PROBABILIDAD Y ESTADÍSTICA
EXAMEN JULIO 2005

DATOS DEL ESTUDIANTE

No. de Examen	Apellidos y Nombres	Cédula

- **La duración total del examen es de 3 horas y media.**
- **El puntaje mínimo para aprobar es 50 puntos.**
- **Publicación de resultados: viernes 5 de agosto - 14:00 hs.**
- **Muestra de exámenes: martes 9 de agosto - 16:00 hs.**

Problema 1 (35 puntos)

Considere una moneda cargada de manera que

$$\mathbf{P}(\text{Cara}) = q \quad \text{y} \quad \mathbf{P}(\text{Número}) = 1 - q,$$

donde $q \in (0, 1)$.

Sea X la variable aleatoria que cuenta el número de veces que hay que tirar la moneda hasta que hayan salido los dos resultados (por ejemplo, si la secuencia de tiradas es: $CCCN$, entonces $X = 4$; si la secuencia de tiradas es: NNC, entonces $X = 3$).

a) **(10 puntos)** Calcule la función de probabilidad de la variable X.

b) **(8 puntos)** Calcule la probabilidad de que la primer tirada haya sido *Cara* dado que $X = 10$.

c) **(10 puntos)** Pruebe que el valor esperado de X es

$$\mathbf{E}(\mathbf{X}) = \frac{1}{q(1-q)} - 1.$$

d) **(7 puntos)** Suponga que se repite n veces el procedimiento y que el número medio de tiradas hasta que hayan salido los dos resultados es $\overline{X}_n = 5.25$. Estime el valor del parámetro q, sabiendo que es más probable obtener *Cara* que *Número*.

Problema 2 (30 puntos)

Considere una sucesión de variables aleatorias $X_1, X_2, \ldots, X_n$ i.i.d. con distribución uniforme en el intervalo $[\theta - 1, \theta + 1]$, donde θ es un parámetro desconocido. Denotamos $\overline{X}_n$ el promedio de estas variables.

a) **(7 puntos)** Calcule $\mathbf{E}(\overline{X}_n)$ y $\mathbf{Var}(\overline{X}_n)$.

b) **(8 puntos)** Construya un estimador T_n del parámetro θ que sea consistente e insegado y calcule su error cuadrático medio, esto es, $\mathbf{E}([T_n - \theta]^2)$.

c) **(8 puntos)** Sabiendo que $n = 300$ y $\overline{X}_n = 2.5$, construya un intervalo de confianza aproximado al 95% para el parámetro θ.

d) **(7 puntos)** Calcule $\mathbf{P}\left(\max\{X_1, X_2, \ldots, X_n\} \leq \theta + \frac{1}{2}\right)$.

Problema 3 (35 puntos)

Nota: En todas las pruebas de hipótesis utilice el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a 0.1

Los siguientes datos corresponden a los tiempos de duración (medidos en meses) de lamparitas marca ACME:

3.33	6.71	2.53	7.15	16.82	3.56	0.17	2.15	3.51	7.24

a) **(7 puntos)** Implemente el test de correlación de rangos de Spearman para estudiar la aleatoriedad de la muestra.

b) **(10 puntos)** Estudie si es razonable suponer que los datos corresponden a una distribución exponencial (realice una sola prueba de hipótesis).

c) **(8 puntos)** ¿Es razonable suponer que la mediana de la muestra es $m = 4$? (realice una sola prueba de hipótesis).

d) **(10 puntos)** Se dispone ahora de una nueva muestra *i.i.d. e independiente* de la anterior (no hay que verificar estos supuestos), correspondiente a tiempos de duración de lamparitas de marca desconocida:

5.6	3.03	1.46	16.52	0.37	0.58	3.38	9.15	0.9	4.44

Implemente un test de comparación de muestras para concluir si es razonable suponer que los nuevos datos tienen la misma distribución que los anteriores.

