Examen General de Estadística
Semestre 2018-2
Viernes 22 de junio de 2018
De 9:00 a 14:00 hrs

Instrucciones: Las tres primeras preguntas, correspondientes a Inferencia Estadística, son obligatorias. De las seis preguntas restantes, correspondientes a Inferencia Bayesiana y Modelos Lineales, responde solamente tres. Es decir, solamente debes resolver seis problemas: los tres primeros y otros tres a ser seleccionados de entre los últimos seis. Todos los problemas tienen el mismo valor.

Tiempo máximo de examen: 5 horas.

Inferencia Estadística

1. Sean $X_1, X_2, \cdots, X_n$ variables aleatorias independientes e idénticamente distribuidas de acuerdo a la densidad:

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 \exp(-\theta x),$$

con $x > 0$ y $\theta > 0$.

   a) Exhibe los estimadores máximo-verosímiles para θ, para $\frac{1}{\theta}$ y para $\log(\theta)$.

   b) Di para cuál de estas funciones de θ existe un estimador insesgado con varianza mínima. Exhibe dicho estimador (puedes argumentar usando la Cota de Cramér-Rao, o bien utilizando el Teorema de Rao-Blackwell). ¿Qué propiedades generales (asintóticas) tiene este estimador?

2. Sea $X_1, \ldots, X_n$ una muestra aleatoria de la distribución con densidad

$$f(x;a,\theta) = \theta a^\theta x^{-(\theta+1)} I_{(a,\infty)}(x), \qquad \theta > 0, \;\; a > 0.$$

   a) Exhibe la función de verosimilitud $l(a,\theta)$ y encuentra el estimador máximo-verosímil de θ cuando a es conocido.

b) Encuentra el estimador máximo-verosímil de a cuando θ es conocido (observa que la función de verosimilitud con θ fijo, vista como función de a, es creciente en a sobre el intervalo $(0, x_{(1)})$ y cero si $a \geq x_{(1)}$, siendo $x_{(1)}$ el mínimo de las x_i's; es decir, de la muestra observada).

c) Habiendo maximizado la verosimilitud para θ con a fija, y sabiendo cómo se maximiza respecto de a, entonces es inmediato identificar los estimadores máximo verosímiles de θ y a. Exhíbelos.

d) Exhibe una expresión para la estadística suficiente (minimal) para el problema en cuestión y argumenta cómo se podría mostrar que esa estadística es completa.

3. Considera una distribución exponencial negativa; esto es, con densidad

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

Su función de distribución, para $x > 0$, está dada por $F(x;\theta) = 1 - e^{-x\theta}$ y es cero si $x \leq 0$.

a) Demuestra que si se quiere obtener la distribución exponencial truncada (también referida como condicionada), a $x \in (0, 1)$ (no importa si los extremos del intervalo se incluyen), ésta tiene función de distribución

$$F_c(x;\theta) = \frac{1 - e^{-x\theta}}{1 - e^{-\theta}},$$

obviamente para valores de x en (0,1), siendo cero si $x \leq 0$ y uno si $x \geq 1$. Esa distribución tiene una densidad (es continua) y además tiene la propiedad de que como función de distribución paramétrica está bien definida aún si $\theta < 0$.

b) Demuestra que, si $\theta \to 0$, esta distribución converge a la distribución uniforme $U(0, 1)$. Por ello, como familia paramétrica para modelar distribuciones en el intervalo (0,1) resulta interesante pues las densidades son decrecientes si $\theta > 0$, son crecientes si $\theta < 0$ y reconstruye la uniforme si $\theta = 0$.

c) Considera $X_1, \cdots, X_n$ una muestra aleatoria de la distribución exponencial truncada. Demuestra que $t(\underline{X}) = \sum X_i$ es suficiente para θ; de hecho, demuestra que esta distribución es miembro de la familia exponencial (con dimensión 1).

d) Para hacer sencillo este inciso, considera $n = 1$ y exhibe la solución al problema de contraste de hipótesis $H_0 : \theta = 0$ vs $H_1 : \theta > 0$. Argumenta si la prueba obtenida es uniformemente más potente.

e) Repite d) pero ahora con $H_1 : \theta < 0$ y también para $H_1 : \theta \neq 0$.

Inferencia Bayesiana

1. Sea X una variable aleatoria (univariada) que representa algún atributo de interés de una población finita. Por ejemplo, X puede ser el ingreso mensual de los individuos en la población. Supongamos que la población consta de N individuos, así que lo más que podemos observar es $\{X_1, \ldots, X_N\}$ (en el caso de un censo).

   Supongamos que la distribución inicial de $\{X_1, \ldots, X_N\}$ puede escribirse como
   $$p(x_1, \ldots, x_N) = \int \prod_{i=1}^{N} p(x_i|\theta)\, p(\theta)\, d\,\theta,$$
   para algunas distribuciones $p(x|\theta)$ y $p(\theta)$.

   a) Supongamos que se observa una muestra de $\{X_1, \ldots, X_N\}$, digamos $\{x_1, \ldots, x_n\}$ con $n < N$. Encuentra la distribución final
   $$p(x_{n+1}, \ldots, x_N | x_1, \ldots, x_n)$$
   a partir de $p(\theta)$, la muestra observada y la función de densidad de las variables no observadas.

   b) Si en el inciso anterior suponemos que $p(x|\theta) = N(x|\theta, 1)$ y $p(\theta) = N(\theta|\mu, 1/\tau)$, donde la media μ y la precisión τ son conocidas, encuentra
   $$p(x_{n+1}|x_1, \ldots, x_n) \qquad \text{y} \qquad p(x_{n+1}, x_{n+2}|x_1, \ldots, x_n).$$

   c) Continuando con el inciso anterior, encuentra la media inicial y la media final de $S = \sum_{i=1}^{N} X_i$.

