Examen de Inferencia
UTDT – Septiembre 2023

Instrucciones: El examen tiene una duración de 2 horas y 45 minutos, y es individual. Debes **justificar todas tus respuestas**; la puntuación asignada dependerá del nivel de profundidad con que plantees y respondas a cada ejercicio. *Buena suerte*!

Nombre y Apellido: .... **Legajo**:...........................

1. (45pts) Considere $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ donde X tiene la siguiente función de densidad:

$$f(x;\theta) = \frac{1}{6\theta^4} x^3 e^{-\frac{x}{\theta}}, \qquad \text{para } x > 0 \text{ y con } 0 < \theta < \infty.$$

Luego se cumple que $E(X) = 4\theta$ y $E(X^2) = 20\theta^2$ (no debe demostrar estas igualdades).

(a) (2pts) Demuestre que X sigue un modelo estadístico que pertenece a la familia exponencial e indique (dar la expresión de) un estadístico suficiente para el parámetro θ.

(b) (5pts) Obtener el estimador de momentos $\widetilde{\theta}_n$ del parámetro θ y demostrar que es consistente.

(c) (5pts) Dar la expresión del error estandard de $\widetilde{\theta}_n$ (si no resolvió el punto anterior, considere $\widetilde{\theta}_n = \overline{X}_n/2$).

(d) (7pts) Demostrar que $\widehat{\theta}_n = \frac{1}{4n}\sum_{i=1}^n X_i$ es el estimador máximo verosímil de θ. No olvide verificar la condición de segundo orden (condición suficiente) para un máximo.

(e) (5pts) Calcule la función de error cuadrático medio de $\widehat{\theta}_n$.

(f) (7pts) Indicar si el estimador máximo verosímil es el estimador UMVUE de θ (si no resolvió el punto anterior, asuma de aquí en adelante que: $\widehat{\theta}_n$ es insesgado para θ y que $Var(\widehat{\theta}_n) = \theta^2/n$).

(g) (5pts) Sabiendo que se cumplen las condiciones para que $\sqrt{n}(\widehat{\theta}_n - \theta) \to_d N(0, v_\theta)$ (donde v_θ es la varianza asintótica de $\widehat{\theta}_n$) y que ud. está interesado en estimar $\psi \equiv \ln(\theta)$, se pide:

   i. Indique la expresión que tiene $\widehat{\psi}_n$ (el estimador máximo verosímil del parámetro ψ).

   ii. Determine a donde converge en distribución $\sqrt{n}(\widehat{\psi}_n - \psi)$ indicando claramente los parámetros de la distribución límite.

Nota: *Si no pudo resolver el inciso anterior, utilice de aquí en adelante que* $v_\psi = 1/4$*, donde la cantidad* v_ψ *representa el valor de la varianza asintótica del estimador* $\widehat{\psi}_n$.

(h) (5pts) De una muestra de tamaño $n = 100$ se obtuvo que $\sum_{i=1}^{100} x_i = 400$. Construya el intervalo de confianza de nivel aproximado del 95% (centrado en $\widehat{\psi}_n$) para el parámetro ψ.

(i) (4pts) Indique si las afirmaciones siguientes son verdaderas o falsas (no hace falta justificar):

- "*La probabilidad de que el parámetro desconocido* ψ *pertenezca al intervalo que construiste con los datos del punto (h) es de aproximadamente* 0.95".
- "*Con los datos del punto (h), para el test de Wald con nivel de significatividad aproximado de* $\alpha = 0.05$ *no se alcanza a rechaza la hipótesis nula* $H_0 : \psi = 0$".

2. (25pts) Sea $\underline{X} = \{X_1, \ldots, X_n\} \overset{iid}{\sim} N(\theta, \theta^2)$ con $\theta > 0$; se pretende contrastar $H_0 : \theta = 1$ vs $H_1 : \theta \neq 1$ utilizando el estadístico de contraste $T_n(\underline{X}) = \sqrt{n}(\overline{X}_n - \theta)/\theta$ y la función de test:

$$\delta = \begin{cases} 1 & \text{si } |T_n(\underline{X})| \geq 0.5, \\ 0 & \text{en otro caso.} \end{cases}$$

   (a) (3pts) Determinar el nivel de significatividad α del test.

   (b) (8pts) Hallar la expresión general de la función $\beta(\theta)$ para todo $\theta \neq 1$.

   (c) (8pts) De una muestra de tamaño $n = 16$ se observó que $\overline{x}_n = 1.5$:

      i. Indique si el estadístico de contraste cae en la zona de rechazo.

      ii. Computa el p–valor asociado a esta muestra. ¿Se alcanzaría a rechazar H_0 si tu nivel de significatividad para el test fuera de $\alpha = 0.01$?