SOLUCIÓN

Problema 1

a) Consideramos los sucesos

$$\begin{aligned} C &= \text{“salió Cara en el primer lanzamiento”} \\ N &= \text{“salió Número en el primer lanzamiento”} \end{aligned}$$

Luego, para $n \geq 2$:

$$\begin{aligned} \mathbf{P}(X = n) &= \mathbf{P}([X = n] \cap C) + \mathbf{P}([X = n] \cap N) \\ &= \mathbf{P}([X = n] | C)\mathbf{P}(C) + \mathbf{P}([X = n] | N)\mathbf{P}(N). \end{aligned}$$

Ahora bien, si salió Cara en el primer lanzamiento, para que X sea igual a n hay que tener Cara en los siguientes $n-2$ lanzamientos y en el último debe salir Número, esto es

$$\mathbf{P}([X = n] | C) = q^{n-2}(1-q)$$

De la misma manera

$$\mathbf{P}([X = n] | N) = (1-q)^{n-2} q.$$

Así que

$$\mathbf{P}(X = n) = q^{n-1}(1-q) + (1-q)^{n-1} q \qquad \forall n \geq 2. \tag{1}$$

Otra manera: Si $X = n$, para $n \geq 2$, entonces ha ocurrido uno u otro de los siguientes eventos mutuamente excluyentes:

- $A = CCCC \ldots N$, esto es, $n-1$ Caras y al final Número,
- $B = NNNN \ldots C$, esto es, $n-1$ Números y al final Cara.

Está claro que $\mathbf{P}(A) = q^{n-1}(1-q)$ y que $\mathbf{P}(B) = (1-q)^{n-1}q$. Por lo tanto:

$$\mathbf{P}(X = n) = \mathbf{P}(A) + \mathbf{P}(B) = q^{n-1}(1-q) + (1-q)^{n-1} q \qquad \forall n \geq 2.$$

b) De la definición de probabilidad condicional tenemos

$$\mathbf{P}(C | [X = 10]) = \frac{\mathbf{P}(C \cap [X = 10])}{\mathbf{P}([X = 10])},$$

de manera que usando (1) y las probabilidades calculadas queda:

$$\mathbf{P}(C | [X = 10]) = \frac{q^9(1-q)}{q^9(1-q) + (1-q)^9 q}.$$

c)

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{n=2}^{\infty} n\mathbf{P}(X=n) = \sum_{n=2}^{\infty} n\left[q^{n-1}(1-q)+(1-q)^{n-1}q\right] \\
&= \sum_{n=2}^{\infty} nq^{n-1}(1-q) + \sum_{n=2}^{\infty} n(1-q)^{n-1}q.
\end{aligned}$$

Sabemos que si $Y \sim Geo(p)$, entonces

$$\mathbf{E}(Y) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = \frac{1}{p},$$

por lo tanto

$$\sum_{n=2}^{\infty} n(1-p)^{n-1}p = \frac{1}{p} - p.$$

Usando este resultado obtenemos

$$\mathbf{E}(X) = \frac{1}{(1-q)} - (1-q) + \frac{1}{q} - q = \frac{1}{q(1-q)} - 1$$

d) Por el método de los momentos

$$\mathbf{E}(X) = \overline{X}_n,$$

esto es

$$\frac{1}{q(1-q)} - 1 = 5.25 \Leftrightarrow \begin{cases} q = 0.2 \\ q = 0.8 \end{cases}$$

Por otra parte sabemos que es más probable obtener *Cara* que *Número* , así que se cumple

$$q > 1-q \Leftrightarrow \frac{1}{2} < q,$$

de manera que

$$q = 0.8$$

Problema 2

a)

$$\begin{aligned}
\mathbf{E}(\overline{X}_n) &= \mathbf{E}\left(\frac{X_1+X_2+\ldots+X_n}{n}\right) = \frac{1}{n}\left(\mathbf{E}(X_1)+\mathbf{E}(X_2)+\ldots+\mathbf{E}(X_n)\right) \\
&= \frac{1}{n}n\mathbf{E}(X) = \mathbf{E}(X) = \frac{\theta+1+\theta-1}{2} = \theta.
\end{aligned}$$

Por otra parte, usando que las variables son independientes:

$$\begin{aligned}
\mathbf{Var}(\overline{X}_n) &= \mathbf{Var}\left(\frac{X_1+X_2+\ldots+X_n}{n}\right) = \frac{1}{n^2}\left(\mathbf{Var}(X_1)+\mathbf{Var}(X_2)+\ldots+\mathbf{Var}(X_n)\right) \\
&= \frac{1}{n}\mathbf{Var}(X) = \frac{1}{n}\frac{(\theta+1-(\theta-1))^2}{12} = \frac{1}{3n}.
\end{aligned}$$

b) Para que T_n sea un estimador insesgado y consistente de θ debe cumplir $\mathbf{E}(T_n) = \theta$ y $T_n \to \theta$. El promedio cumple la primer condición, como hemos visto en la parte anterior, y la segunda es precisamente la Ley Fuerte de los Grandes Números. Por lo tanto tomamos

