Examen General de Estadística
Semestre 2020-2
Julio 27, 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 Estadística

1. Sea $X_{(n)} = \text{máx}(X_1, \ldots, X_n)$ donde $X_1, \ldots, X_n$ es una muestra aleatoria de una distribución exponencial con parámetro uno $(F(x) = 1 - e^{-x})$. a) Encuentre la distribución de $X_{(n)} - \log n$, en términos de F. Llámela $G_n(y)$. b) Encuentre el límite de $G_n(y)$ cuando n tiende a ∞, $(G(y))$ ¿Es una función de distribución?

2. Sea X una v.a. que describe el tiempo de vida (en horas) de una partícula radiactiva. Suponga que la f.d.p. de X está dada por:

$$f(x|\theta) = \theta e^{-x\theta}, \quad \text{para} \quad x > 0.$$

   Una m.a. de n partículas se pone bajo observación, pero el experimento se detiene cuando la k-ésima partícula expira; es decir, no se espera hasta que todas las partículas hayan cesado su actividad, sino solo hasta que k de ellas (con k fijada de antemano) lo hayan hecho. Los datos consisten en las k mediciones $Y_1, \ldots, Y_k$, y de las $n-k$ mediciones restantes sólo se sabe que exceden el valor Y_k.

   2.1. Encuentre el estimador de máxima verosimilitud para el tiempo medio de vida de las partículas, i.e. $1/\theta$.

   2.2. Exhiba una cantidad pivotal para $1/\theta$.

3. Sea X variable aleatoria con densidad

$$f(x:\theta) = \begin{cases} \theta x^{\theta-1} & x \in (0,1) \\ 0 & e.o.c \end{cases}$$

donde $\theta > 0$.

3.1. Para contrastar $H_0 \theta \leq 1$ vs $H_1 : \theta > 1$ se toma una muestra aleatoria de tamaño 2 de X y se usa la región de rechazo $\mathcal{C} = \{(x_1, x_2) | x_1 x_2 \geq 3/4\}$. Encuentre la función potencia y el tamaño de esta prueba.

3.2. Con una muestra aleatoria de tamaño 2 de X, encuentre la prueba uniformemente más potente de tamaño $\alpha = (1/2)(1 - \log 2)$ para contrastar $H0 : \theta \leq 1$ vs $H_1 : \theta > 1$.

2. Inferencia Bayesiana

1. En un embarazo doble, los bebés pueden ser monocigóticos (C) o dicigóticos (D). Los monocigóticos se originan del mismo óvulo, son muy parecidos entre sí y siempre son del mismo sexo, mientras que los dicigóticos son engendrados a partir de óvulos distintos y pueden ser de sexos distintos. (En términos coloquiales, a los primeros se les llama "gemelos idénticos" y a los segundos se les llama "mellizos".)

(a) Sea M el evento "bebé de sexo masculino" y F el evento "bebé de sexo femenino". Entonces,

(i) $\Pr[\{F, F\} | C] = 0.50$. ¿Cuál es el valor de $\Pr[\{M, M\} | C]$?

(ii) $\Pr[\{F, F\} | D] = 0.25$. ¿Cuál es el valor de $\Pr[\{M, M\} | D]$?

(iii) $\Pr[\{F, M\} | D] = 0.50$. ¿Cuál es el valor de $\Pr[\{F, M\} | C]$?

(b) Expresa a $\Pr[\{M, M\}]$ y a $\Pr[\{F, F\}]$ en términos de $\Pr[C]$; es decir, de la probabilidad de que los gemelos sean monocigóticos.

(c) Supón que nacen dos niñas. También en términos de $\Pr[C]$, calcula la probabilidad de que sean dicigóticas, es decir $\Pr[D | \{F, F\}]$. ¿Cuál es el valor de esta probabilidad cuando $\Pr[C] = 0.50$?

