Inferencia

Instrucciones: El examen tiene una duración de 3 horas, es a libro cerrado e individual; solo puedes consultar tu hoja de fórmulas. Debes justificar tus respuestas; la puntuación dependerá del nivel de profundidad con que plantees y respondas los ejercicios. *Buena suerte*!

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

1. (25pts) Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ donde X sigue una distribución normal $N(0, \theta)$ con media $E(X) = \mu = 0$ conocida y **varianza** $Var(X) = \theta > 0$ desconocida. Recuerde que la densidad de la variable aleatoria $X \sim N(0, \theta)$ se escribe como:

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/(2\theta)},$$

y además se cumple (ud. no necesita demostrar) que: $E(X^2) = \theta$ y $E(X^4) = 3\theta^2$.

   (a) (5pts) Demostrar que el estimador de máxima verosimilitud de θ es $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

   (b) (5pts) Demostrar que $\widehat{\theta}_n$ es insesgado y consistente para θ.

   (c) (10pts) Demostrar que la información de Fisher asociada al modelo estadístico vale $I_1(\theta) = \frac{1}{2\theta^2}$ y que $\widehat{\theta}_n$ es el estimador insesgado de varianza uniformemente mínima (UMVUE) de θ.

   (d) (5pts) De una muestra de tamaño $n = 100$ se tiene que $\sum_{i=1}^{100} x_i^2 = 500$. Con esta información construya un intervalo de confianza aproximada del 95% para θ centrado en $\widehat{\theta}_n$.

2. (25pts) Sea $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, donde $\sigma^2 = 4$; se pretende contrastar

$$H_0 : \mu = 1 \text{ vs } H_1 : \mu \neq 1$$

utilizando como estadístico de contraste $W_n \equiv \sqrt{n}(\overline{X}_n - \mu)$.

   (a) (3pts) Tomando en cuenta que $\frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ cuando $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$; indique qué distribución sigue el estadístico de contraste W_n del test bajo la hipótesis nula.

   (b) (7pts) Determine el nivel de significatividad α del test si la región de rechazo fuera

$$\mathcal{R} = \{\overline{X}_n : |\sqrt{n}(\overline{X}_n - 1)| \geq 3\}.$$

   (c) (9pts) De una muestra de tamaño $n = 16$ se observó que $\overline{x}_n = 2$:

      i. (3pts) ¿Hay evidencia suficiente para rechazar H_0 si se usa la regla dada en el inciso (b)?

      ii. (3pts) Calcule el p–valor asociado a esta muestra.

      iii. (3pts) En base a la respuesta del inciso anterior: ¿Existe evidencia suficiente para rechazar H_0 si el nivel de significatividad para el test fuera de $\alpha = 0.01$?

   (d) (6pts) Calcule la probabilidad de cometer el error de tipo II cuando $\mu = 1.5$ para la regla de decisión del inciso (b) y el tamaño de muestra $n = 16$.

3. (20pts) Considere una muestra $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Unif}(\theta, 1)$ donde θ es un parámetro desconocido que verifica que $0 < \theta < 1$; luego la densidad del modelo estadístico se escribe como:

$$f(x;\theta) = \frac{1}{1-\theta}\mathbb{1}_{\{[\theta,1]\}}(x),$$

y se tiene que $E(X) = \frac{1+\theta}{2}$ y $Var(X) = \frac{(1-\theta)^2}{12}$.

(a) (10pts) Grafique la función de verosimilitud para demostrar que el estimador máximo verosímil del parámetro θ es $\widehat{\theta}_n = \min\{X_1, \ldots, X_n\}$. **Sólo se considerará puntaje cuando se justifique detalladamente cómo se construye la función de verosimilitud.**

(b) (7pts) Demuestre que el estimador de momentos de θ (basado en el primer momento poblacional) es $\widetilde{\theta}_n = 2\overline{X}_n - 1$ y construya la función de error cuadrático medio de dicho estimador.

(c) (3pts) Siendo los valores observados de X_i en una muestra de tamaño $n = 5$:

$$x_1 = 0.52, \quad x_2 = 0.78, \quad x_3 = 0.10, \quad x_4 = 0.33, \quad x_5 = 0.89$$

Indicar las estimaciones de los parámetros $\mu \equiv E(X)$ y $\sigma^2 \equiv Var(X)$ utilizando el **estimador de máxima verosimilitud** $\widehat{\theta}_n = \min\{X_1, \ldots, X_n\}$.

