Modelos Estadísticos para la Regresión y la Clasificación

Continuación Práctico 3 - Estimación de máxima verosimilitud

Micaela Long

Instituto de Matemática y Estadística Prof. Rafael Laguardia (IMERL)
Facultad de Ingeniería, Universidad de la República, Uruguay

6 de setiembre de 2024

Práctico 23/8:

- Práctico 2 Ejercicio 3: Descomposición de la varianza.
- Repaso teórico estimación
- Práctico 3 Ejercicio 1
- Práctico 3 Ejercicio 2
- Práctico 3 Ejercicio 3 (calculamos estimador de momentos)

Ejercicio 3: Momentos vs. Máxima Verosimilitud

Sea $X_1, \ldots, X_n$ un muestreo i.i.d. de una distribución normal $N(\theta, \theta)$.

1. Calcular los estimadores de momentos $\hat{\theta}_M$ y de máxima verosimilitud $\hat{\theta}_{MLE}$ de θ
2. Nos enfocamos en esta parte en el estimador de máxima verosimilitud.

   a) Probar que $\hat{\theta}_{MLE}$ es sesgado pero asíntóticamente insesgado

   b) Probar que $\mathrm{Var}(\hat{\theta}_{MLE})$ alcanza la cota de Cramér-Rao.

Estimador de momentos

- **Momento**: esperanza de una potencia de una variable aleatoria.
- Momento de primer orden es $\mathbb{E}(X)$, el momento de segundo orden es $\mathbb{E}(X^2)$, etc.
- Estimador de momentos se basa en la idea de hacer coincidir los **momentos muestrales** (medidas estadísticas calculadas a partir de los datos) con **los momentos poblacionales** (valores esperados derivados de la distribución teórica) para encontrar el parámetro desconocido.

$$\mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- En nuestro ejercicio $\theta = \mathbb{E}(X)$, por lo que el estimador de momentos es:

$$\hat{\theta}_M = \frac{1}{n}\sum_{i=1}^{n} X_i$$

En otros ejemplos podríamos necesitar momentos de mayor orden.

Ejercicio 3
Parte 1

Queremos encontrar el estimador MLE:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \mathcal{L}(\theta) = \arg\max_{\theta} \log\left(\mathcal{L}(\theta)\right)$$

Para todo i se tiene $X_i \sim \mathcal{N}(\theta, \theta)$, luego

$$f_\theta(x_i) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(x_i-\theta)^2}{\theta}}$$

La función de verosimilitud es:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(x_i-\theta)^2}{\theta}}$$

Vamos a maximizar la log verosimilitud

$$l(\theta) = \log\left(\mathcal{L}(\theta)\right) = \log\left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(x_i-\theta)^2}{\theta}}\right] = -\frac{n}{2}\log(2\pi\theta) + \sum \frac{-\frac{1}{2}(X_i-\theta)^2}{\theta}$$

Ejercicio 3
Parte 1

Para hallar el máximo derivamos $l(\theta)$ e igualamos a 0

$$0 = \frac{\partial l(\theta)}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} X_i^2 - \frac{n}{2}$$

Esto se cumple si y solo si

$$\theta^2 + \theta - \sum_{i=1}^{n} \frac{X_i^2}{n} = 0$$

Luego, como $\theta = \mathrm{Var}(X_i) > 0$

$$\hat{\theta}_{MLE} = -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{X_i^2}{n}}$$

Ejercicio 3
Parte 2

Queremos ver que el estimador es:

1. Sesgado
2. Asintóticamente insesgado
3. Asintóticamente eficiente (alcanza la cota de CR en el límite)

1) Para ver que es sesgado, calculamos $\mathbb{E}(\hat{\theta}_{MLE})$

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_{MLE}) = \mathbb{E}\left(-\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{X_i^2}{n}}\right) &\overset{(1)}{<} -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{\mathbb{E}(X_i^2)}{n}} \\
&\overset{(2)}{=} -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{\theta + \theta^2}{n}} \\
&= \theta
\end{aligned}$$

donde en:

(1) usamos que $\sqrt{x}$ es estrictamente cóncava, y que vale la **desigualdad de Jensen**: si f es estrictamente cóncava entonces $\mathbb{E}(f(X)) < f(\mathbb{E}(X))$

(2) $\mathbb{E}(X_i^2) = \mathrm{Var}(X_i) + \mathbb{E}(X_i)^2 = \theta + \theta^2$.