   (d) (6pts) Piense ahora que este mismo contraste se llevó a cabo múltiples veces en $m = 6$ poblaciones independientes. Lo p-valores registrados para éstos test fueron los siguientes:

$$p_1 = 0.045,\ p_2 = 0.001,\ p_3 = 0.265,\ p_4 = 0.005,\ p_5 = 0.033, \text{ y } p_6 = 0.195,$$

   donde p_i indica el p-valor relativo al test i-ésimo. Utilizando la metodología propuesta por Benjamini–Hochberg indique cuales de los test deberían ser rechazados si pretendemos acotar la tasa de falsos rechazos (FDR) en el 5%.

3. (30pts) La variable aleatoria discreta X representa la cantidad de días que trascurren desde que un individuo se contagia de cierta enfermedad hasta que aparece el primer síntoma de la misma. Asumiendo un modelo Geométrico para X, esto es:

$$X \sim f(x; \theta) = \begin{cases} (1-\theta)^x \theta, & \text{con } 0 < \theta < 1 \text{ y } x \in \{0, 1, 2, \ldots\}, \\ 0 & \text{en otro caso,} \end{cases}$$

se pretende hacer inferencia Bayesiana para el parámetro $E(X) = (1-\theta)/\theta$ (lo que se pretende estimar es la cantidad promedio de días que toma la aparición del primer síntoma de la enfermedad). Se elige como prior para θ una densidad Beta(α, β), esto es:

$$\theta \,|\, \boldsymbol{\eta} = (\alpha, \beta) \sim \pi(\theta; \boldsymbol{\eta}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \text{ para } 0 \leq \theta \leq 1,$$

donde $\boldsymbol{\eta} = (\alpha, \beta)$ son los hiperparámetros del modelo. Luego se tiene que: $E_\eta(\theta) = \alpha/(\alpha+\beta)$, $\text{Mo}_\eta(\theta) = (\alpha-1)/(\alpha+\beta-2)$ si $\beta, \alpha > 1$ y $V_\eta(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ (son la esperanza, moda y varianza a priori respectivamente). Con esta información responder a las siguientes preguntas:

   (a) (8pts) Dada una muestra $\mathcal{D}_n = \{X_1 = x_1, \ldots, X_n = x_n\}$ y sabiendo que el modelo planteado en el enunciado es un modelo conjugado, identificar la distribución a posteriori $\pi(\theta \,|\, \mathcal{D}_n, \boldsymbol{\eta})$ y computar las expresiones de la media, moda y la varianza a posteriori de θ.

   (b) (5pts) Indicar a donde convergen (en probabilidad) la media a posteriori cuando el tamaño de la muestra $n \to \infty$ (Ayuda: Recuerde que por la LGN se tiene que $\overline{X}_n \to_P E(X)$).

   (c) (7pts) Si elegimos $\alpha = \beta = 1$ la prior resulta no informativa. Dar la expresión de la moda a posteriori de θ e indicar que relación tiene ésta con el estimador máximo verosímil de θ.

   (d) (10pts) Una investigadora que trabaja con este modelo, se plantea las siguientes hipótesis: $H_0 : E(X) \geq 2$ vs $H_1 : E(X) < 2$. Con los datos de la muestra $\mathcal{D}_n$ y la prior $\boldsymbol{\eta}_0$ que eligió la investigadora para codificar sus creencias sobre θ se calcularon las siguientes probabilidades:

$$\int_{1/3}^{1} \pi(\theta; \boldsymbol{\eta}_0) \mathtt{d}\theta = 0.25, \quad \text{y} \quad \int_{1/3}^{1} \pi(\theta \,|\, \mathcal{D}_n, \boldsymbol{\eta}_0) \mathtt{d}\theta = 0.95.$$

   Explique que representan estas dos probabilidades y como se resuelve el test utilizando las mismas. Indique también a que conclusiones arriba la investigadora con sus datos y sus creencias iniciales sobre el parámetro θ.

Anexo: Si $Z \sim N(0,1)$, luego

- $P(Z \leq 2.10) = 0.982, \quad P(Z \leq -2.10) = 0.007.$
- $P(Z \leq 2.00) = 0.977, \quad P(Z \leq -2.00) = 0.023.$
- $P(Z \leq 1.96) = 0.975, \quad P(Z \leq -1.96) = 0.025.$
- $P(Z \leq 1.64) = 0.950, \quad P(Z \leq -1.64) = 0.050.$
- $P(Z \leq 0.52) = 0.698, \quad P(Z \leq -0.52) = 0.301.$
- $P(Z \leq 0.50) = 0.698, \quad P(Z \leq -0.50) = 0.301.$
- $P(Z \leq 0.37) = 0.691, \quad P(Z \leq -0.37) = 0.309.$
- $P(Z \leq 0.35) = 0.636, \quad P(Z \leq -0.35) = 0.363.$