2. Supón que $X_i \sim \text{Poisson}(x_i|\theta_i)$, $i = 1, \ldots, n$, y que estas variables aleatorias son condicionalmente independientes dados los valores de los parámetros. Supón además que $\theta_1, \ldots, \theta_n$ se distribuyen de acuerdo a una ley exponencial con media $1/\phi$, y que son condicionalmente independientes dado el valor del hiperparámetro ϕ. Finalmente, considera una distribución inicial para ϕ de la forma $p(\phi) \propto 1$.

(a) Sea $\psi = (1 + \phi)^{-1}$. Demuestra que la distribución final de ψ es proporcional a

$$\psi^{n\bar{x}-2}(1-\psi)^n$$

siempre que $n\bar{x} > 1$.

(b) Usando este hecho, encuentra el estimador bayesiano de θ_i, $i = 1, \ldots, n$, basado en una función de pérdida cuadrática.

(c) ¿Qué pasa cuando $n\bar{x} \leq 1$?

3. Sea $X_1, X_2, \ldots, X_n$ una muestra de variables aleatorias i.i.d. de una distribución exponencial negativa $\text{Exp}(x|\theta)$, parametrizada de manera que $E(X|\theta) = 1/\theta$.

(a) Utilizando una función de pérdida cuadrática con respecto a $\mu = 1/\theta$, es decir,

$$L(\hat{\theta}, \theta) = \left(\hat{\theta}^{-1} - \theta^{-1}\right)^2$$

y suponiendo una distribución inicial conjugada para θ, encuentra el estimador bayesiano de θ.

(b) Dado que θ es un parámetro de escala, podría ser más apropiado utilizar una función de pérdida que sea invariante ante cambios de escala. Encuentra el estimado bayesiano de θ utilizando ahora la función de pérdida

$$L^*(\hat{\theta}, \theta) = \theta^2 \left(\hat{\theta}^{-1} - \theta^{-1}\right)^2$$

y suponiendo, al igual que antes, una distribución inicial conjugada para θ.

(c) ¿Cómo se comparan estos dos estimadores cuando $n \to \infty$?

3. Modelos Lineales

1. Si $Y = X\beta + \epsilon$ tiene una distribución Normal con media $X\beta$ y varianza $\sigma^2 V$. Demustra que si $\hat{\beta} = (X'V - 1X)^{-1}X'V^{-1}Y$:

a) $SCR = (Y - X\hat{\beta})'V^{-1}(Y - X\hat{\beta})$ se distribuye como $\sigma^2\chi^2_{n-p}$.

b) Este estimador es insesgado para σ^2 y con mínima varianza.

c) Si $\hat{Y} = X\hat{\beta} = PY$, P es idempotente pero no necesariamente simétrica.

d) $\hat{\beta}$ es el MELI de β.

2. Considera el modelo $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ con $i = 1, ..., I, j = 1, ..., J, k = 1, ..., K$. Prueba que

Si los efectos son considerados fijos tenemos

a) $E(CM\alpha) = \frac{JK}{I-1}\sum_{i=1}^{I}(\bar{\alpha}_i - \bar{\alpha}_. + \bar{\gamma}_{i.} - \bar{\gamma}_{..})^2 + \sigma^2$

b) $E(CM\gamma) = E(CM\alpha\beta) = \frac{K}{(I-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{\gamma}_{ij} - \bar{\gamma}_{i.} - \bar{\gamma}_{.j} + \bar{\gamma}_{..})^2 + \sigma^2$

Mientras que en un modelo con solo efectos aleatorios lo que tenemos es

a) $E(CM\alpha) = JK\sigma^2_\alpha + K\sigma^2_\gamma + \sigma^2$

b) $E(CM\gamma) = E(CM\alpha\beta) = K\sigma^2_\gamma + \sigma^2$

3. Prueba que en el modelo de análisis de covarianza $Y = X\beta + Z\gamma + \epsilon$, el estimador del modelo de regresión se puede obtener del modelo $Y = PZ\gamma + \epsilon$, donde $P = I - X(X'X)^{-1}X'$ y además se puede demostar que es equivalente el modelo $Y = R_z\gamma + \epsilon$ donde R_z son los residuos del modelo $Z = X\beta + \epsilon$, donde $X\beta$ corresponde al modelo de análisis de varianza.