Luego $\mathbb{E}(\hat{\theta}_{MLE}) \neq \theta$, por lo que el estimador es sesgado.

Ejercicio 3
Parte 2

2) Para ver que es asintóticamente insesgado, tenemos que considerar $\lim_{n\to\infty} \mathbb{E}(\hat{\theta}_{MLE})$.

$$\hat{\theta}_{MLE} = -\frac{1}{2} + \sqrt{\frac{1}{4} + \sum_{i=1}^{n} \frac{X_i^2}{n}} = -\frac{1}{2} + \sqrt{\frac{1}{4} + u} = g(u)$$

donde $u = \sum_{i=1}^{n} \frac{X_i^2}{n}$.

Notar que por la Ley (fuerte) de los Grandes Números (LGN), en un conjunto de probabilidad 1 vale

$$\begin{aligned} \lim_{n\to\infty} u = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{X_i^2}{n} &= \mathbb{E}(X_i^2) \\ &= \mathrm{Var}(X_i) + \mathbb{E}(X_i)^2 \\ &= \theta + \theta^2 \end{aligned}$$

Si definimos $u_0 = \theta + \theta^2$, podemos aproximar g por

$$g(u) \approx g(u_0) + g'(u_0)(u - u_0)$$

donde $g'(u) = \frac{1}{2\sqrt{\frac{1}{4}+u}}$.

Notar que estamos usando dos resultados importantes: el desarrollo de Taylor de orden 1, y la LGN.

Ejercicio 3
Parte 2

Es decir

$$
\begin{aligned}
\hat{\theta}_{MLE} = g(u) &\approx g(u_0) + g'(u_0)(u - u_0) \\
&= -\frac{1}{2} + \sqrt{\frac{1}{4} + u_0} + \frac{1}{2\sqrt{\frac{1}{4} + u_0}}(u - u_0) \\
&= -\frac{1}{2} + \sqrt{\frac{1}{4} + \theta + \theta^2} + \frac{1}{2\sqrt{\frac{1}{4} + \theta + \theta^2}}\left(\sum_{i=1}^{n} \frac{X_i^2}{n} - (\theta + \theta^2)\right) \\
&= \theta + \frac{1/2}{\theta + 1/2}\left(\sum_{i=1}^{n} \frac{X_i^2}{n} - (\theta + \theta^2)\right)
\end{aligned}
$$

Luego

$$
\begin{aligned}
\mathbb{E}(\hat{\theta}_{MLE}) \underset{n\to\infty}{\to} \mathbb{E}&\left(\theta + \frac{1/2}{\theta + 1/2}\left(\sum_{i=1}^{n} \frac{X_i^2}{n} - (\theta + \theta^2)\right)\right) \\
&= \theta + \frac{1/2}{\theta + 1/2}\left(\sum_{i=1}^{n} \frac{\mathbb{E}(X_i^2)}{n} - (\theta + \theta^2)\right) \\
&= \theta
\end{aligned}
$$

Entonces $\hat{\theta}_{MLE}$ es asintóticamente insesgado.

