Estadística (M) - Segundo cuatrimestre 2019

Práctica 1 - Estimación Puntual

Se recomienda empezar por los ejercicios marcados sin †.

A) *Estimadores basados en momentos.*

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución

   - $Bi\,(1, \theta)\,,\ 0 < \theta < 1.$
   - $\mathcal{P}\,(\theta)\,,\ \theta > 0.$
   - $\mathcal{E}\,(\theta)\,,\ \theta > 0.$
   - $\mathcal{G}\,(\theta)\,,\ 0 < \theta < 1.$

   En cada uno de estos casos, encontrar:

   (a) Un estimador de θ basado únicamente en el primer momento.

   (b) Un estimador de θ basado únicamente en el segundo momento.

   Verificar si los respectivos estimadores cumplan las restricciones impuestas sobre el parámetro. Por ejemplo, para el caso de la Binomial, verificar si la imagen del estimador obtenido esté contenido en el intervalo (0,1).

2. **Estimadores de momentos para** $q(\theta)$. Supongamos que tenemos una muestra aleatoria $X_1, \ldots, X_n$ y queremos un estimador basado en momentos para $q(\theta)$, donde θ es un parámetro desconocido de la distribución de la muestra y q es una función conocida. Consideremos las siguientes dos estrategias:

   (a) Si tenemos $\widehat{\theta}$ un estimador de momentos para θ, estimar $q(\theta)$ con $q(\widehat{\theta})$.

   (b) Buscar una función g tal que $\mathbb{E}[g(X_1)] = q(\theta)$ y luego aproximar $\mathbb{E}[g(X_1)]$ con $\frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

   Aplicar las estrategias (a) y (b) en los siguientes casos:

   (i) Sea $X_1, \ldots, X_n$ una m.a. de una distribución $\mathcal{E}\,(\theta)$ y se quiere estimar $q\,(\theta) = P_\theta\,(X_1 \geq 1)$.

   (ii) Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N\,(0, \sigma^2)\,$, con $\sigma > 0$ y se quiere estimar σ.

† 3. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas. Hallar el estimador de los momentos de μ si

   (a) la densidad de X_1 está dada por:

   $$f\,(x) = \frac{1}{\mu^4} x I_{[0,\mu^2]}\,(x) + \frac{1}{\mu^4}(2\mu^2 - x) I_{[\mu^2, 2\mu^2]}\,(x) \qquad \mu > 0$$

   (b) la función de probabilidad de X_1 está dada por:

k	1	2	4
$p\,(k)$	μ^2	$2\mu\,(1-\mu)$	$(1-\mu)^2$

   $0 < \mu < 1$

B) *Estimadores de máxima verosimilitud.*

1. Dada una muestra aleatoria $X_1, \ldots, X_n$ para cada una de las distribuciones consideradas en el ejercicio A.1, encontrar los respectivos estimadores de máxima verosimilitud (EMV) de θ.

2. Sea $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu, \sigma^2)$. Consideremos el parámetro bivariado $\theta = (\mu, \sigma^2)$.

   (a) Encontrar el EMV de θ.

   (b) Dado $\xi \in \mathbb{R}$ encontrar el EMV de $p = P_\theta (X_1 > \xi)$.

   (c) Dado $p \in (0,1)$, encontrar el EMV del valor ξ tal que $P_\theta (X_1 > \xi) = p$.

   (d) Hallar el EMV de μ cuando σ^2 es conocido. ¿Es razonable que no dependa de σ^2?

   (e) Hallar el EMV de σ^2 cuando μ es conocido. ¿Es razonable que dependa de μ?

3. Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}[0, \theta]$

   (a) Hallar el EMV de θ.

   (b) Hallar el estimador de los momentos de θ.

4. Sea $X_1, \ldots, X_n$ una m.a. de una distribución exponencial desplazada, cuya densidad es
$$f(x) = e^{-(x-\theta)} I_{[\theta,\infty)}(x).$$

   (a) Encontrar el EMV de θ.

   (b) Encontrar el estimador de los momentos de θ.

5. Sean $X_1, \ldots, X_n$ una m.a. de una distribución $N(\mu_1, \sigma^2)$ e $Y_1, \ldots, Y_m$ una m.a. de una distribución $N(\mu_2, \sigma^2)$, independientes entre sí.

   (a) Hallar el EMV de $\theta = (\mu_1, \mu_2, \sigma^2)$.

