Examen General de Estadística Semestre 2015-I

19 de enero de 2015

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** resuelva solamente tres. Es decir resuelva solamente seis preguntas: las tres primeras, y otras tres a ser seleccionadas de entre las últimas seis. Tiempo de máximo examen: cinco horas.

INFERENCIA

1. En el intervalo (0,θ) considera la densidad $f(x;\theta) = a(\theta)c(x)$, con el parámetro θ positivo y desde luego la función $c(x)$ no-negativa.

   Observa que $a(\theta)$ es el recíproco de la integral $\int_0^\theta c(x)dx$, y esa integral por construcción es creciente en θ.

   Demuestra que el estimador máximo verosímil del parámetro θ es $T = max(X_i)$, usualmente denotado $X_{(n)}$. Demuestra también que T es suficiente, y es minimal; esto segundo mostrando que dos muestras equivalentes en verosimilitud (las verosimilitudes, como funciones del parámetro son esencialmente las mismas), necesariamente tienen el mismo máximo.

2. Considera una distribución normal en que la media y la varianza están relacionadas, siendo la varianza igual al cuadrado de la media. Esto es es una normal, $n(\theta, \theta^2)$ con θ real.

   Supón una muestra independiente de tamaño n, $X_1, X_2, \cdots, X_n$ y analizando la verosimilitud $l(\theta, \underline{x})$, para la muestra observada $\underline{x}$, demuestra que el vector $T = (\bar{X}, S^2)$ es suficiente y también minimal, pero no es una estadística completa (la expresión para S^2 es la del estimador máximo verosímil en el caso $n(\mu, \sigma^2)$). Para esto último basta exhibir un ejemplo de una función (no nula) de T cuyo valor esperado sea cero para todo θ real.

   Como comentario, la densidad anterior se conoce como un ejemplo con parámetro "curvo".

3. Considera una muestra independiente de una exponencial negativa, parametrizada como un caso particular de una gamma(α, β) con $\alpha = 1$; esto es con densidad

$$f(x, \beta) = \frac{1}{\beta} e^{-x/\beta}$$

Observa que ${X_1}^2/2$ resulta ser insesgado para β, por lo cual se puede considerar a $\sum_1^n({X_i}^2/2n)$ como un estimador insesgado también, y que utiliza toda la muestra.

Por otro lado, la estadística suficiente minimal (y completa) es $T = \sum_1^n X_i$. Obtén el estimador insesgado de varianza mínima de β, función obviamente de T y compara las varianzas de ambos.

MODELOS LINEALES

1. El término regresión se utiliza casi siempre cuando las observaciones tienen cada una una distribución con parámetro que es una función lineal (con constantes conocidas) de un vector de parámetros desconocidos; y el estudio se centra en estimar los parámetros desconocidos. Lo típico es tener observables Y_i con distribución normal de varianza constante (desconocida) σ^2 y quizás una media de la forma $\alpha+\beta x_i$, con las x_i conocidas pero con α y β parámetros desconocidos (es el caso de un modelo lineal simple; si fuera la esperanza sólo βx_i el modelo es el de una recta que pasa por el orígen).

   Imagina $Y_1, Y_2, \cdots, Y_n$ independientes, cada una distribuida de acuerdo a una Poisson en donde la media de Y_i es expresable como $x_i\lambda$, con las x_i conocidas pero λ desconocida. Se puede escribir la verosimilitud $l(\lambda, \underline{y})$ y obtener el estimador máximo verosímil de λ. Obtén dicho estimador y habrás hecho un ejercicio que en espiritu replica lo que se hace en regresión.

   Imagina que en realidad cada Y_i que sabemos que es Poisson tiene una media expresable como $E(Y_i) = \delta + x_i\lambda$, con δ y λ desconocidos y las x_i's conocidas. Se puede encontrar la verosimilitud $l(\delta, \lambda, \underline{y})$ y con ella pretender encontrar los estimadores máximo-verosímiles. ¿Puedes decir qué ecuaciones habría que resolver?

   Si en el ejercicio anterior sólamente sabemos que el valor esperado de Y_i es λ_i, con las n λ_i's desconocidas, también podemos escribir la verosimilitud y obtener los estimadores máximo verosímiles de cada λ_i. Obtenlos también y observa de que en esta caso no tenemos <u>nada parecido</u> a un planteamiento de regresión.

2. 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 los valores (que en promedio) tomaría la variable (dependiente) y, cuando los valores de las variables independientes están representados por el vector (fila) $\underline{x}^{'}$.