Ejercicio 3
Parte 2

3) De la misma calculamos $\mathrm{Var}(\hat{\theta}_{MLE})$

$$
\begin{aligned}
\mathrm{Var}(\hat{\theta}_{MLE}) &\approx \mathrm{Var}\left(\theta + \frac{1/2}{\theta + 1/2}\left(\sum_{i=1}^{n} \frac{X_i^2}{n} - (\theta + \theta^2)\right)\right) \\
&= \left(\frac{1/2}{\theta + 1/2}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} \frac{X_i^2}{n} - (\theta + \theta^2)\right) \\
&= \left(\frac{1/2}{\theta + 1/2}\right)^2 \sum_{i=1}^{n} \frac{\mathrm{Var}(X_i^2)}{n^2} \\
&= \frac{1/4}{n(\theta + 1/2)^2} \mathrm{Var}(X_i^2)
\end{aligned}
$$

Para calcular $\mathrm{Var}(X_i^2)$ usamos la siguiente propiedad:

$$
Z \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow \mathrm{Var}(Z^2) = 4\mu^2\sigma^2 + 2\sigma^4
$$

Luego, como $X_i \sim \mathcal{N}(\theta, \theta)$

$$
\begin{aligned}
\mathrm{Var}(\hat{\theta}_{MLE}) \approx \frac{1/4}{n(\theta + 1/2)^2}(4\theta^3 + 2\theta^2) &= \frac{1/4}{n(\theta + 1/2)^2} 4\theta^2 \left(\theta + \frac{1}{2}\right) \\
&= \frac{\theta^2}{n(\theta + 1/2)}
\end{aligned}
$$

En el límite la aproximación es una igualdad.

Ejercicio 3
Parte 2

Queda probar que la cota de Cramér-Rao es $\frac{\theta^2}{n(\theta+1/2)}$. Recordar que la cota de CR es:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{\mathcal{I}(\theta)} = \frac{1}{-\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial \theta^2}\right)}$$

Antes calculamos

$$\frac{\partial l(\theta)}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n} X_i^2 - \frac{n}{2}$$

Derivando respecto a θ obtenemos

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n} X_i^2$$

Luego

$$\begin{aligned} -\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial \theta^2}\right) &= -\frac{n}{2\theta^2} + \frac{1}{\theta^3}\sum_{i=1}^{n} \mathbb{E}(X_i^2) \\ &= -\frac{n}{2\theta^2} + \frac{n(\theta+\theta^2)}{\theta^3} \\ &= \frac{n(\theta+1/2)}{\theta^2} \end{aligned}$$

Entonces

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{-\mathbb{E}\left(\frac{\partial^2 l(\theta)}{\partial \theta^2}\right)} = \frac{\theta^2}{n(\theta+1/2)} = \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_{MLE})$$

Ejercicio 4: Sesgo de un estimador

Un estimador $\hat{\theta}$ de un parámetro θ tiene distribución normal. Se sabe además que $\text{ECM}(\hat{\theta}) = 8$ y $P(\hat{\theta} \leq \theta) = 0{,}8413$. Hallar el sesgo de $\hat{\theta}$.

Sugerencia: Usar

$$X \sim \mathcal{N}(0,1) \Rightarrow \mathbb{P}(X \leq 1) = 0{,}8413$$

Ejercicio 5

Comparación de estimadores. Ana y Beto saben que el delivery llega en un tiempo uniforme en el intervalo $[0, a]$, en donde 0 es el momento en el que hacen el pedido, y a es un valor desconocido. Se deciden a estimar a, y para esto disponen de un muestreo aleatorio $X_1, \dots, X_n$. Proponen los siguientes estimadores:

$$A_n = \text{máx}\{X_1, \dots, X_n\} \quad \text{y} \quad B_n = 2\bar{X}_n.$$

1. Calcular el sesgo y el error cuadrático medio de B_n. Determinar si B_n es consistente.
2. Demostrar que la densidad de A_n es

$$p_A(x) = \begin{cases} \frac{nx^{n-1}}{a^n} & \text{si } 0 \leq x \leq a, \\ 0 & \text{si no.} \end{cases}$$

3. Calcular el sesgo y el error cuadrático medio de A_n. ¿Es asintóticamente insesgado?

Ejercicio 5

Sugerencias:

1. Para determinar si B_n es consistente, es suficiente probar que $EMC(B_n) \to 0$ con $n \to \infty$
2. Notar que

$$F_{A_n}(x) = \mathbb{P}(max\{X_1, \cdots, X_n\} \leq x) = \mathbb{P}(X_1 \leq x, \cdots, X_n \leq x)$$

   y que las X_i son independientes y $X_i \sim U[0, a]$.

