Tema 3: Estimadores de máxima verosimilitud

(basado en el material de A. Jach (http://www.est.uc3m.es/ajach/)
y A. Alonso (http://www.est.uc3m.es/amalonso/))

- Planteamiento del problema: motivación
- Método de obtención del E.M.V.
- Propiedades del E.M.V.

Planteamiento del problema

Hasta ahora hemos visto:

- Población y muestra
- Estadísticos y parámetros
- Estimadores y sus propiedades

¿Cómo conseguir estimadores con buenas propiedades?

- A veces tenemos un estimador natural: $\overline{X}$ para μ, S^2 para σ^2, etc.
- A veces podemos deducirlo de forma intuitiva: $\frac{n+1}{n}\max\{X_1, \ldots, X_n\}$ para **b**, siendo la distribución de interés $U(0, b)$ (Ejemplo 1, Tema 2).
- ¿Y en el resto de los casos?

Planteamiento del problema

Ejemplo 1.

En una urna hay 4 bolas que pueden ser blancas o negras. La proporción, θ, de bolas blancas en la urna es desconocida.
Nuestro objetivo es estimar el valor de θ. Para ello, extraemos de la urna 2 bolas **con reemplazamiento**. *Supongamos que la primera bola extraída es blanca (B) y la segunda es negra (N).*

¿Qué valor de θ te resulta más verosímil?

verosímil.

1. adj. Que tiene apariencia de verdadero.
2. adj. Creíble por no ofrecer carácter alguno de falsedad.

`http://www.rae.es`

Planteamiento del problema

Ejemplo 1 (cont.) Buscaremos el valor de θ (entre todos los posibles) que haga más verosímil (más probable) el resultado que hemos obtenido.

Para ello calcularemos

$$P(B, N|\theta)$$

y elegiremos el valor de θ que nos de una probabilidad mayor.

Es decir, buscamos el valor de θ que maximiza la función de probabilidad conjunta $P(B, N|\theta)$, también llamada verosimilitud.

La estimación así obtenida se llama estimación de máxima verosimilitud (EMV): usamos la información disponible en la muestra para elegir el valor del parámetro para el cuál es más probable haber obsservado ese resultado muestral.

Verosimilitud

Definición 1.

Sea $X_1, \ldots, X_n$ una muestra aleatoria (no necesariamente simple) de una población X con función de probabilidad P_θ (o con función de densidad f_θ). Para cada muestra particular $(x_1, \ldots, x_n)$, la función de verosimilitud se define como la función de probabilidad (o de densidad) conjunta de $(X_1, \ldots, X_n)$ evaluada en $(x_1, \ldots, x_n)$.

$$\mathcal{L}(\theta) = \mathcal{L}(x_1, \ldots, x_n; \theta) = P_\theta(X_1 = x_1, \ldots, X_n = x_n) \quad \text{si } X \text{ es discreta}$$

$$\mathcal{L}(\theta) = \mathcal{L}(x_1, \ldots, x_n; \theta) = f_\theta(x_1, \ldots, x_n) \quad \text{si } X \text{ es continua}$$

Observaciones:

- La notación $\mathcal{L}(\theta)$ indica que $\mathcal{L}$ es una función de θ y no de $(x_1, \ldots, x_n)$.
- θ puede ser un escalar o un vector ($\theta = (\theta_1, \ldots, \theta_k)$).
- El subíndice θ en la función de probabilidad o de densidad indica que dicha función depende del valor del parámetro. Utilizaremos también las notaciones $P(\cdot\,; \theta)$ y $f(\cdot\,; \theta)$.

Verosimilitud

Ejemplo 1 (cont.) Escribir la función de verosimilitud en el caso de extraer dos bolas de la urna **con reemplazamiento**.

Al extraer una bola al azar

$$X = \begin{cases} 1 & \text{si es blanca} \\ 0 & \text{si es negra} \end{cases} \quad \sim B(\theta)$$

Tenemos una m.a.s. de tamaño 2, es decir, $X_1, X_2 \sim B(\theta)$ i.i.d.