4. (30pts) Para hacer inferencia Bayesiana respecto del parámetro $\mu \equiv E(X)$, se asume un modelo geométrico para la variable aleatoria de interés X; luego la densidad del modelo se escribe como:

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

donde $0 < \theta < 1$ es un parámetro desconocido y se cumple que $\mu(\theta) = E(X) = (1-\theta)/\theta$. Para la prior se elije: $\theta \sim \text{Beta}(\eta)$, donde $\eta = (\alpha, \beta)$ es un vector de hiperparámetros:

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

por lo que media y varianza **apriori** valen $E(\theta) = \frac{\alpha}{\alpha+\beta}$ y $Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ respectivamente.

(a) (10pts) Determinar la forma general de la distribución a posteriori para θ.

(b) (5pts) Computa las expresiones de la media y la varianza a posteriori de θ como funciones de los hiperparámetros η y de los datos de la muestra $\mathcal{D}_n = \{X_1, \ldots, X_n\}$.

(c) (5pts) ¿Es correcto afirmar que se trata de un modelo Bayesiano conjudago? ¿Porqué?

(d) (10pts) Se pretende contrastar las hipótesis $H_0 : E(X) \le 1$ vs $H_1 : E(X) > 1$. Con los datos de la muestra $\mathcal{D}_n$ y la prior $\boldsymbol{\eta}_0$ que eligió para codificar las creencias sobre θ se calcularon las siguientes probabilidades:

$$\int_0^{1/2} \pi(\theta;\boldsymbol{\eta}_0)\mathtt{d}\theta = 0.25, \quad \text{y} \quad \int_0^{1/2} \pi(\theta \,|\, \mathcal{D}_n, \boldsymbol{\eta}_0)\mathtt{d}\theta = 0.95,$$

donde $\pi(\theta;\boldsymbol{\eta}_0)$ representa la distribución apriori y $\pi(\theta \,|\, \mathcal{D}_n, \boldsymbol{\eta}_0)$ la distribución aposteriori. Explique que representan estas dos probabilidades y cómo deben ser empleadas para resolver el contraste planteado. Indique también a que conclusión llega respecto del parámetro $E(X)$.

Cuantiles de $Z \sim N(0,1)$

$$
\begin{array}{rl}
P(Z \leq -4.00) \approx 0 & P(Z \leq 4.00) \approx 1 \\
P(Z \leq -2.50) = 0.006 & P(Z \leq 2.50) = 0.994 \\
P(Z \leq -2.00) = 0.023 & P(Z \leq 2.00) = 0.977 \\
P(Z \leq -1.96) = 0.025 & P(Z \leq 1.96) = 0.975 \\
P(Z \leq -1.76) = 0.039 & P(Z \leq 1.76) = 0.961 \\
P(Z \leq -1.64) = 0.050 & P(Z \leq 1.64) = 0.950 \\
P(Z \leq -1.50) = 0.067 & P(Z \leq 1.50) = 0.933 \\
P(Z \leq -1.41) = 0.079 & P(Z \leq 1.41) = 0.921 \\
P(Z \leq -1.15) = 0.125 & P(Z \leq 1.15) = 0.875 \\
P(Z \leq -1.07) = 0.142 & P(Z \leq 1.07) = 0.858 \\
P(Z \leq -1.00) = 0.159 & P(Z \leq 1.00) = 0.841 \\
P(Z \leq -0.94) = 0.174 & P(Z \leq 0.94) = 0.826 \\
P(Z \leq -0.67) = 0.251 & P(Z \leq 0.67) = 0.749 \\
P(Z \leq -0.50) = 0.301 & P(Z \leq 0.50) = 0.698 \\
P(Z \leq -0.25) = 0.401 & P(Z \leq 0.25) = 0.599 \\
P(Z \leq -0.10) = 0.461 & P(Z \leq 0.10) = 0.539 \\
P(Z \leq -0.05) = 0.481 & P(Z \leq 0.05) = 0.520 \\
P(Z \leq -0.025) = 0.490 & P(Z \leq 0.025) = 0.510
\end{array}
$$