Examen General de Estadística 2014 I

Enero 23 2014. 9-14 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. 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. Considere una muestra independiente de tamaño n de una $n(0, \sigma^2)$ con σ^2 desconocida.

   a) Encuentre el estimador máximo verosímil de σ^2, verifique que es insesgado, encuentre su varianza y compárela con la cota de Cramér Rao. ¿Es óptimo?

   b) Diga cómo construir un intervalo de confianza para σ al $(1-\alpha)100\%$ utilizando tablas de una distribución ji cuadrada.

2. Para probar (contrastar) hipótesis, en el caso paramétrico, un procedimiento muy utilizado (óptimo en ciertas circunstancias) es el del “cociente de verosimilitudes generalizadas” que consiste en obtener el cociente entre la máxima verosimilitud bajo H_0 (la hipótesis nula) y la máxima verosimilitud no restringida (o sea bajo $H_0 \cup H_1$), y se rechaza cuando ese cociente es chico.

   Considere el siguiente problema de “prueba de hipótesis”: Para $x_1, x_2, \cdots, x_n$ una realización de variables independientes idénticamente distribuidas según la función de distribución F, de la cual lo único que se sabe es que es una discreta sobre el conjunto $\{0, 1, 2, \cdots\}$, se desea probar H_0: $F = F_0$ en que la F_0 es totalmente conocida y corresponde a una distribución que tiene probabilidades $p_0, p_1, \cdots$ en el soporte, con (obviamente), $p_j \geq 0$ para toda j y $\sum_{j\geq 0} p_j = 1$.

   Si Λ denota al cociente de verosimilitudes generalizadas, observe que el numerador es simplemente la verosimilitud bajo la hipótesis nula, que es $\prod_j {p_j}^{N_j}$ con N_j el número de observaciones, $\{x_i's\}$, iguales a j.

   a) Muestre que el denominador de Λ, resultante de maximizar sobre $H_0 \cup H_1$ es:

   $$\prod_j (N_j/n)^{N_j},$$

   con lo cual Λ queda bien definido.

   b) Existe un concepto interesante en probabilidad que es la Divergencia Logarítmica de Kullback-Liebler que mide “que tan lejos” está la densidad discreta dada por las $\{p_j\}'s$ desde el punto de vista de la densidad dada por las, digamos $\{q_j\}'s$ que es:

$$D({p_j}'s;{q_j}'s) = \sum_j q_j \log(q_j/p_j)$$

Se puede demostrar que es nonegativa, es cero sí y solo si las dos densidades son idénticas y requiere, para estar bien definida, que si una $q_j>0$ entonces $p_j>0$ (o sea el soporte de p contiene al soporte de q). Si no es el caso, la divergencia logarítmica no está definida.

c) Muestre que $-\log(\Lambda)$ coincide con $D({p_j}'s; N_j/n's)$, que está bien definida pues si una $N_j/n> 0$ entonces debe ser cierto que la correspondiente $p_j > 0$ (al menos bajo H_0)

d) Muestre (usando un desarrollo de Taylor a segundo orden) que $-2\log(\Lambda)$ resulta en una expresión tipo X^2 de Pearson en que en el denominador aparecen, en lugar de los $n{p_j}'s$, los ${N_j}'s$.

Si el problema de bondad de ajuste fuera continuo; esto es, se sabe que F es una (absolutamente) continua y F_0 es una continua conocida, el uso de Λ no conduce a nada pues su denominador no está definido. Correspondería al máximo sobre todas las densidades f de la verosimilitud $L(f) = \prod_{i=1}^n f(x_i)$, y pueden imaginar f's con valores tan grandes como quieran en los x_i's.

3. Considere 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 $n(\mu_1, {\sigma_1}^2)$. 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". Suponga que la hipótesis nula es cierta; esto es, $\mu_1 = \mu_2 (= \mu)$, con μ desconocida.

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

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

MODELOS LINEALES

1. Sea un modelo de regresión lineal múltiple:

$$Y = X\beta + \epsilon,$$

   con X matríz n por p $(n > p)$.

