Examen General de Estadística Semestre 2014-II

25 de junio 2014

Instrucciones: Las tres primeras preguntas correspondientes a **Inferencia Estadística** son obligatorios. De las seis preguntas correspondientes a **Inferencia bayesiana** y **Modelos lineales** resuelva solamente 3. Es decir resuelva solamente seis preguntas: las tres primeras, y otras tres a ser seleccionadas de entre las últimas seis. Tiempo de máximo examen: 5 horas.

INFERENCIA

1. Sea $X_1, \cdots, X_n$ una muestra aleatoria de una distribución normal con ambos parámetros desconcidos. Para la parametrización con media y varianza. (μ, σ^2) encuentra la matriz cota a la de varianza-covarianza de los estimadores insesgados del vector (μ, σ^2); la cota conocida como de Cramér y Rao.

   a) Supón que nos interesa otra parametrización basada en el coeficiente de variación. denótalo por $\tau = \frac{\sigma}{\mu}$ y exhibe, usando la matriz cota antes descrita, la cota inferior a la varianza de los estimadores insesgados de τ.

   b) Alternativamente, obtén la matriz cota a la de varianza covarianza de los estimadores insesgados del vector (σ, τ) y verifica que se obtiene el mismo resultado para lo preguntado en el inciso anterior.

2. Existe una familia de densidades discretas para la variable aleatoria X, en el recorrido $\{0, 1, \cdots, K\}$ en el que K puede ser finito. o bien ∞, llamada de series de potencias en que la densidad es:

$$P[X = j; \theta] = f(j : \theta) = a_j\theta^j / b(\theta),$$

   con las $a_j \geq 0$, y θ en $(0, \infty)$ o en un subintervalo abierto de los reales positivos(en que $b(\theta)$ sea finito).

   (a) Muestra que la estadística suficiente minimal T, asociada a una muestra de tamaño m *iid* es la suma de las observaciones y que se está dentro de una familia exponencial.

   (b) Identifica K y el θ equivalente al parámetro p de la Bernoulli(p), para la binomial(n, p), para la binomial negativa(r, p), (con r entero positivo) y la Poisson(λ).

   (c) Observa que para los casos de K finita, el observar la estadística T, que permite inferir sobre θ equivale a utilizar la cantidad $(mK - T)$ para inferir sobre $1/\theta$; esto es, si X tiene densidad de las de serie de potencias con parámetro θ, entonces $K - X$ tiene densidad de series de potencias con parámetro $1/\theta$.

3. Supón un problema (simple) de bondad de ajuste en el que con base en una muestra *iid* $X_1, \cdots, X_n$, de una función de distribución continua F se desea probar $H_0 : F = F_0$ con F_0 totalmente especificada.

   Es usual utilizar para este tipo de problema una estadística que se base en la discrepancia entre la llamada función de distribución empírica, denotada F_n y la identificada por la hipótesis nula, F_0. Recuerda que

$$F_n(x) = \frac{\# X_i's \leq x}{n},$$

for

   y recuerda también que por ser F_0 contínua, podemos transfromar cadas X_i con F_0, digamos $U_i = F_0(X_i)$, (que por cierto es uniforme en (0,1)), y entonces,

$$\# X_i's \leq x = \# U_i's \leq u,$$

   con $u = F_0(x)$.

   Supón que la muestra que se observó es $x_1, \cdots, x_n$ y que los correspondientes valores para las transformadas son $u_1, \cdots, u_n$, que ordenados de menor a mayor son $u_{(1)}, \cdots, u_{(n)}$.

   Encuentre una expresión, función de las $u_(i)'s$ para la estadística de Cramér-von Mises:

$$n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x).$$

   Te sugerimos efectuar el cambio de variable $u = F_0(x)$ y partir la correspondiente integral entre 0 y 1 en los intervalos: $(0, u_{(1)})$, $(u_{(1)}, u_{(2)})$, hasta $(u_{(n)}, 1)$.

MODELOS LINEALES

1. Considera un modelo lineal normal de regresión

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

   con las $x_i' \in R^p$, conocidas.

   Si la matríz X $n \times p$ $(n > p)$ cuyas filas son las $\underline{x}_i'$ es de rango completo, $\hat{\beta}$, es única, $\hat{\sigma^2}$ está bien definida (son los máximo verosímiles) y se considera el "modelo ajustado":

$$\hat{y} = \underline{x}'\hat{\beta}.$$

   Dicho modelo se utiliza principalmente para estimar valores que tomaría la variable (dependiente) y, si los valores de las variables independientes están representados por el vector (fila) $\underline{x}'$.