Examen General de Estadística
Semestre 2020-1
Enero 20, 2020.
9:00-14:00 hrs

Instrucciones: Las tres primeras preguntas correspondientes a Inferencia Estadística son obligatorias. De las seis preguntas correspondientes a Inferencia Bayesiana y Modelos Lineales; resuelve solamente tres; todas las preguntas de estas dos secciones tienen en mismo valor. Es decir, resuelve solamente seis preguntas: las tres primeras, y otras tres a ser seleccionadas de entre las últimas seis. Tiempo máximo de examen: 5 horas.

1. Inferencia

Instrucciones: Justifique sus respuestas con los procedimientos correspondientes.
1. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población representada por una variable aleatoria X de una población con distribución $N(\theta, 1)$.

a) Encuentre la Cota Inferior de Cramér-Rao para la varianza de los estimadores insesgados de θ^2 y $P(X > 0)$. [2 puntos]

b) Encuentre el estimador máximo verosimil para $P(X > 0)$. [1 punto]

c) ¿Existe algún estimador insesgado para $P(X > 0)$? De ser el caso encuéntrelo. [1 punto]

d) ¿Existe un UMVUE para $P(X > 0)$? De ser el caso encuéntrelo y exhíbalo. [5 puntos]

2. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con f.d.p.

$$f(x|\theta) = \frac{2x}{\theta^2}\, 1_{[0,\theta]}(x),$$

encuentre un intervalo de confianza del $(1-\alpha) \times 100\,\%$ de longitud óptima para θ (o en su defecto el intervalo de menor longitud que sea posible obtener). [6 puntos]

3. Suponga que $X_1, X_2, \ldots, X_n$ una m.a. de una población con distribución Beta$(\mu, 1)$ y $Y_1, Y_2, \ldots, Y_m$ una muestra aleatoria de una población con distribución Beta$(\theta, 1)$. Además, suponga que las X's y las Y's son independientes.

(a) Obtenga la región crítica para realizar el contraste

$$H_0 : \theta = \mu \;\; Vs \;\; H_1 : \theta \neq \mu.$$

[4 puntos]

(b) Muestre que la región crítica de la parte anterior, puede basarse en la estadística

$$T = \frac{\sum_{i=1}^{n} \log(X_i)}{\sum_{i=1}^{n} \log(X_i) + \sum_{j=1}^{m} \log(Y_i)}.$$

[4 puntos]

2. Inferencia Bayesiana

Recuerde que Gamma(a, b) denota a la distribución Gamma con función de densidad dada por

$$f_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}.$$

1. Sea Y una variable aleatoria con distribución Gamma tal que su media y su varianza son iguales a un entero positivo M (desconocido). Suponga que la distribución inicial de M es Geométrica, con función de densidad

$$p(m) = \theta^{m-1}(1-\theta) \quad (m = 1, 2, \ldots),$$

donde $0 < \theta < 1$.

(a) Demuestre que la distribución final de $M - 1$ dado que $Y = y$, es Poisson con media $\lambda = y\,\theta$.

(b) Considere la función de pérdida

$$L(d, M) = \frac{(M-d)^2}{M} \quad (0 < d < \infty).$$

Muestre que la pérdida esperada, respecto a la distribución final encontrada en el inciso anterior, está dada por

$$\bar{L}(d) = 1 + \lambda - 2d + \frac{(1-e^{-\lambda})d^2}{\lambda}.$$

(c) Encuentre el estimador bayesiano de M con base en esta función de pérdida.

2. Se desea aproximar la densidad conjunta $p(\theta, \phi)$ a través de una densidad de la forma

$$g(\theta, \phi) = g_1(\theta) g_2(\phi),$$

bajo la cual θ y ϕ son independientes.

Demuestre que la mejor aproximación, en el sentido que minimiza la divergencia de Kullback-Leibler entre $p(\theta, \phi)$ y $g(\theta, \phi)$, es tal que

$$g_1(\theta) = \int p(\theta, \phi) d\phi \text{ y } g_2(\phi) = \int p(\theta, \phi) d\theta.$$

3. Sea $\{X_1, \ldots, X_n\}$ una muestra aleatoria con distribución Pareto(β, γ), cuya función de densidad está dada por

$$f_{\beta,\gamma}(x) = \frac{\beta \gamma^\beta}{x^{\beta+1}} I_{[x \geq \gamma]}, \quad \beta, \gamma > 0.$$

Suponga que observamos $Y = \text{mín}\{X_1, \ldots, X_n\}$ y que γ es conocida pero β no.

(a) Muestre que la distribución Gamma(r, θ), con $r, \theta > 0$, es una familia conjugada para la distribución Pareto(β, γ), con γ fija. Es decir, muestre que la distribución final de β es una distribución Gamma.
(*Hint:* La estadística Y también sigue una distribución *Pareto*.)

(b) Suponiendo que $\beta \sim$ Gamma(r, θ), encuentre el estimador de Bayes bajo una función de pérdida cuadrática.

3. Modelos Lineales

1. Sea u un vector particionado, cuyo vector de medias y su matriz de covarianzas son: $E(u) = E \begin{pmatrix} y \\ x \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}$, con $y \in \mathcal{R}^p$; $x \in \mathcal{R}^q$. Sea $A_{q \times p}$ una matriz de constantes.

(a) Demuestre que $E(x'Zy) = tr(A\Sigma_{xy}) + \mu_x' A \mu_y$.

(b) Si $u \sim N(\mu, \Sigma)$, encuentre la distribución de $x|y$.

2. Suponga que $y = X\beta + \epsilon$, con $\text{Cov}(y) = \sigma^2 V$ y que de forma errónea se usa el estimador de mínimos cuadrados ordinarios $\hat{\beta}^* = (X'X)^{-1}X'y$ en lugar del estimado correcto $\hat{\beta} = (X'V^{-1}X)^{-1}X'v^{-1}y$. Suponiendo una distribución normal con media cero para los errores:

(a) Encuentre los estimadores máximo verosímiles de β y σ^2.

(b) Encuentre la $E(\hat{\beta})$ y $\text{Cov}(\hat{\beta})$.

(c) Demuestre que $\hat{\beta}^*$ es insesgado, pero que $\text{Cov}(\hat{\beta}^*) \neq \sigma^2(X'X)^{-1}$.

3. En el modelo $y = X\beta + \epsilon$ con $E(y) = X\beta$ y con X de rango $q < p \leq n$,

(a) Enuncie bajo que condiciones la función $\lambda'\beta$ es estimable.

(b) Usando todas las condiciones anteriores demuestre que si $\mathbf{x}' = (1, x_1, \ldots, x_n)$ y $\mathbf{z}' = (x, c_1x_1, \ldots, c_nx_n)$

$$\hat{y} = \hat{\beta}'\mathbf{x} = \hat{\beta}_z{}'\mathbf{z},$$

donde $\hat{\beta}_z$ es el estimador de mínimos cuadrados de la regresión de $\mathbf{y}$ en $\mathbf{z}$.