Para $\hat{\beta}$, solución a las ecuaciones normales, se define el coeficiente de determinación R^2 como el % de variación total, explicada por β, por:

$$R^2 = \frac{\hat{\beta}' X' Y}{||Y||^2}.$$

Este coeficiente de determinación, corresponde al cuadrado de la correlación máxima que se puede obtener entre una combinación lineal de las p columnas de X y el vector Y. Suele cambiarse en el caso en que la regresión tiene un término constante (esto es, cuando una de las columnas de X; sin perder generalidad, la primera, es una columna de 1's), por el coeficiente de determinación "ajustado", denotado por $R^{\star 2}$ está dado por,

$$R^{\star 2} = \frac{\hat{\beta}' X' Y - (\sum_i y_i)^2/n}{||Y||^2 - (\sum_i y_i)^2/n}.$$

Este coeficiente ajustado mide el porcentaje de variación (ya corregida por la presencia de una media) explicado por la presencia de los (p-1) regresores restantes.

a) Para X de rango completo y con un término constante (la primera columna de 1's en X) , muestre que la prueba F que se construye para probar H_0: $\beta_2 =, \beta_3 =, \cdots, \beta_p = 0$, (las $p-1$ últimas componentes del vector β) son todas iguales a cero, es:

$$F = \frac{R^{\star 2}}{(1 - R^{\star 2})} \frac{(n-p)}{(p-1)}.$$

2. Si $\underline{x}_i'$ es la i-ésima fila de X en un modelo de regresión como el descrito en el problema anterior, y $\hat{\epsilon}_i = y_i - \underline{x}_i' \hat{\beta}$ denota al i-ésimo residuo ("pelón", o sea sin estandarizar ni "estudentizar"), muestre que una condición necesaria y suficiente para que $\sum_{i=1}^n \hat{\epsilon}_i = 0$ (para cualquier conjunto de observaciones y_i), es que la matriz X tenga, como una de sus columnas, una de 1's, o de una constante $\neq 0$ que multiplica a una columna de 1's. Se sugiere mostrar la suficiencia de manera directa y la necesidad con un contra ejemplo en que se muestre que la aseveración de que "en cualquier modelo lineal la suma de los residuos pelones es cero", es falsa (mientras más sencillo el contraejemplo, mejor).

3. Para una clasificación cruzada a dos criterios sin interacción, por ejemplo el modelo asociado a un diseño aleatorizado en bloques:

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

con $i = 1, \cdots, t$ y $j = 1, \cdots, b$, y los $\epsilon' s$ independientes no correlacionados de media cero y varianza constante se define un <u>contraste</u> en los τ's como cualquier combinación lineal $\sum_i a_i \tau_i$ con $\sum_i a_i = 0$.

Demuestre que todo contraste en los τ's es estimable (Gauss-Markov) y exhiba el mejor estimador lineal insesgado (MELI) de un contraste dado, así como la expresión para la varianza de ese estimador.

INFERENCIA BAYESIANA

1. Una bola está en una de n cajas y está en la i-ésima caja con probabilidad p_i. Si está en la caja i, una extracción al azar de esa caja descubrirá la bola con probabilidad α_i.

   a) Demuestre que la probabilidad de que la bola esté en la caja j, dado que una extracción al azar de la caja i no la descubrió es

   $$\frac{p_j}{1 - \alpha_i p_i} \qquad (j \neq i).$$

   b) Encuentre la misma probabilidad condicional, pero ahora en el caso $j = i$.

2. Suficiencia

   a) Enuncie y discuta la definición bayesiana de una *estadística suficiente*.

   b) ¿Cuál es la relación entre esta definición y la correspondiente definición clásica?

   c) Sean X y Y dos variables aleatorias con función de densidad conjunta $f(x, y)$. La información mutua de X y Y se define como

   $$I(X;Y) = \int\int f(x,y) \log\left\{\frac{f(x,y)}{f(x)f(y)}\right\} dx\, dy.$$

   Ahora sea $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ una muestra de variables aleatorias i.i.d. con función de densidad $f(x|\theta)$. Por otro lado, sea $f(\theta)$ la función de densidad de la distribución inicial sobre Θ. Finalmente, sea $T = T(\boldsymbol{X})$ una estadística. Demuestre que $I(T(\boldsymbol{X}); \Theta) = I(\boldsymbol{X}; \Theta)$ si y sólo si $T(\boldsymbol{X})$ es suficiente para Θ.

   d) Sea $X_1, \ldots, X_n$ una muestra de observaciones i.i.d. con distribución $\text{Ga}(\alpha, \theta)$, con α conocido. Demuestre, utilizando la definición bayesiana del inciso (a), que $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ es una estadística suficiente para θ.

3. Considere un problema de decisión estadístico con una función de pérdida acotada $L(d, \theta)$ y una función de densidad $p(\theta)$, donde $d \in \mathcal{D}$ y $\theta \in \Theta$. Sea

   $$L^*(d, \theta) = a\,\{L(d,\theta) - \inf_{d \in \mathcal{D}} L(d, \theta)\},$$

   con $a > 0$. Demuestre que la decisión bayesiana óptima bajo la función de pérdida $L^*(d, \theta)$ es la misma que la que se obtiene bajo la función de pérdida $L(d, \theta)$.