3. Usar la descomposición sesgo varianza del $EMC(A_n)$

Ejercicio 6

Estimador de mínima varianza. Se considera la condición (C):

$$\frac{\partial \ell(\theta, x)}{\partial \theta} = I(\theta)(T(x) - \theta).$$

1. Probar que si un estimador $T(x)$ satisface la condición (C), entonces $I(\theta)$ es la información de Fisher de θ.
2. Probar que si $\hat{\theta}_{\text{MLE}}$ es el estimador de máxima verosimilitud de θ y $T(x)$ cumple con la condición (C), entonces $T(x) = \hat{\theta}_{\text{MLE}}$.

Sugerencia: Para la parte 2, recordar que el estimador de máxima verosimilitud maximiza la función de log verosimilitud (derivada de la log verosimilitud es 0)

Ejercicio 7: El método delta

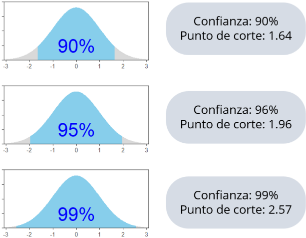
Sea X una variable aleatoria con la siguiente densidad de probabilidad:

$$p(x; a) = (a+1)x^a \quad \text{si } 0 < x < 1; \quad 0 \text{ en otro caso.}$$

Sea $X_1, \ldots, X_n$ un muestreo aleatorio de X.

1. Hallar el estimador de máxima verosimilitud $\hat{a}$ de a.
2. Hallar la densidad de $Y = -\ln(X)$ y calcular $\mu = \mathbb{E}(Y)$ en función de a.
3. Verificar que $\hat{a} = g(Y_n)$, con $g(y) = \frac{1}{y} - 1$, y probar que $\hat{a}$ es consistente.
4. Usando el desarrollo $\hat{a} - a = g(Y_n) - g(\mu) \approx g'(\mu)(Y_n - \mu)$ para n grande:
    1. Probar que $\hat{a}$ es asintóticamente insesgado.
    2. Hallar la varianza asintótica de $\hat{a}$.
    3. ¿Es $\hat{a}$ asintóticamente normal?
5. Se dispone de la siguiente muestra de X:
   0.56, 0.82, 0.71, 0.87, 0.33, 0.36, 0.93, 0.94, 0.89, 0.42.
   Hallar un intervalo de confianza asintótico para a al nivel 0.9.

Ejercicio 7
Parte 5



$$\mathbb{P}(a_1 \leq a \leq a_2) = 0{,}9$$

$$a_1 = \hat{a} - z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{a})}$$

$$a_2 = \hat{a} + z_{\alpha/2}\sqrt{\mathrm{Var}(\hat{a})}$$

Figura: https://fernandoblancopsy.com

- En este caso $z_{\alpha/2} = 1{,}64$
- $\hat{a}, \mathrm{Var}(\hat{a})$ se calculan a partir de los datos.
- $\mathrm{Var}(\hat{a})$ es la varianza asintótica (porque nos pide el intervalo de confianza asintótico).

Ejercicio 8: Transformación de estimadores eficientes
Supongamos que $\hat{\theta}$ es un estimador eficiente de θ (insesgado y alcanza la cota de Cramér-Rao).

1. Consideramos la transformación $f(\theta) = a\theta + b = \alpha$. Probar que $\hat{f}(\theta) = a\hat{\theta} + b$ es un estimador eficiente de $\alpha = f(\theta)$.
2. Si $f(\theta) = \theta^2$, muestre que $\hat{f}(\theta) = \hat{\theta}^2$ deja de ser eficiente para θ^2 pero sí lo es asintóticamente. Se sugiere considerar $\hat{\theta} = \bar{X}$ estimador de μ con una muestra $X_1, \ldots, X_n$ iid de $N(\mu, \sigma^2)$.

