = 25$ es

$$L_{22,\ 53,\ 25}(p) = \frac{100! \times 2^{53}}{22!\ 53!\ 25!} p^{97}(1-p)^{103}, \quad 0 < p < 1.$$

$*$ $\ln L_{22,\ 53,\ 25}(p) = \ln \dfrac{100! \times 2^{53}}{22!\ 53!\ 25!} + 97 \ln p + 103 \ln(1-p).$

$*$ $\dfrac{d \ln L_{22,\ 53,\ 25}(p)}{dp} = \dfrac{97}{p} - \dfrac{103}{1-p} = \dfrac{97 - 200p}{p(1-p)} = 0.$

$$\Downarrow$$

$$\widehat{p}(22, 53, 25) = \frac{97}{200} = 0.485.$$

Problema 8

En un muestreo aleatorio simple de una variable aleatoria con distribución $\mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$, se observa que no se obtiene un valor menor que -1 hasta la quinta observación. Dar una estimación máximo verosímil de μ.

Notando $X \to \mathcal{N}(\mu, 1)$, la variable observada es:

Y : *Número de observaciones de X hasta obtener una menor que -1.*

Por tanto, si $p = P_\mu(X < -1)$:

$$P_p(Y = y) = (1-p)^{y-1}p, \quad y \in \mathbb{N}.$$

Disponemos de una observación de la variable Y. A partir de ella, podemos estimar el parámetro p, y como p depende de μ y viceversa, aplicar el *Teorema de Invarianza* para estimar μ.

Estimación máximo verosímil de p: Como en probó en el Problema 5, el estimador máximo verosímil de p basado en una muestra de tamaño n de Y es $1/\overline{Y}$. En nuestro caso la muestra es de tamaño 1 y el valor observado es 5, de forma que la estimación de p es $\widehat{p} = 1/5$.

Estimación máximo verosímil de μ:

Notando Φ a la función de distribución de la $\mathcal{N}(0,1)$:

$$p = P_\mu(X < -1) = P\left(X - \mu < -1 - \mu\right) = \Phi(-1-\mu)$$

$$\Rightarrow -1 - \mu = \Phi^{-1}(p) \Rightarrow \ \mu = -1 - \Phi^{-1}(p).$$

$$\Downarrow \ \textit{T. Invarianza}$$

$$\widehat{\mu} = -1 - \Phi^{-1}(1/5).$$

9. Sea X = tiempo de vida de un filamento $\sim exp(\theta)$. Se toma $\underline{X}$ m.a.s. de tam n y se ven los filamentos quemados en tiempo T. Dar el EMV de $E_\theta[X]$ a partir de la info de $\underline{X}$.

$X \sim exp(\theta) \Rightarrow E_\theta[X] = 1/\theta$. Voy a buscar $\hat{\theta}(T)$ dato que me dan.

$P_\theta[X \leq T] = \int_0^T \theta e^{-\theta x}\,dx = 1 - e^{-\theta T}, \ \forall \theta \in \mathbb{R}^+$

tamaño de la muestra

prob. de ser quemado antes de T

Defino Y = nº de filamentos quemados en tiempo $T \sim B(n, 1-e^{-\theta T})$

La función de verosimilitud asoc. a la obs. del nº de filamentos quemados en tiempo $\leq T$, llamemos a ese nº k, es: $L_k(\theta) =$

$P_\theta[Y = k] = \binom{n}{k}(1-e^{-\theta T})^k (e^{-\theta T})^{n-k}, \ \forall \theta$

$\frac{\delta}{\delta\theta}[\ln L_k(\theta)] = \frac{\delta}{\delta\theta}\left[\ln\binom{n}{k} + k\ln(1-e^{-\theta T}) + \overbrace{(n-k)\ln e^{-\theta T}}^{-(n-k)\theta T}\right] =$

$k \cdot \frac{\theta T e^{-\theta T}}{\ldots} - (n-k)\theta T = \frac{k\theta T e^{-\theta T} - n\theta T + k\theta T + n\theta T e^{-\theta T} - k\theta T e^{-\theta T}}{\ldots} = 0$

$$\frac{\delta}{\delta\theta}\left[\ln L_k(\theta)\right] = \frac{\delta}{\delta\theta}\left[\ln\binom{n}{k} + k\ln(1-e^{-\theta T}) + \overbrace{(n-k)\ln e^{-\theta T}}\right] =$$

$$k\cdot\frac{\theta T e^{-\theta T}}{1-e^{-\theta T}} - (n-k)\theta T = \frac{k\theta T e^{-\theta T} - n\theta T + k\theta T + n\theta T e^{-\theta T} - k\theta T e^{-\theta T}}{1-e^{-\theta T}} = 0$$

$$\Rightarrow \theta T(-n+k+ne^{-\theta T}) = 0 \Rightarrow e^{-\theta T} = \frac{n-k}{n} \Rightarrow \ln e^{-\theta T} =$$

$$\ln\frac{n-k}{n} \Rightarrow \theta = -\frac{\ln(n-k)-\ln(n)}{T} \Rightarrow \hat{\theta}(k) = \frac{\ln(n-k)-\ln(n)}{T}$$

Por el T. Invarianza: $\hat{E}_\theta[X] = \dfrac{T}{\ln(n-k)-\ln(n)}$.

