

Se recomienda empezar por los ejercicios marcados sin *.

A) Riesgo, Error Cuadrático Medio, Estimadores Insesgados.

1. Sea X_1, \dots, X_n una muestra aleatoria de una distribución con parámetro θ . Sea $\hat{\theta}_n = \delta(\mathbf{X})$ un estimador de θ con varianza finita. Si $b(\hat{\theta}_n)$ es el sesgo de $\hat{\theta}_n$, probar que $\text{ECM}(\hat{\theta}_n) = V(\hat{\theta}_n) + b(\hat{\theta}_n)^2$.
2. Sea X_1, \dots, X_n una m.a. de una distribución tal que existe $V(X) = \sigma^2$.

(a) Demostrar que

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

es un estimador insesgado de σ^2 .

(b) Sea X_1, \dots, X_n es una m.a. de una distribución $N(\mu, \sigma^2)$, calcular el $\text{ECM}(s^2)$ y el $\text{ECM}(\tilde{\sigma}^2)$ donde $\tilde{\sigma}^2 = (n-1)s^2/(n+1)$.

Sugerencia: puede usar que si X_1, \dots, X_n son i.i.d con distribución $N(\mu, \sigma^2)$, entonces

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

(c) Probar que, aunque $\text{ECM}(s^2) > \text{ECM}(\tilde{\sigma}^2)$, la razón entre los ECM tiende a 1 cuando $n \rightarrow \infty$.

(d) Mostrar que la contribución del sesgo de $\tilde{\sigma}^2$ a $\text{ECM}(\tilde{\sigma}^2)$ es despreciable cuando $n \rightarrow \infty$ (en el sentido de que la razón entre $b(\tilde{\sigma}^2)^2$ y $\text{ECM}(\tilde{\sigma}^2)$ tiende a 0).

3. Dada una m.a. X_1, \dots, X_n de una distribución $\mathcal{U}(0, \theta)$, sea $\hat{\theta}_n$ el EMV de θ y $\tilde{\theta}_n$ el estimador de θ basado en el primer momento.

(a) Probar que $\tilde{\theta}_n$ es insesgado y que $\hat{\theta}_n$ es asintóticamente insesgado.

(b) Calcular el ECM de ambos estimadores. ¿Qué estimador preferiría desde este punto de vista?

(c) ¿Cómo modificaría el estimador $\hat{\theta}_n$ para que quede insesgado? Llamemos $\hat{\theta}_n^{\text{mod}}$ a este estimador modificado.

(d) (Para hacer con R) Para cada uno de los valores $n = 6, 10, 20, 40, 80, 200$ hacer lo siguiente:

i. Generar una muestra de tamaño n de una distribución $\mathcal{U}(0, \theta_0)$ con $\theta_0 = 3$.

ii. Evaluar los tres estimadores $\tilde{\theta}_n, \hat{\theta}_n$ y $\hat{\theta}_n^{\text{mod}}$ en dicha muestra.

iii. Repetir los dos pasos anteriores $m = 1000$ veces, obteniendo así, para cada uno de los tres estimadores, replicaciones $\theta_1^*, \dots, \theta_m^*$.

iv. Para cada estimador, computar el ECM

$$\frac{1}{m} \sum_{i=1}^m (\theta_i^* - \theta_0)^2$$

Finalmente, graficar n vs. el ECM de cada uno de los tres métodos anteriores (todo en un mismo gráfico). Hacer boxplots con los valores estimados $\theta_1^*, \dots, \theta_m^*$. ¿Tienen sentido estos resultados respecto de los resultados teóricos de los primeros items?

* 4. Un estimador δ se dice **inadmisible** para una cierta función de riesgo $R(\delta, \theta)$ si existe otro estimador δ^* tal que $R(\delta, \theta) \geq R(\delta^*, \theta)$ para todo $\theta \in \Theta$ y $R(\delta, \theta) > R(\delta^*, \theta)$ para algún $\theta \in \Theta$. Un estimador es **admissible** si no es inadmissible.

(a) Probar que los estimadores Bayes (respecto a la pérdida cuadrática) son admisibles para el ECM, o sea, para el riesgo $R(\delta, \theta) = \mathbb{E}[(\delta(\mathbf{X}) - \theta)^2]$.

(b) Consideremos el riesgo cuadrático pesado, o sea $R(\delta, \theta) = \mathbb{E}[w(\theta)(\delta(\mathbf{X}) - \theta)^2]$ con w una función positiva. Sea X_1, \dots, X_n una muestra aleatoria con $\mathbb{E}[X_1] = \mu$ y $\text{VAR}(X_1) = \sigma^2 < \infty$.

i. Probar que los estimadores de la forma $a\bar{X} + b$ con $a > 1$ son inadmisibles respecto a esta función de riesgo.

ii. Probar que los estimadores de la forma $\bar{X} + b$ con $b \neq 0$ son inadmisibles respecto a esta función de riesgo.

* 5. Sea X_1, \dots, X_n una m.a. de una distribución tal que existe $E(X_1) = \mu$. Consideremos la clase de estimadores

$$\mathcal{D} = \left\{ \delta(\mathbf{X}) = \sum_{i=1}^n \alpha_i X_i : \alpha_i \in \mathbb{R}, \sum_{i=1}^n \alpha_i = 1 \right\}.$$

(a) Probar que todo $\delta \in \mathcal{D}$ es un estimador insesgado de μ .