   (b) Hallar el EMV de $\alpha = \mu_1 - \mu_2$.

† 6. Se tienen observaciones independientes $X_1, \ldots, X_n$ de poblaciones normales con la misma media μ pero con varianzas $\sigma_1^2, \ldots, \sigma_n^2$ respectivamente.

   (a) ¿Es posible estimar todos los parámetros por máxima verosimilitud?

   (b) Suponiendo $\sigma_1^2, \ldots, \sigma_n^2$ conocidos, hallar el EMV de μ. Interpretar.

† 7. Si $X_1, \ldots, X_n$ es una m.a. de una distribución $\mathcal{U}[\theta - 1/2, \theta + 1/2]$, mostrar que cualquier T tal que $X_{(n)} - 1/2 \leq T \leq X_{(1)} + 1/2$ es un EMV de θ.

   Aclaración: $X_{(1)} = \min\{X_1, \ldots, X_n\}$ y $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

† 8. El número de microorganismos que se encuentran por grupo sobre una superficie sigue la siguiente distribución
$$p(x) = \theta I_{\{1\}}(x) + \frac{1-\theta}{k-1} I_{\{2,3,\ldots,k\}}(x)$$
   donde $0 < \theta < 1$. Supongamos que se examinan en forma independiente n grupos y se cuentan el número de microorganismos $X_1, \ldots, X_n$ en cada uno de ellos. Calcular el estimador de máxima verosimilitud de θ.

C) *Estimadores Bayesianos con pérdida cuadrática.*

En esta sección, todos los estimadores Bayes pedidos son respecto a la pérdida cuadrática. Más específicamente, sea $\mathbf{X} = (X_1, \ldots, X_n)$ es una m.a. tal que $X_1|_{\boldsymbol{\theta}=\theta} \sim F_\theta$ y $\boldsymbol{\Theta} \sim \Lambda$ es la distribución a priori de $\boldsymbol{\Theta}$. Dado un estimador $\delta(\mathbf{X})$ de $q(\theta)$, llamamos riesgo Bayes a

$$r(\Lambda, \delta) = \int \mathbb{E}_\theta[(\delta(\mathbf{X}) - q(\theta))^2] \, d\Lambda(\theta)$$

y estimador Bayes para $q(\theta)$ a

$$\delta_\Lambda(\mathbf{X}) = \underset{\delta}{\operatorname{argmin}}\, r(\Lambda, \delta).$$

En este caso, el estimador Bayes puede ser obtenido simplemente como la esperanza de la distribución a posteriori:

$$\delta_\Lambda(\mathbf{X}) = \mathbb{E}[q(\boldsymbol{\Theta})|\mathbf{X}].$$

1. (a) Sea $\boldsymbol{\theta}$ con distribución a priori Λ y $X_1, \ldots, X_n$ una m.a. tal que $X_i|_{\boldsymbol{\theta}=\theta} \sim F_\theta$. Se desea estimar $q(\theta)$ utilizando la función de pérdida cuadrática. Demostrar que si el estimador de Bayes $\delta_\Lambda(\mathbf{X})$ es un estimador insesgado de $q(\theta)$, es decir

   $$E\left(\delta_\Lambda(\mathbf{X}) \mid \boldsymbol{\theta} = \theta\right) = q(\theta) \quad \forall \theta \in \Theta,$$

   entonces $E\left[\left(\delta_\Lambda(\mathbf{X}) - q(\boldsymbol{\theta})\right)^2\right] = 0$.

   (b) Mostrar que $\bar{X}$ no es un estimador de Bayes de θ para ninguna distribución a priori Λ cuando $X|_{\boldsymbol{\theta}=\theta} \sim N(\theta, 1)$ y cuando se usa la pérdida cuadrática.

2. Sea X una v.a. tal que $X|_{\boldsymbol{\theta}=\theta} \sim \mathcal{U}(0, \theta)$ y $\boldsymbol{\theta} \sim \Gamma(2, 1)$. Encontrar el estimador Bayes de θ. Aclaración: notar que tenemos una sola variable aleatoria, es decir, una muestra con $n = 1$.

3. Sea $X_1, \ldots, X_n$ una m.a. tal que $X_i|_{\boldsymbol{\theta}=\theta} \sim Bi(1, \theta)$ y $\boldsymbol{\theta} \sim \Lambda$ continua. Consideremos la función de pérdida cuadrática.