Para una muestra particular cualquiera (x_1, x_2):

$$\begin{aligned} \mathcal{L}(x_1, x_2; \theta) &= P(X_1 = x_1, X_2 = x_2; \theta) \overset{indep.}{=} P(X_1 = x_1; \theta) \cdot P(X_2 = x_2; \theta) \\ &= \left[\theta^{x_1} (1-\theta)^{1-x_1}\right] \cdot \left[\theta^{x_2} (1-\theta)^{1-x_2}\right] \end{aligned}$$

Si realizamos la extracción **sin reemplazamiento**, ¿cuál sería la función de verosimilitud?

Verosimilitud

El caso más frecuente (y el que váis a tener en el examen) es aquél en el que las v.a. son i.i.d., es decir, tenemos una muestra aleatoria simple.

- X v.a. discreta $\sim P(\cdot\,;\theta)$. $X_1,\ldots,X_n$ m.a.s. de X:

$$\begin{aligned}\mathcal{L}(x_1,\ldots,x_n;\theta) &= P(X_1=x_1\,,\ldots,\,X_n=x_n;\theta)\\ &\overset{indep.}{=} P(X_1=x_1;\theta)\cdot\ldots\cdot P(X_n=x_n;\theta)\\ &\overset{i.d.}{=} \prod_{i=1}^{n} P(X=x_i;\theta)\end{aligned}$$

- X v.a. continua $\sim f(\cdot\,;\theta)$. $X_1,\ldots,X_n$ m.a.s. de X:

$$\begin{aligned}\mathcal{L}(x_1,\ldots,x_n;\theta) &= f(x_1,\ldots,x_n\,;\theta)\\ &\overset{indep.}{=} f_{X_1}(x_1\,;\theta)\cdot\ldots\cdot f_{X_n}(x_n\,;\theta)\\ &\overset{i.d.}{=} \prod_{i=1}^{n} f(x_i\,;\theta)\end{aligned}$$

Verosimilitud

Ejemplo 2.

Sea $X_1, \ldots, X_n$ *una m.a.s. de* $X \sim N(\mu, \sigma)$.

Función de densidad de X: $\quad f(x\,;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$

Función de verosimilitud de $\theta = (\mu, \sigma)$:

$$\begin{aligned}
\mathcal{L}(x_1,\ldots,x_n\,;\mu,\sigma) &= f(x_1,\ldots,x_n\,;\theta) \overset{indep.}{=} \prod_{i=1}^{n} f_{X_i}(x_i\,;\mu,\sigma) \overset{i.d.}{=} \prod_{i=1}^{n} f(x_i\,;\mu,\sigma) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}
\end{aligned}$$

$$x_1,\ldots,x_n \in \mathbb{R}$$

Estimador de máxima verosimilitud

Definición 2.

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población X con función de verosimilitud $\mathcal{L}(\theta)$. Para cada muestra particular $(x_1, \ldots, x_n)$, la estimación de máxima verosimilitud de θ es el valor $\hat{\theta}_{MV}$ que maximiza la verosimilitud. Es decir:

$$\mathcal{L}(x_1, \ldots, x_n; \hat{\theta}_{MV}) = \max_{\theta} \mathcal{L}(x_1, \ldots, x_n; \theta)$$

El estimador de máxima verosimilitud, $\hat{\theta}_{MV}(X_1, \ldots, X_n)$, es aquél que evaluado en cada muestra particular nos da la estimación de máxima verosimilitud ($\hat{\theta}_{MV}(x_1, \ldots, x_n)$).

Observaciones:

- Muchas veces se utiliza el término estimador de máxima verosimilitud para denotar tanto el estimador como cualquier estimación particular.
- Ambos se denotan por $\hat{\theta}_{MV}$ y se abrevian por E.M.V.

Estimador de máxima verosimilitud

Sea $\theta = (\theta_1, \dots, \theta_k)$. Procedimiento para obtener el E.M.V. de θ_j dada una muestra particular $(x_1, \dots, x_n)$:

1. Escribir la verosimilitud: $\mathcal{L}(\theta) = \mathcal{L}(x_1, \dots, x_n; \theta)$

2. Escribir el logaritmo de la verosimilitud: $\ell(\theta) = \ln \mathcal{L}(\theta)$
   $\ell(\theta)$ se llama función soporte o log-verosimilitud.