$$T_n = \overline{X}_n.$$

Con esto, el error cuadrático medio resulta:

$$\mathbf{E}([T_n - \theta]^2) = \mathbf{E}([\overline{X}_n - \mathbf{E}(\overline{X}_n)]^2) = \mathbf{Var}(\overline{X}_n) = \frac{1}{3n}.$$

c) Por el Teorema Central del Límite sabemos que

$$\mathbf{P}\left(-z_{\frac{\alpha}{2}} \leq \frac{\overline{X}_n - \mathbf{E}(\overline{X}_n)}{\sqrt{\mathbf{Var}(\overline{X}_n)}} \leq z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha$$

y por lo tanto

$$\mathbf{P}\left(\overline{X}_n - \sqrt{\frac{1}{3n}} z_{\frac{\alpha}{2}} \leq \theta \leq \overline{X}_n + \sqrt{\frac{1}{3n}} z_{\alpha}\right) \approx 1 - \alpha.$$

En este caso

$$1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05$$

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.95.$$

Así, un intervalo de confianza aproximado al 95% para el parámetro θ es

$$2.5 - \sqrt{\frac{1}{3(300)}}\,(1.96) \leq \theta \leq 2.5 + \sqrt{\frac{1}{3(300)}}\,(1.96)\,,$$

esto es

$$I = [2.4347 \;\;,\;\; 2.5653].$$

d)

$$\mathbf{P}\left(\max\{X_1, X_2, \ldots, X_n\} \leq \theta + \frac{1}{2}\right) = \left(\mathbf{P}\left(X \leq \theta + \frac{1}{2}\right)\right)^n = \left(\int_{\theta-1}^{\theta+\frac{1}{2}} \frac{1}{2} dx\right)^n = \left(\frac{3}{4}\right)^n.$$

Problema 3

a) **Test de rangos de Spearman**

Se ordenan los datos de menor a mayor. La tabla muestra los datos ordenados X_i^*, los índices i (esto es, el lugar que ocupaban los datos originalmente) y los rangos $R(X_i)$.

X_i^*	i	$R(X_i)$
0.17	7	1
2.15	8	2
2.53	3	3
3.33	1	4
3.51	9	5
3.56	6	6
6.71	2	7
7.15	4	8
7.24	10	9
16.82	5	10

El estadístico del test resulta

$$r_s = 1 - 6\frac{\sum_{i=1}^{10}(R(X_i) - i)^2}{10(10^2-1)} = 0.0061$$

La tabla del test para $n = 10$ muestra que $\alpha^* = 0.5$. Es razonable, entonces, suponer que la muestra es aleatoria.

b) Para estudiar si es razonable suponer que los datos corresponden a la distribución exponencial realizamos el test de ajuste de Lilliefors.

Test de Lilliefors

El estadístico del test es

$$D_n = \sup_{t\in\mathbf{R}} |F(t) - F_n(t)|$$

donde $F_n(t)$ es la función de distribución empírica de la muestra y $F(t)$ es la distribución exponencial de parámetro $\lambda = 1/\overline{X}_{10} = 0.18808$. En este caso el valor que se obtiene para el estadístico es $D_n = 0.2326$.

La tabla del test de Lilliefros para $n = 10$ nos da $\alpha^* > 0.20$ y por lo tanto es razonable suponer que la muestra ajusta a la distribución $\exp(0.18808)$.

c) Realizamos el Test de Signos para estudiar si es razonable suponer que la mediana de los datos es $m = 4$.

El estadístico del test es

$$U = \sum_{i=1}^{10} \mathbf{1}_{\{X_i \leq m\}}$$

En este caso $U = 6$, y utilizando la tabla de la distribución Bin(10,0.5) o la tabla del test de signos se tiene que $\alpha^* = 0.8281$. Por lo tanto es razonable suponer que la mediana es $m = 4$.

d) **Test de Kolmogorov-Smirnov de comparación de muestras**

El estadístico del test es

$$D = \sup_{t \in \mathbf{R}} |F_m^X(t) - F_n^Y(t)|$$

donde ${F_m}^X(t)$ y $F_n^Y(t)$ son las funciones de distribución empírica de la primera y de la segunda muestra respectivamente. El valor que se obtiene para el estadístico es $D = 0.30$

La tabla del test para $m = n = 10$ nos da $\alpha^* > 0.20$ y por lo tanto es razonable suponer que las dos muestras tienen la misma distribución.