2. Sea $X_1, X_2, \ldots, X_n$ una muestra de variables aleatorias i.i.d. de una distribución exponencial negativa $\mathrm{Exp}(x|\theta)$, 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 conforme $n \to \infty$?

3. Sea $\boldsymbol{X} = (X_1, \ldots, X_k) \sim \text{Multinomial}(\boldsymbol{x}|\boldsymbol{\theta}, n)$ y supón que $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ tiene una distribución inicial Dirichlet con parámetro $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$. Entonces la distribución final de $\boldsymbol{\theta}$ dada una observación $\boldsymbol{X} = \boldsymbol{x}$ es también Dirichlet, pero con parámetro $\boldsymbol{\alpha} + \boldsymbol{x}$.

a) Demuestra que la distribución marginal de θ_i $(i = 1, \ldots, k)$ es

$$\text{Beta}\left(\theta_i \,\middle|\, \alpha_i + x_i, \sum_{j \neq i} (\alpha_j + x_j)\right).$$

b) Encuentra el estimador de Bayes para $\boldsymbol{\theta}$ bajo la función de pérdida

$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{k} \theta_i \, \log(\theta_i / \hat{\theta}_i).$$

c) Demuestra que la moda de la distribución final de $\boldsymbol{\theta}$ es

$$\tilde{\boldsymbol{\theta}} = \frac{1}{\sum_{i=1}^{k} (\alpha_j + x_j) - k} \cdot (\alpha_1 + x_1 - 1, \ldots, \alpha_k + x_k - 1).$$

Esta expresión también se puede considerar un estimador bayesiano de $\boldsymbol{\theta}$. ¿Cómo se compara $\tilde{\boldsymbol{\theta}}$ con el estimador obtenido en el inciso anterior?

Modelos Lineales

1. Considera variables aleatorias $Y_1, Y_2, \cdots, Y_n$ independientes, donde Y_i tiene distribución normal $N(\alpha + \beta z_i, 1)$, siendo las $z_i' s$ constantes conocidas y α, β parámetros desconocidos.

   a) Obtén la función de verosimilitud $l(\alpha, \beta)$ y explicita los estimadores máximo-verosímiles de los dos parámetros. Calcula la matriz de varianza-covarianza de estos estimadores y di bajo qué condiciones la covarianza es cero.

   b) Calcula la matriz de información de Fisher (la esperada; ¡Ojo porque en este caso no hay una por unidad muestral!). Comenta si la cota inferior para la matriz de varianza-covarianza de los estimadores insesgados de α y β se alcanza con los estimadores máximo-verosímiles.

2. Considera el modelo de regresión

$$y_i = \underline{x}_i' \beta + e_i, \quad i = 1, \cdots, n, \quad e_i \sim N(0, \sigma^2)$$

   con las $\underline{x}_i' \in R^p$ conocidas. Supón que la matriz X es de rango máximo. Los llamados residuos, $\hat{e}_i = y_i - \hat{y}_i$, se basan en la diferencia entre el valor observado (y_i) y el valor que el modelo estima $(\hat{y}_i)$. Ahora bien, $\hat{y}_i$ es el i-ésimo elemento del vector $\underline{\hat{y}} = X(X'X)^{-1}X'\underline{y}$, que es la llamada matriz "sombrero" multiplicando al vector $\underline{y}$.

   a) Demuestra que cualquier elemento de la diagonal de la matriz

$$I - X(X'X)^{-1}X'$$

   deberá ser no-negativo. Esto se basa exclusivamente en resultados de álgebra lineal.

   b) Demuestra también que

$$\mathrm{Var}(\hat{e}_i) = \sigma^2(1 - \underline{x}_i'(X'X)^{-1}\underline{x}_i).$$

   c) La propiedad de que cualquier elemento de la diagonal de la matriz sombrero es menor o igual que 1, permite aseverar que

$$\max_i(\underline{x}_i'(X'X)^{-1}\underline{x}_i) \leq 1.$$

Esto, aplicado al caso de un modelo lineal simple en donde $\underline{x}_i' = (1, z_i)$, donde las z_i son los valores de la variable "independiente", reproduce una desigualdad muy conocida en análisis numérico (la de Samuelson). Exhibe dicha desigualdad.

3. Considera el siguiente modelo

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad i = 1, 2, \qquad j = 1, \ldots, n_i$$

con los supuestos usuales para las ϵ_{ij} y con $n_1 = 2$, $n_2 = 3$ y las y_{ij} dadas por

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 14 \\ 16 \\ 19 \\ 23 \\ 15 \end{pmatrix}$$

a) Encuentra la mejor estimación lineal insesgada de $\tau_1 - \tau_2$.

b) Contrasta las hipótesis

$$H_o : \tau_1 = \tau_2 \qquad \text{vs.} \qquad H_a : \tau_1 \neq \tau_2$$

con nivel $\alpha = 0.05$. El cuantil de orden 0.95 de una distribución $F_{(1,3)}$ es 216.

c) ¿Qué hubieras hecho si las tablas disponibles para la distribución F sólo tuvieran grados de libertad para el numerador mayores o iguales que para el denominador? Dicho de otra forma, ¿qué hubieras hecho si pudieras obtener cualquier cuantil de una distribución $F_{(3,1)}$ pero no de $F_{(1,3)}$?