3. Obtener el θ_j tal que
$$\frac{\partial}{\partial \theta_j} \ell(\theta) = 0$$
   y denotarlo por $\hat{\theta}_j$.

4. Comprobar que realmente es un máximo, es decir, que
$$\frac{\partial^2}{\partial \theta_j^2} \ell(\theta)|_{\theta_j = \hat{\theta}_j} < 0$$

Estimador de máxima verosimilitud

Ejemplo 3.

Sea $X_1, \ldots, X_n$ *una m.a.s. de* $X \sim N(\mu, \sigma)$. *Obtener el E.M.V. de* μ.

1. *Escribir la verosimilitud (Ejemplo 2):*

$$\mathcal{L}(\mu) = \left(\frac{1}{\sqrt{2\pi}\,\sigma} \right)^n exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

2. *Escribir la función soporte:*

$$\ell(\mu) = \ln \mathcal{L}(\mu) = -n \ln \left(\sqrt{2\pi}\,\sigma \right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Estimador de máxima verosimilitud

Ejemplo 3 (cont.)

3. Obtener el valor de μ tal que $\frac{\partial}{\partial\mu}\ell(\mu) = 0$.

$$\frac{\partial}{\partial\mu}\ell(\mu) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n} -2\,(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\left(-n\mu + \sum_{i=1}^{n} x_i\right)$$

$$\begin{aligned} \frac{\partial}{\partial\mu}\ell(\mu) = 0 \quad &\Leftrightarrow \quad \frac{1}{\sigma^2}\left(-n\mu + \sum_{i=1}^{n} x_i\right) = 0 \Leftrightarrow -n\mu + \sum_{i=1}^{n} x_i = 0 \\ &\Leftrightarrow \quad n\mu = \sum_{i=1}^{n} x_i \Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \end{aligned}$$

Por tanto $\hat{\mu} = \overline{x}$.

4. Comprobar que realmente es un máximo, es decir, que $\frac{\partial^2}{\partial\mu^2}\ell(\mu)|_{\mu=\hat{\mu}} < 0$

$$\frac{\partial^2}{\partial\mu^2}\ell(\mu) = \frac{-n}{\sigma^2} < 0 \quad \forall\mu \quad \Rightarrow \quad \frac{\partial^2}{\partial\mu^2}\ell(\mu)|_{\mu=\hat{\mu}} < 0$$

Por tanto la estimación máximo verosímil para una muestra dada es $\hat{\mu}_{MV} = \overline{x}$, y el estimador máximo verosímil es $\hat{\mu}_{MV} = \overline{X}$.

Estimador de máxima verosimilitud

Ejemplo 4.

Sea $X_1, \ldots, X_n$ una m.a.s. de X. Obtener:

- *El E.M.V. de σ^2 si $X \sim N(\mu, \sigma)$ con μ conocida.*
- *El E.M.V. de (μ, σ^2) si $X \sim N(\mu, \sigma)$.*
- *El E.M.V. de p si $X \sim B(p)$.*
- *El E.M.V. de λ si $X \sim \mathcal{E}xp(\lambda)$.*
- *El E.M.V. de λ si $X \sim \mathcal{P}(\lambda)$.*

Propiedades del E.M.V.

Los E.M.V. tienen muy buenas propiedades.

1. INVARIANZA:
   Si $\hat{\theta}_{MV}$ es el estimador máximo verosímil de θ, entonces $h(\hat{\theta}_{MV})$ es el estimador máximo verosímil de $h(\theta)$.

Ejemplo 5.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$. Sabemos que $\hat{\mu}_{MV} = \overline{X}$.
¿Quiénes son los E.M.V. de 3μ, μ^2 y $1/\mu$?
Por el principio de invarianza tenemos que $\widehat{3\mu}_{MV} = 3\overline{X}$, $\widehat{\mu^2}_{MV} = \overline{X}^2$ y $\widehat{1/\mu}_{MV} = 1/\overline{X}$.

2. CONSISTENCIA:
   Bajo ciertas condiciones generales, $\hat{\theta}_{MV}$ es un estimador consistente de θ.

3. INSESGADEZ ASINTÓTICA:
   Se verifica que $\lim_{n\to\infty} E[\hat{\theta}_{nMV}] = \theta$.

Propiedades del E.M.V.

