Certamen 1 MAT-360: INFERENCIA ESTADÍSTICA

Problema 1. [30 puntos] Suponga que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño n proveniente de una familia con función de cuantía dada por

$$f(x,\theta) = \begin{cases} \theta, & x = -1 \\ (1-\theta)^2\theta^x, & x = 0, 1, 2, \ldots \end{cases}$$

donde $0 < \theta < 1$.
a) Encuentre un estimador suficiente minimal para θ.
b) Es su estimador encontrado en a) completo?
c) Encuentre el estimador máximo verosímil de θ.

Problema 2. [25 puntos] Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una población normal con media μ y varianza σ^2, ambos parámetros desconocidos. Encuentre un estimador insesgado de ψ_p, el quantil de orden $100p$ de la población y calcule su varianza.

Hint: El cuantil de orden p satisface la ecuación $P[X \leq \psi_p] = p$.

Problema 3. [25 puntos] Suponga que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de tamaño n proveniente de una familia con función de densidad dada por

$$f(x,\theta) = \frac{a}{\theta^a} x^{a-1}, 0 < x < \theta,$$

donde $a > 0$ es una constante conocida.
a) Encuentre un estadísto suficiente y completo para θ.
b) Calcule

$$\mathbb{E}\left[\frac{X_{(1)}}{X_{(n)}}\right]$$

Problema 4. [20 puntos] Sea $p = \{p_i\}$ una distribución de probabilidad y $q = \{q_i\}$ el verdadero modelo, entonces la distancia de Kulback-Leibler se define como:

$$D_{KL}(p||q) = \sum_i p_i \log(\frac{p_i}{q_i}).$$

a) Demuestre que $D_{KL}(p||q) \geq 0$.
b) Suponga que observamos las frecuencias de un histograma, digamos $c = \{c_i\}$, donde $n = \sum_i c_i$. Este histograma mide la frecuencia de cada clase. El histograma normalizado cuenta las frecuencias relativas $\frac{c_i}{n}$ como una distribución de probabilidad $p_i = \frac{c_i}{n}$.

Supongamos que la probabilidad de observar un histograma $c = \{c_i\}$ sigue una distribución de probabilidad multinomial con respecto a un verdadero modelo $q = \{q_i\}$. Es decir,

$$L(c|q) = \frac{n!}{\prod_i c_i!} \prod_i q_i^{c_i}.$$

Note que $n = \sum_i c_i$ es el número total de observaciones. Definamos el promedio de verosimilitudes como $\overline{L} = \log L(c|q)^{\frac{1}{n}}$. Demuestre que para n grande

$$\overline{L} = -D_{KL}(p||q).$$