PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS SOCIALES
ESPECIALIDAD DE ECONOMÍA

ESTADÍSTICA INFERENCIAL
EXAMEN PARCIAL

CÓDIGO: EST-241
PROFESOR: José Flores Delgado
HORARIO: 522
FECHA: 12 de octubre de 2013
DURACIÓN DE LA PRUEBA: tres horas.
Se permite el uso del formulario del curso.
Se calificarán solamente las caras derechas de los cuadernillos.

Ejercicio 1 (3 puntos)

Se tiene que $X \sim exp(8)$ e $Y|X = x \sim P(2x)$, $\forall x > 0$. (1 punto)

a) Halle $E(Y|X = x)$ y $E(XY|X = x)$.

b) Obtenga el modelo condicional de $X|Y = y$, para $y = 0, 1, \ldots$ (2 puntos)

Ejercicio 2 (4 puntos)

Sean $X_1, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con $E(X_1) = 0$; $E(X_1^2) = m_2$; $E(X_1^3) = m_3$; y $E(X_1^4) = m_4$. Determine $V(\sum_{i=2}^{n}(X_i - X_1)^2)$.

$X_1 + X_n = X_2 + \ldots$

Ejercicio 3 (6 puntos)

Considere el modelo de regresión lineal dado por $Y_i = \alpha + \beta x_i + \epsilon_i$, para $i = 1, \ldots, n$, donde $\epsilon_1, \ldots, \epsilon_n$ son variables aleatorias, independientes y cada una con media 0 y varianza σ^2, α y β son constantes desconocidas y $x_1, \ldots, x_n$ constantes conocidas.
Para cada $j \in \{1, \ldots, n\}$, sea

$$\hat{Y}_j = \sum_{i=1}^{n}(a_i + x_j b_i)Y_i,$$

Cov(ε, ε

$1 - \bar{X}(b_1 + \ldots + b_n)$

$\frac{x_1 - \bar{X} + \ldots X_n - \bar{X}}{\sigma}$

con $b_i = \dfrac{x_i - \bar{X}}{\sum_{k=1}^{n}(x_k - \bar{X})^2}$ y $a_i = \dfrac{1}{n} - b_i\bar{X}$, para $i = 1, \ldots, n$.

$\frac{1}{n} - b_1\bar{X} + \frac{1}{n} - b_n\bar{X}$

a) Halle $E(Y_i)$ y $V(Y_i)$, para $i = 1, \ldots, n$. (1 punto)

b) Halle $E(\hat{Y}_j)$ y $V(\hat{Y}_j)$, para $j = 1, \ldots, n$. (1 punto)

c) Halle $Cov(Y_j, \hat{Y}_j)$, para $j = 1, \ldots, n$. (1,5 punto)

d) $V(Y_j - \hat{Y}_j)$, para $j = 1, \ldots, n$. (1 punto)

e) $E(Y_j - \hat{Y}_j)^2$ para $j = 1, \ldots, n$. (0,5 punto)

f) Determine $E(\widehat{\sigma^2})$, donde $\widehat{\sigma^2} = \dfrac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2}{n-2}$. (1 punto)

Ejercicio 4 (4 puntos)

El problema que se describe a continuación es un caso particular de la teoría propuesta en *Cooner et al.* (2007)[1]. El objetivo general es describir y modelar el tiempo hasta la ocurrencia de algún evento de interés, E.
El número de factores de riesgo latentes, que podrían originar la ocurrencia del evento E, es una variable aleatoria discreta positiva, N, con distribución de Poisson truncada con parámetro 5, es decir,

$$f_N(n) = \frac{5^n}{(e^5 - 1)\, n!}, \; n = 1, 2, \ldots$$

Para cada $n \in \{1, 2, \ldots\}$, sea T_n el tiempo de activación del factor de riesgo n (es decir, el tiempo que transcurre hasta que se presenta este factor de riesgo). Estos tiempos $T_1, T_2, \ldots$ son variables aleatorias independientes e independientes de N y cada $T_i \sim expg(3; 2)$, con $expg(\alpha, \beta)$ el modelo probabilístico exponencial generalizado; propuesto en *Gupta* y *Kundu* (1999)[2], que se caracteriza por la función de distribución acumulada siguiente:

$$F(x) = (1 - e^{-\beta x})^{\alpha}, \; x > 0.$$

Además, el evento de interés E ocurre en el instante que se activa el último factor de riesgo; es decir, si T es el tiempo hasta la ocurrencia del evento E; entonces, se cumple que

$$T := \text{máximo}\{T_1, \ldots, T_N\}.$$

Con esta información determine el modelo probabilístico de T.
Siga los pasos siguientes:

i) Para cada $n \in \mathbb{N}^+$, considere $Y_n = \text{máximo}\{T_1, \ldots, T_n\}$ y determine $F_{Y_n}(t)$, $t > 0$. Concluya que $Y_n \sim expg(3n; 2)$.
Note que $Y_n \le t \Leftrightarrow T_1 \le t \cap \ldots \cap T_n \le t$.

ii) Use la definición de la función de distribución acumulada, el principio de sustitución y el resultado del paso anterior para determinar $F_{T|N=n}(t)$, $t > 0$; $n = 1, 2, \ldots$
Concluya que $T|N = n \sim expg(\overset{\alpha}{3n}; \overset{\beta}{2})$, $n = 1, 2, \ldots$

iii) Determine $f_{T|N=n}(t)$, $t > 0$; $n = 1, 2, \ldots$

iv) A partir del modelo marginal de N y el condicional de T dado $N = n$, determine $f_{N,T}(n, t)$, $n \in \mathbb{N}^+$, $t > 0$.

v) Por último, a partir del modelo conjunto anterior determine $f_T(t)$, $t > 0$.

Ejercicio 5 (3 puntos)

Sea el modelo de regresión múltiple $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, donde $\boldsymbol{\epsilon} \sim N_n(\overset{\mu_\epsilon}{\mathbf{0}}; \sigma^2\mathbf{I})$, $\boldsymbol{\beta} = (\beta_j)_{k\times 1}$ es un vector columna de parámetros y $\mathbf{X} = (x_{ij})_{n\times k}$ una matriz no aleatoria de valores conocidos con rango columna completo. El estimador usual de $\boldsymbol{\beta}$ está dador por $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

a) Halle el vector de medias y la matriz de varianzas-covarianzas de $\mathbf{Y}$. (1 punto)

b) Halle el vector de medias y la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\beta}}$. (2 puntos)

J.F. San Miguel, 12 de octubre de 2013.

[1] Cooner, F., Banerjee, S., Carlin, B.P. and Sinha, D. Flexible cure rate modeling under latent activation schemes. Journal of the American Statistical Association 102(478), 560-572.

[2] Gupta and Kundu. Theory and methods: Generalized exponential distributions. Australian and New Zealand Journal of Statistics, 41(2):173-188, 1999.