Examen General de Estadística
Semestre 2016-1
Martes 19 de Enero
9:00-14:00hrs.

Instrucciones: Las tres primeras preguntas correspondientes a Inferencia Estadística son obligatorias. De las seis preguntas correspondientes a Modelos Lineales e Inferencia Bayesiana resuelve solamente tres. 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.

(Inferencia Estadística)

1. Sea $X_1, \cdots, X_n$ una muestra aleatoria de una distribución exponencial con densidad
$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x > 0, \quad \theta > 0$$
a) Demuestra que $X_1^2/2$ es insesgado para $g(\theta) = \theta^2$, por lo que $\tilde{\theta}^2 = \sum {X_i}^2/2n$ es también insesgado.

   b) Demuestra que $t(X) = \sum X_i$ es suficiente para θ, y argumenta sobre si es completa.

   c) Encuentra el mejor estimador para $g(\theta)$, denótalo $g(\hat{\hat{\theta}})$.

   d) Exhibe las varianzas de ambos estimadores.

2. Sea $X_1, \cdots, X_n$ una muestra aleatoria con distribución Poisson con parámetro λ_i, es
$$f(x_i;\lambda_i) = \frac{\lambda_i^{x_i} e^{-\lambda_i}}{x_i!} \quad x_i = 0, 1, \cdots, \quad \lambda > 0$$
a) Encuentra el estimador máximo verosímil del vector $\underline{\lambda} = (\lambda_1 \cdots, \lambda_n)$.

   Considera la prueba con hipótesis nula:
$$H_0 \; \lambda_i = \alpha \; d_i, \; \begin{array}{l} d_i \text{ constante conocidas} \\ \alpha > 0 \text{ desconocido} \end{array},$$

y alternativa H_1 que simplemente niega H_0.

b) Encuentra el estimador máximo verosímil de $\underline{\lambda}$ bajo H_0.

c) Encuentra el cociente de verosimilitudes generalizadas para probar si H_0 es cierta.

3. Considera dos muestras independientes (e independientes entre sí); la primera de tamaño n de una $n(\mu_1, {\sigma_1}^2)$, y la segunda de tamaño m de una $n(\mu_2, {\sigma_2}^2)$. El problema clásico de comparación de medias; esto es el probar $H_0 : \mu_1 = \mu_2$ sin suponer que las varianzas son iguales, es el problema denominado, "de Behrens-Fisher". Supón que la hipótesis nula es cierta; esto es, $\mu_1 = \mu_2 (= \mu)$, con μ desconocida.

   a) Describe la verosimilitud e identifica (una) estadística suficiente. Demuestra que la estadística formada por las dos medias muestrales y las dos varianzas muestrales (de dimensión cuatro), es suficiente minimal. Observa como contrasta esto con el hecho de que la dimensión del parámetro bajo H_0, es tres.

   b) Muestra que la estadística suficiente minimal no es completa. Para ello exhibe una función de la estadística suficiente minimal que tenga esperanza cero siempre, pero que no es la función nula.

(Modelos Lineales)

1. Supón que para n observaciones tanto de la variable "dependiente" y y las potenciales variables independientes x_1, x_2 y x_3, la matriz de correlaciones estimadas para el vector $\begin{pmatrix} y \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}$ es $\begin{pmatrix} 1 & .8 & -.4 & 0 \\ .8 & 1 & -.56 & 0 \\ -.4 & -.56 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

   a) Si tuvieras que elegir una sola x_j para la regresión lineal simple $y_i = \alpha + \beta x_j + e_i \quad i = 1, \cdots, n$

   ¿cuál eligirías? (justifica tu respuesta).

   b) Dada la elección que hiciste en a), ahora calcula el coeficiente de correlación parcial entre y y las restantes candidatas y di qué variable adicional eliges para una regresión lineal $y_i = \alpha + \beta x_j + \gamma x_k, + e_i, i = 1, \cdots, n$

c) Habiendo ajustado el modelo en b) puedes exhibir el coeficiente de correlación múltiple (entre y y el par (x_j, x_k))

2. Considera una clasificación cruzada de dos criterios sin interacción con modelo:

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk},$$

con i =1,2, con j =1,2,3 y $k = 1, \cdots, n_{ij}$, con las n_{ij} sencillitas todas iguales a 1 excepto $n_{22} = 0$.

Es decir, se tienen 5 observaciones nada más.

a) Di de las tres cantidades paramétricas siguientes, si son o no son estimables en el sentido Gauss-Markov:

$\tau_1 - \tau_2$, $\beta_1 - \beta_2$ y $\beta_2 - \beta_3$.

b) Ahora repite tus respuestas si en adición nos dicen que $n_{12} = 0$, o sea nos quedaron sólo 4 observaciones.

c) En lugar de tener como en b) $n_{12} = 0$, ese es 1; y el que es cero es n_{13}, otra vez sólo 4 observaciones. Qué respondes respecto a la estimabilidad de las tres cantidades?

3. Considera el modelo de rango completo

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i \qquad i = 1, \ldots, n$$

Sean R_p^2 y R_{p-1}^2 los coeficientes de determinación múltiple del modelo completo y del modelo que sólo considera las primeras $p-1$ variables. Demuestra que

$$\frac{R_p^2 - R_{p-1}^2}{1 - R_{p-1}^2} = \frac{d}{1+d}$$

donde

$$d = \frac{t^2}{n-p}$$

y

$$t^2 = \frac{SCE_{p-1} - SCE_p}{SCE_p}\,(n-p),$$

SCE_p y SCE_{p-1} son las Sumas de Cuadrados del Error del modelo completo y del modelo que sólo considera las primeras $p-1$ variables.

(Inferencia Bayesiana)

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

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

donde $0 < \theta < 1$.

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

(b) Considera la función de pérdida

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

Muestra 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) Encuentra el estimador bayesiano de M correspondiente.

2. Se desea aproximar la densidad conjunta $f(\theta, \phi)$ 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.

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

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

3. Supón que $s_m = \{x_1, \ldots, x_m\}$ es una muestra aleatoria de una distribución Poisson $P(x|\theta)$. Supón además que

$$p(\theta) = \omega_1 Ga(\theta|a_1, b_1) + \omega_2 Ga(\theta|a_2, b_2) + \omega_3 Ga(\theta|a_3, b_3),$$

donde $\omega_1, \omega_2, \omega_3 \geq 0$, $\omega_1 + \omega_2 + \omega_3 = 1$, y $a_1, a_2, a_3, b_1, b_2, b_3 > 0$, todos valores conocidos. Aquí, $Ga(\theta|a, b)$ denota una densidad Gamma con media a/b.

(a) Demuestra que esta familia de distribuciones iniciales es conjugada para la distribución Poisson, encontrando los nuevos pesos y los valores de los hiperparámetros de la distribución final correspondiente.

(b) Supón ahora que Y es condicionalmente independiente de s_m dado θ. Encuentra la densidad predictiva

$$p(y|s_m) = \int_0^\infty p(y|\theta)p(\theta|s_m)\, d\theta.$$

(c) Finalmente, supón que $m = 5$ y $s_m = \{0, 2, 2, 2, 0\}$, y que $(\omega_1, \omega_2, \omega_3) = (0.4, 0.2, 0.4)$, $(a_1, b_1) = (2, 1)$, $(a_2, b_2) = (1, 2)$ y $(a_3, b_3) = (2, 2)$. Usando los resultados anteriores, calcula $p(y = 1|s_m)$.