(b) Supongamos que existe $V(X) = \sigma^2$. Mostrar que $\bar{X} = \arg \min_{\delta \in \mathcal{D}} V(\delta)$.

* 6. (Para hacer con R) Volvemos a los estimadores James-Stein de la práctica anterior. Vamos a comparar los estimadores Bayes y EMV para $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$. Para cada uno de los valores $n = 6, 10, 20, 40, 80, 200$ hacer lo siguiente:

(a) Generar los μ_1, \dots, μ_n independientemente a partir de una distribución $N(2, 4)$ y a partir de esto obtener X_1, \dots, X_n donde $X_i | \mu_i \sim N(\mu_i, 1)$.

(b) Estimar $\boldsymbol{\mu}$ con el método de máxima verosimilitud y con el estimador de James-Stein.

(c) Para cada estimador, obtener

$$e = \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i - \mu_i)^2.$$

(d) Replicar los últimos 3 pasos $m = 1000$ veces, obteniendo para cada estimador los errores e_1, \dots, e_m . Estimar el ECM para cada uno de los dos estimadores promediando estas m repeticiones.

(e) Graficar (en un único gráfico) n vs. el estimador del ECM obtenido en el item anterior para cada metodo.

B) Consistencia y distribución asintótica

1. (a) Sea δ_n un estimador de $q(\theta)$. Probar que si $\text{ECM}_\theta(\delta_n) \rightarrow 0$, entonces δ_n es débilmente consistente.

(b) Sea X_1, \dots, X_n una m.a. tal que $E_\theta(X) = \theta$ y $V_\theta(X) < \infty$. Consideremos el siguiente estimador aleatorizado de θ :

$$\delta_n = (1 - \varepsilon_n) \bar{X} + \varepsilon_n n$$

donde $\varepsilon_n \sim \text{Bi}(1, 1/n)$ y es independiente de las X_i . Probar que δ_n es débilmente consistente, aunque $\text{ECM}_\theta(\delta_n) \rightarrow \infty$.

2. Sea X_1, \dots, X_n una muestra aleatoria de una distribución $\mathcal{E}(\theta)$. Probar que el EMV de θ es fuertemente consistente y hallar su distribución asintótica.
3. Sea X_1, \dots, X_n una muestra aleatoria de una distribución $\mathcal{P}(\lambda)$. Consideremos $\delta_n = \bar{X}$ y

$$\delta_n^* = -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{1}{n} \sum_{i=1}^n X_i^2}$$

- (a) Analizar si estos estimadores de λ son fuertemente consistentes.
- (b) Hallar sus distribuciones asintóticas. Decir si alguno de los dos es asintóticamente eficiente. (NOTA: $E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$)
4. Sea X_1, \dots, X_n una muestra aleatoria de una distribución $U(0, \theta)$.
- (a) (Para hacer con el R) Hacer histogramas para los valores $\theta_1^*, \dots, \theta_n^*$ para cada n y cada uno de los tres estimadores del ejercicio A.3 de esta práctica. ¿A qué familia intuye que pertenecen las distribuciones asintóticas de dichos estimadores?
- (b) Probar que el EMV de θ es débilmente y fuertemente consistente y hallar su distribución asintótica.
- (c) Probar que el estimador de momentos para θ es fuertemente consistente y hallar su distribución asintótica.
- (d) ¿Cuál de estos dos estimadores preferiría? ¿Por qué?
5. Sea X_1, \dots, X_n una m.a. de una distribución exponencial desplazada, cuya densidad es

$$f(x) = e^{-(x-\theta)} I_{[\theta, \infty)}(x).$$

- (a) Probar que el EMV de θ es débilmente y fuertemente consistente y hallar su distribución asintótica.
- (b) Probar que el estimador de momentos para θ es consistente y hallar su distribución asintótica.
- (c) ¿Cuál de estos dos estimadores preferiría? ¿Por qué?
- * 6. Sea X_1, \dots, X_n una m.a. de una distribución $F_0(x - \theta)$ con F_0 una función de distribución fija y $\theta \in \mathbb{R}$. Sea δ_n un estimador de θ equivariante por traslaciones (i.e.: para cualquier constante $c \in \mathbb{R}$ se verifica que $\delta_n(x_1 + c, \dots, x_n + c) = \delta_n(x_1, \dots, x_n) + c$).
- (a) Mostrar que $\delta_n - \theta$ tiene una distribución que no depende de θ .
- (b) Deducir que la varianza asintótica de δ_n (si existe) no depende de θ .
- * 7. Sea X_1, \dots, X_n una muestra aleatoria de una distribución F con densidad $f(x - \mu)$ donde $f(x)$ satisface: i) $f(x) = f(-x)$, ii) $f(x)$ continua en $x = 0$.
- (a) Sea $Z_{a,i} = I\{X_i \leq a\}$. Mostrar que $Z_{a,i}$ tiene distribución $\text{Bi}(1, F(a))$.
- (b) Sea $Z_a = \sum Z_{a,i}$. Mostrar que Z_a tiene distribución $\text{Bi}(n, F(a))$.
- (c) Probar que
- $$P(Z_a \geq \lfloor \frac{n}{2} \rfloor + 1) \leq P(\text{mediana}(X_1, \dots, X_n) \leq a) \leq P(Z_a \geq \lfloor \frac{n}{2} \rfloor)$$
- (d) Mostrar que $\text{mediana}(X_1, \dots, X_n)$ es un estimador débilmente consistente de μ .