Problema 9

En la producción de filamentos eléctricos la medida de interés, X, es el tiempo de vida de cada filamento, que tiene una distribución exponencial de parámetro θ. Se eligen n de tales filamentos de forma aleatoria e independiente, pero, por razones de economía, no conviene esperar a que todos se quemen y la observación acaba en el tiempo T. Dar el estimador máximo verosímil para la media de X a partir del número de filamentos quemados durante el tiempo de observación.

$$f_\theta(x) = \theta e^{-x\theta}, \quad x>0, \quad (\theta>0)$$

Se pretende estimar $E_\theta[X] = 1/\theta$ para lo cual basta encontrar el estimador máximo verosímil de θ y aplicar el *Teorema de Invarianza*.

Para la estimación, se toman n filamentos de forma independiente y se observa cuántos se queman antes del tiempo T. Por lo tanto, la variable observada es

Y : *Número de filamentos quemados hasta el tiempo* $T \rightarrow \{B(n,p);\ p\in(0,1)\}$

$$p = P_\theta(X<T) = 1-e^{-\theta T}.$$

Estimación máximo verosímil de p: Como en probó en el Problema 4, el estimador máximo verosímil de p basado en una muestra de $Y \to B(n,p)$ es $\overline{Y}/n$. En nuestro caso, tenemos una sola observación de Y y, por tanto, $\widehat{p}(Y) = Y/n$.

Estimación máximo verosímil de θ:

$$e^{-\theta T} = 1 - p \ \Rightarrow\ -\theta T = \ln(1-p) \ \Rightarrow\ \theta = -\frac{\ln(1-p)}{T}.$$

$$\Downarrow \ \textit{T. Invarianza}$$

$$\widehat{\theta}(Y) = -\frac{\ln(1 - Y/n)}{T}.$$

$$\Downarrow \ \textit{T. Invarianza}$$

$$\widehat{E_\theta[X]}(Y) = -\frac{T}{\ln(1 - Y/n)}.$$

Nota:

* $Y = n$ daría como estimación de la media el valor 0, que no tiene sentido. La conclusión razonable en ese caso sería decir que el tiempo de vida medio es menor que T.

* $Y = 0$ daría como estimación de la media el valor $+\infty$, que no tiene sentido. Lo razonable sería concluir que el tiempo de vida medio es mayor que T.

⑩ Sean $X_1, \dots, X_n$ observaciones independientes de una variable $X \sim \Gamma(p,a)$; $p, a > 0$. Estimar ambos parámetros mediante el método de los momentos.

$f_{p,a}(x) = \dfrac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}$, $x>0$ Y sabemos que $E[X] = \dfrac{p}{a}$, $Var[X] = \dfrac{p}{a^2}$

$$\left.\begin{array}{l} E[X] = m_1 \\ Var[X] = m_2 - m_1^2 \end{array}\right\} \overset{\text{Mét.}}{\underset{\text{mom.}}{\Longrightarrow}} \left.\begin{array}{l} A_1 = \frac{p}{a} \\ A_2 - A_1^2 = \frac{p}{a^2} \end{array}\right\} \overset{a>0}{\Longleftrightarrow} \left.\begin{array}{l} p = a A_1 \\ p = a^2 (A_2 - A_1^2) \end{array}\right\} \Longleftrightarrow a A_1 = a^2 (A_2 - A_1^2) \Longleftrightarrow a = \frac{A_1}{A_2 - A_1^2}$$

$$p = \frac{A_1^2}{A_2 - A_1^2}$$

luego $\hat{a}(X_1, \dots, X_n) = \dfrac{\sum_{i=1}^n X_i}{\sum_{i=1}^n (X_i - \overline{X})^2}$, $\hat{p}(X_1, \dots, X_n) = \dfrac{(\sum_{i=1}^n X_i)^2}{n \sum_{i=1}^n (X_i - \overline{X})^2}$

Aplicación. Neumáticos con vidas útiles de 35200, 41000, 44700, 38600 y 41500 Kilómetros.

Suponiendo que estos datos son observaciones independientes de una v. a. con dist. exp. de parámetro θ, dar una estimación de dicho parámetro por el método de los momentos.

$X \sim \exp(\theta) \underset{\theta>0}{\iff} f_\theta(x) = \theta e^{-\theta x}, \quad x \geq 0. \quad E[X] = \frac{1}{\theta}.$

Mét. momentos: $\bar{X} = \frac{1}{\theta} \Rightarrow \theta = \frac{1}{\bar{X}}, \quad \hat{\theta}(X_1, \dots, X_n) = \frac{1}{\bar{X}}$

Para la realización muestral dada: $\bar{x} = \frac{35200+41000+44700+38600+41500}{5} = 40200$

Por tanto $\hat{\theta}(35200, \dots, 41500) = \frac{1}{40200} \simeq 2'4875 \cdot 10^{-5}$