Dí brevemente qué ocurre si la matríz X no es de rango completo. ¿Existe un "estimador" $\hat{\beta}$? ¿o sólo existe una solución al las ecuaciones normales? ¿Es única esa solución?. En adición agrega si $\hat{\sigma^2}$ está bien definida (¿es única?); y finalmente discute si se puede hablar de un "modelo ajustado" para predecir una y cuando los valores de las variables independientes sean los representados en $\underline{x}'$.

¿Podrá usarse el modelo ajustado para cualquier $\underline{x}'$? Argumenta tu respuesta. (Sugerencia: una propiedad de la predicción que se haga usando el modelo ajustado es que sea un estimador insesgado de la media de la distribución de la variable sobre la cual se hace la predicción).

3. Considera
$$y_i = \underline{x}_i'\beta + e_i \quad i = i, \cdots, n \quad, e_i \sim n(0, \sigma^2)$$
con las $\underline{x}_i' \in R^p$, conocidas y los e_i independientes.

Si la matríz X $n \times p$ $(n > p)$ cuyas filas son las $\underline{x}_i'$ es de rango completo, $\hat{\beta}$, solución a las ecuaciones normales, es única y toda combinación lineal $\lambda'\beta$ es estimable (Gauss Markov).

Supón que X no es de rango máximo y que una observación futura e independiente de las n $y_i's$ iniciales, y^f, con
$$y^f = \underline{x^f}'\beta + e^f,$$
se hará; y se desea establecer un intervalo de confianza para su media. Dí cómo se construiría ese intervalo en el caso de que al aumentar con $\underline{x^f}'$; como la $(n+1)$-ésima fila, a la matríz X, su rango aumenta (o sea la fila $\underline{x^f}'$, no pertenece al espacio fila de la matríz original X). Dí si en ese caso, consideras que no es posible construir un intervalo para la media. Argumenta tu respuesta.

ESTADÍSTICA BAYESIANA

1. Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una muestra de observaciones i.i.d. de una distribución Poisson $Po(x|\lambda)$, con $\lambda > 0$. Supongamos que λ tiene una distribución inicial gamma $\mathrm{Ga}(\lambda|\alpha, \beta)$ con $\alpha > 1$ y $\beta > 0$.

   (a) Encuentra la distribución final de λ.

   (b) Calcula
$$\delta(\lambda_1||\lambda_2) = \sum_{x=0}^{\infty} Po(x|\lambda_2)\ \log\left(\frac{Po(x|\lambda_2)}{Po(x|\lambda_1)}\right);$$

es decir, la discrepancia de Kullback-Leibler entre dos modelos Poisson cualesquiera identificados por los valores λ_1 y λ_2.

(c) Encuentra el estimador óptimo bayesiano bajo la función de pérdida

$$L(\hat{\lambda}, \lambda) = \delta(\hat{\lambda}||\lambda)$$

y con respecto a la distribución final de λ. Denota este estimador como $\hat{\lambda}^*_{KL}$.

2. Continuando con el problema anterior:

(a) Encuentra el estimador MAP (*maximum a posteriori*); es decir, el valor de λ que maximiza la densidad final de λ. Denota este estimador como $\hat{\lambda}^*_{MAP}$.

(b) Demuestra que $\hat{\lambda}^*_{MAP}$ también puede obtenerse como el valor de λ que minimiza

$$G(\lambda) = n\,\delta(\lambda||\bar{x}) + \beta\,\delta(\lambda||\lambda^*_0),$$

donde $\bar{x}$ denota la media muestral y $\lambda^*_0 = (\alpha - 1)/\beta$ es la moda de la distribución inicial de λ.

(c) Demuestra que $\hat{\lambda}^*_{KL}$ y $\hat{\lambda}^*_{MAP}$ son asintóticamente equivalentes cuando $n \to \infty$.

3. Sea $\{X_1, \ldots, X_n\}$ una muestra de observaciones i.i.d. de la distribución normal $\mathrm{N}(x|\lambda, \lambda)$, con $\lambda > 0$, de manera que la media y la varianza son iguales.

(a) Encuentra la distribución inicial de Jeffreys para λ.

(b) Muestra que este modelo pertenece a la familia exponencial, especificando la estadística suficiente minimal así como el parámetro canónico correspondiente.

(c) Determina la forma del *kernel* de las distribuciones conjugadas para λ. ¿Se puede obtener la distribución inicial de Jeffreys como límite de estas distribuciones conjugadas?