PROBABILIDAD Y ESTADÍSTICA
EXAMEN DICIEMBRE 2004

DATOS DEL ESTUDIANTE

Número de Examen	Nombre	Cédula

- La duración del examen es de **3 horas y media**.
- El puntaje mínimo para salvar el examen es de **50 puntos**.
- **Publicación de resultados:** Miércoles 22 de diciembre - 12:00hs.
- **Muestra de exámenes:** Miércoles 22 de diciembre - 18:00hs.

Problema 1 (30 puntos)

Dos apostadores, A y B, arrojan cada uno un dado: el dado A (esto es, el dado arrojado por el apostador A) es un dado equilibrado, mientras que el dado B (el dado arrojado por el apostador B) tiene la siguiente distribución de probabilidad:

$$\mathbf{P}(1) = \mathbf{P}(2) = \mathbf{P}(3) = \frac{1}{6} + \alpha \ , \ \mathbf{P}(4) = \mathbf{P}(5) = \mathbf{P}(6) = \frac{1}{6} - \alpha,$$

donde $0 < \alpha < \frac{1}{6}$.

Gana una apuesta aquel apostador que obtiene el máximo puntaje (en caso de empate no hay ganador).

a) **(10 puntos)** Calcule la probabilidad de que el jugador A gane una apuesta y la probabilidad de que haya un empate.

b) **(8 puntos)** Suponga que se realizan 6 apuestas. Calcule la probabilidad de que haya al menos 2 empates.

c) **(12 puntos)** Suponga que la variable aleatoria X denota el número de apuestas que se realizan hasta que gana el jugador B. ¿Cuál es la distribución de X? Calcule el mínimo valor de α para el cual se cumple que el valor esperado de X es superior o igual a 3.

Problema 2 (35 puntos)

Considere una variable aleatoria X absolutamente continua con densidad de probabilidad

$$f_X(x) = \begin{cases} \frac{2a}{x^2} & \text{si } 1 \leq x \leq b \\ 0 & \text{en caso contrario,} \end{cases} \tag{1}$$

donde $a > \frac{1}{2}$ y $b > 1$.

a) **(5 puntos)** Calcule b como función de a.

b) **(10 puntos)** Calcule $\mathbf{E}(X)$, $\mathbf{Var}(X)$ y m_X, la mediana de X.

Considere ahora una muestra, $X_1, X_2, \ldots, X_{100}$, de variables aleatorias *i.i.d.* con densidad dada por (1).

c) **(8 puntos)** Construya un estimador consistente para el parámetro b. Justifique su respuesta.

d) **(12 puntos)** Suponiendo $a = 1$, estime la probabilidad de que el promedio de la muestra esté en el intervalo $[1.8 \log 2, \; 2.1 \log 2]$.

Problema 3 (35 puntos)

Los siguientes datos corresponden a los tiempos de vida (medidos en nanosegundos) de partículas radioactivas emitidas por cierto material:

20.05	1.30	2.54	1.95	9.20	4.20	1.84	1.02	5.60	1.80

Nota: En todos los test use el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a 0.10.

a) **(10 puntos)** Estudie la aleatoriedad de la muestra.

b) **(12 puntos)** Estudie si es razonable suponer que los datos corresponden a una distribución de Pareto, caracterizada por la siguiente función de distribución:

$$F(x) = \begin{cases} 1 - \frac{1}{x} & \text{si } 1 < x \\ 0 & \text{en caso contrario,} \end{cases} \tag{2}$$

c) **(13 puntos)** Se dispone ahora de una nueva muestra *i.i.d. independiente* de la anterior (no hay que verificar estos supuestos) correspondiente a los tiempos de vida de partículas radioactivas emitidas por otro material:

2.60	4.81	12.79	3.82	1.21	1.68	15.50	12.03	1.70	9.40

Implemente un test de Kolmogorov-Smirnov de comparación de muestras para concluir si es razonable o no suponer que los nuevos datos tienen la misma distribución que los anteriores.

PROBABILIDAD Y ESTADÍSTICA
EXAMEN DICIEMBRE 2004: SOLUCIÓN

Problema 1

a) Denotamos con el par (i, j) el evento "en el dado A se obtiene el resultado i y en el dado B se obtiene el resultado j". La probabilidad de que el jugador A gane una apuesta es:

$$\begin{aligned}
\mathbf{P}\{\text{Gane } A\} &= \mathbf{P}\{(2,1),(3,1),(3,2),(4,1),(4,2),(4,3),(5,1),(5,2),\\
&\quad (5,3),(5,4),(6,1),(6,2),(6,3),(6,4),(6,5)\}\\
&= 12 \times \frac{1}{6} \times \left(\frac{1}{6} + \alpha\right) + 3 \times \frac{1}{6} \times \left(\frac{1}{6} - \alpha\right)\\
&= \frac{5}{12} + \frac{3}{2}\alpha,
\end{aligned}$$

donde hemos usado que los dados son independientes.