   (a) Probar que $\bar{X}$ no es un estimador de Bayes de θ para ninguna distribución a priori Λ.

   (b) Hallar el estimador de Bayes δ_Λ cuando $\Lambda = \beta(r, s)$, con $r, s > 0$. Mostrar que δ_Λ es un promedio pesado entre $\bar{X}$ y $\dfrac{r}{r+s}$. Interpretar.

   Recordar: que $\boldsymbol{\theta} \sim \beta(r, s)$ si su función de densidad es

   $$f(\theta) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \theta^{r-1} (1-\theta)^{s-1} I_{(0,1)}(\theta)$$

4. Consideremos una m.a. $X_1, \ldots, X_n$ tal que $X_i|_{\boldsymbol{\theta}=\theta} \sim \mathcal{P}(\theta)$ y $\boldsymbol{\theta} \sim \Gamma(r, \lambda)$, con $r, \lambda > 0$.

   (a) Encontrar el estimador Bayes δ_Λ y calcular $r(\delta_\Lambda, \Lambda)$ el riesgo de Bayes.

   (b) Mostrar que δ_Λ puede escribirse como un promedio pesado entre $\bar{X}$ y $\dfrac{r}{\lambda}$. Interpretar.

   (c) Comparar $r(\delta_\Lambda, \Lambda)$ con $r\left(\bar{X}, \Lambda\right)$. Mostrar que $\dfrac{r\left(\bar{X}, \Lambda\right)}{r(\delta_\Lambda, \Lambda)} \to 1$ cuando $n \to \infty$.

† 5. **Detección de SPAM.** Supongamos que la probabilidad de que un mail sea SPAM es $p \in (0,1)$ (conocida). Supongamos también que los mails que son SPAM provienen de una distribución con densidad $f_1(x)$ y los mails que no lo son provienen de una distribución con densidad $f_0(x)$. Dado un nuevo mail X, la regla **MAP** (maximum a posteriori) clasifica a dicho mail como SPAM si y sólo si

$$\mathbb{P}(\text{el mail es SPAM}|X=x) > \mathbb{P}(\text{el mail no es SPAM}|X=x).$$

Demostrar que esta regla clasifica al mail X como SPAM si y sólo si

$$\frac{f_1(X)}{f_0(X)} > \frac{1-p}{p}.$$

† 6. **Estimadores de James-Stein.**

(a) Supongamos que tenemos una única observación X en el siguiente modelo Bayesiano: $X|\boldsymbol{\mu}_{=\mu} \sim N(\mu,1)$ y $\boldsymbol{\mu} \sim N(M,A)$. Hallar el estimador Bayes de μ.

(b) Sea $X_1,\dots,X_n$ con $n \geq 4$ una muestra aleatoria tal que $X_i|\boldsymbol{\mu}_{i=\mu_i} \sim N(\mu_i,1)$ y $\boldsymbol{\mu}_i \sim N(M,A)$ independientemente para $i=1,\dots,n$. Notar que a diferencia del modelo Bayesiano tradicional, el parámetro μ_i varía con la observación. Sea $\boldsymbol{\mu} = (\mu_1,\dots,\mu_n)^{\mathrm{T}}$ y $\mathbf{X} = (X_1,\dots,X_n)$.

i. Hallar el EMV de $\boldsymbol{\mu}$.

ii. Probar que el estimador Bayes para $\boldsymbol{\mu}$ es

$$\widehat{\boldsymbol{\mu}}_{\text{bayes}} = \mathbf{M} + B(\mathbf{X}-\mathbf{M})$$

donde $\mathbf{M} = (M,M,\dots,M)^{\mathrm{T}}$ y $B = A/(A+1)$.

En la práctica, M y B no son conocidos y se estiman con

$$\widehat{M} = \overline{X} \qquad \text{y} \qquad \widehat{B} = 1 - \frac{n-3}{\sum_{i=1}^n (X_i - \overline{X})^2}.$$

El estimador $\widehat{\boldsymbol{\mu}}_{JS} = \widehat{\mathbf{M}} + \widehat{B}(\mathbf{X}-\widehat{\mathbf{M}})$ donde $\widehat{\mathbf{M}} = (\widehat{M},\dots,\widehat{M})^{\mathrm{T}}$ se conoce como el estimador de James-Stein.