Sugerencias:

1. En la parte 1 utilizar que la cota de CR para $\alpha = f(\theta)$ es

$$\mathrm{Var}(\alpha) \geq \frac{\left(\frac{\partial f}{\partial \theta}\right)^2}{\mathcal{I}(\theta)}$$

2. En la parte 2 utilizar que si $Z \sim \mathcal{N}(\mu, \sigma^2)$ entonces

$$\mathrm{Var}(Z^2) = 4\mu^2\sigma^2 + 2\sigma^2$$

Ejercicio 9: Divergencia de Kullback-Leibler

1. Probar que la divergencia de Kullback-Leibler $D(p\|q) \geq 0$ y que $D(p\|q) = 0 \iff p(x) = q(x)$ para todo x.
2. Calcular $D(f_1\|f_2)$ en los siguientes casos:
   1. $f_1 = \mathcal{N}(2, \sigma^2)$ y $f_2 = \mathcal{N}(4, \sigma^2)$.
   2. $f_1 = \mathcal{N}(\mu_1, \sigma_1^2)$ y $f_2 = \mathcal{N}(\mu_2, \sigma_2^2)$.

Sugerencias:

1. En la parte 1, demostrar que $-D(p\|q) \leq 0$ usando
$$log(x) \leq x - 1$$
   También se puede hacer usando la desigualdad de Jensen.
2. Para demostrar $D(p\|q) = 0 \iff p(x) = q(x)$ para todo x, usar que $\log(x)$ es estrictamente cóncava, y recordar la desigualdad de Jensen para funciones estrictamente cóncavas (la usamos en el ejercicio 3)
3. En la parte 2, hacer cuentas!

Ejercicio 10: Información de Fisher

Calcular la información de Fisher en cada uno de los casos siguientes:

1. Si $X \sim \text{Ber}(p)$.
2. Si $X \sim \exp(\lambda)$.
3. Si $X \sim \mathcal{N}(\mu, 4)$.

Recordar que la información de Fisher se define como

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2 log \mathcal{L}(\theta)}{\partial\theta^2}\right]$$

1) $X \sim \text{Ber}(p)$

$$\mathcal{L}(p) = f_p(x) = p^x(1-p)^{1-x} \qquad x = 0, 1$$

La función de log verosimilitud es

$$l(p) = \log \mathcal{L}(p) = x\log(p) + (1-x)\log(1-p)$$

Luego derivamos dos veces la log verosimilitud

$$\frac{\partial l(p)}{\partial p} = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\frac{\partial^2 l(p)}{\partial p^2} = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

y calculamos la esperanza (respecto a la distribución de X)

$$\begin{aligned}
\mathcal{I}(p) = -\mathbb{E}\left[\frac{\partial^2 log \mathcal{L}(p)}{\partial p^2}\right] &= -\mathbb{E}\left[-\frac{X}{p^2} - \frac{1-X}{(1-p)^2}\right] \\
&= \frac{\mathbb{E}(X)}{p^2} + \frac{1-\mathbb{E}(X)}{(1-p)^2} \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} \\
&= \frac{1}{p(1-p)}
\end{aligned}$$

Ejercicio 11: Cramér-Rao Generalizado

Si $\hat{\theta}_n$ es un estimador de θ con $\mathbb{E}_{X\sim f_\theta}(\hat{\theta}_n) = \tau(\theta)$, entonces

$$\text{Var}(\hat{\theta}_n) \geq \frac{|\tau'(\theta)|^2}{n\mathcal{I}_1(\theta)}$$

Sugerencia: Ver prueba realizada en el teórico (Tema 3, diapositiva 39)
Para estimadores insesgados se utiliza que

$$Cov\left(\hat{\theta}_n, \frac{\partial l(\theta)}{\partial \theta}\right) = 1$$

En este caso la cuenta va a ser la misma, pero

$$Cov\left(\hat{\theta}_n, \frac{\partial l(\theta)}{\partial \theta}\right) = \tau'(\theta)$$

(demostrarlo)

Viernes 13/9:

- Regresión lineal simple.