4. NORMALIDAD ASINTÓTICA:

Bajo ciertas condiciones generales,

$$\sqrt{n}(\hat{\theta}_{MV} - \theta) \overset{A}{\sim} N(0, \sqrt{i(\theta)^{-1}})$$

donde

$$i(\theta) = E\left[\left(\frac{\partial}{\partial\theta} \ln f(X\,;\theta)\right)^2\right]$$

es la cantidad de información de Fisher correspondiente a una observación.

La cantidad de información de Fisher correspondiente a n observaciones es

$$\begin{aligned} I(\theta) &= E\left[\left(\frac{\partial}{\partial\theta} \ln f(X_1,\ldots,X_n\,;\theta)\right)^2\right] \overset{m.a.s.}{=} n\cdot E\left[\left(\frac{\partial}{\partial\theta} \ln f(X\,;\theta)\right)^2\right] \\ &= n\cdot i(\theta) \end{aligned}$$

Propiedades del E.M.V.

4. NORMALIDAD ASINTÓTICA (cont.):

Se tiene que

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln f(X_1,\ldots,X_n\,;\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X_1,\ldots,X_n\,;\theta)\right]$$

La varianza asintótica de $\hat{\theta}_{MV}$ es:

$$\begin{aligned} Var\left[\hat{\theta}_{MV}\right] \overset{A}{=} \frac{1}{n\cdot i(\theta)} = \frac{1}{I(\theta)} &= -\frac{1}{E\left[\frac{\partial^2}{\partial\theta^2}\ln f(X_1,\ldots,X_n\,;\theta)\right]} \\ &\approx -\frac{1}{\frac{\partial^2}{\partial\theta^2}\ell(\theta)\big|_{\theta=\hat{\theta}_{MV}}} \end{aligned}$$

La aproximación final es muy útil en la práctica.

Propiedades del E.M.V.

Ejemplo 6.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$. Obtener la varianza asintótica de $\hat{\mu}_{MV}$.

Sabemos que $$Var\left[\hat{\mu}_{MV}\right] \overset{A}{\approx} -\frac{1}{\frac{\partial^2}{\partial \mu^2}\ell(\mu)|_{\mu=\hat{\mu}_{MV}}} \quad y \quad \frac{\partial^2}{\partial \mu^2}\ell(\mu) = \frac{-n}{\sigma^2}.$$

Por tanto $$Var\left[\hat{\mu}_{MV}\right] \overset{A}{\approx} \frac{\sigma^2}{n}.$$

Observación: en este caso, al tratarse del estadístico media muestral sabemos que su varianza es exactamente $\frac{\sigma^2}{n}$ y no sólo asintóticamente.

Además, al tratarse de la distribución normal, sabemos que la distribución del estadístico media muestral es exactamente normal y no sólo asintóticamente.

Propiedades del E.M.V.

Ejemplo 7.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim \mathcal{E}xp(\lambda)$. Hemos visto, que en este caso, el **estimador** *máximo-verosímil de λ es $\hat{\lambda}_{MV} = \frac{1}{\bar{X}}$.*
¿Cuál su varianza asintótica?

$$Var\left[\hat{\lambda}_{MV}\right] \overset{A}{\approx} -\frac{1}{\frac{\partial^2}{\partial\lambda^2}\ell(\lambda)\big|_{\lambda=\hat{\lambda}_{MV}}} = -\frac{1}{-\frac{n}{\lambda^2}\big|_{\lambda=\hat{\lambda}_{MV}}} = \frac{\hat{\lambda}_{MV}^2}{n}$$

Si observamos la muestra particular (0.37, 2.16, 0.07, 1.86, 7.11, 0.67, 0.49, 0.11), ¿cuál es la **estimación** *máximo-verosímil de λ para esta muestra? ¿Y la aproximación de la varianza asintótica del* **estimador** *máximo-verosímil?*

Estimación máximo-verosímil: $\hat{\lambda}_{MV} = \frac{1}{\bar{x}} = \frac{1}{1.605} \approx 0.623.$

Aproximación de la varianza asintótica de $\hat{\lambda}_{MV}$ (estimador):

$$Var\left[\hat{\lambda}_{MV}\right] \overset{A}{\approx} \frac{\hat{\lambda}_{MV}^2}{n} = \frac{0.623^2}{8} = 0.049$$