De manera análoga:

$$\begin{aligned}
\mathbf{P}\{\text{Empate}\} &= \mathbf{P}\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}\\
&= 3 \times \frac{1}{6} \times \left(\frac{1}{6} + \alpha\right) + 3 \times \frac{1}{6} \times \left(\frac{1}{6} - \alpha\right)\\
&= \frac{1}{6}.
\end{aligned}$$

b) Se realizan 6 apuestas. Denotamos con Y el número de apuestas en las que hay empate. Está claro que Y tiene distribución *Binomial* de parámetros $n = 6$ y $p = \frac{1}{6}$; de manera que

$$\begin{aligned}
\mathbf{P}(Y \geq 2) &= 1 - \mathbf{P}(Y = 0) - \mathbf{P}(Y = 1)\\
&= 1 - \left(\frac{5}{6}\right)^6 - 6 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^5 = 0.263.
\end{aligned}$$

c) La probabilidad de que B gane una apuesta es:

$$\begin{aligned}
\mathbf{P}\{\text{Gane } B\} &= 1 - \mathbf{P}\{\text{Gane } A\} - \mathbf{P}\{\text{Empate}\}\\
&= \frac{5}{12} - \frac{3}{2}\alpha.
\end{aligned}$$

La variable X denota el número de apuestas que se realizan hasta que gana el jugador B; así que X tiene distribución *Geométrica* de parámetro $p = \frac{5}{12} - \frac{3}{2}\alpha$. Se cumple entonces

$$\mathbf{E}(X) = \frac{1}{p} = \frac{1}{\frac{5}{12} - \frac{3}{2}\alpha},$$

y de la condición $\mathbf{E}(X) \geq 3$ se obtiene $\alpha \geq \frac{1}{18}$. Por lo tanto el mínimo valor de α consistente con esta condición es $\alpha = \frac{1}{18}$.

Problema 2

a) De la condición $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (que es la condición de normalización de toda densidad de probabilidad) se obtiene:

$$\int_1^b \frac{2a}{x^2}dx = -\frac{2a}{b} + 2a = 1,$$

de donde resulta $b = \frac{2a}{2a-1}$ (la condición $a > \frac{1}{2}$ garantiza que $b > 1$).

b)

$$\mathbf{E}(X) = \int_1^b x\frac{2a}{x^2}dx = 2a\log b$$

y reemplazando por el valor de b calculado anteriormente resulta

$$\mathbf{E}(X) = 2a\log\left(\frac{2a}{2a-1}\right).$$

Por otra parte

$$\mathbf{E}(X^2) = \int_1^b x^2\frac{2a}{x^2}dx = 2a\,(b-1) = \frac{2a}{2a-1} = b,$$

de manera que

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = \frac{2a}{2a-1} - 4a^2\log^2\left(\frac{2a}{2a-1}\right).$$

Finalmente, m_X se calcula de la condición:

$$\int_1^{m_X} \frac{2a}{x^2}dx = \frac{1}{2},$$

de donde resulta $-\frac{2a}{m_X} + 2a = \frac{1}{2}$ y por lo tanto $m_X = \frac{4a}{4a-1}$.

c) Hay varias maneras de construir un estimador consistente para el parámetro b a partir de la muestra $X_1, X_2, \ldots, X_{100}$. Consideramos dos de esas maneras:

Método de los momentos: Hemos visto que $\mathbf{E}(X^2) = b$, de manera que un estimador para b es

$$\hat{b} = \frac{1}{100}\sum_{i=1}^{100} X_i^2$$

(la Ley Fuerte de los Grandes Números garantiza que $\hat{b} \to b$).

Estimador de Máxima Verosimilitud: La función de verosimilitud es

$$L(b) = \prod_{i=1}^{100} f_X(X_i) = \prod_{i=1}^{100} \frac{2a}{X_i^2}\mathbf{I}_{[1,b]}(X_i),$$

donde $\mathbf{I}_{[1,b]}(x)$ denota la función indicatriz del intervalo $[1,b]$ (esta función vale 1 si $1 \leq x \leq b$ y 0 en caso contrario). Teniendo en cuenta que $2a = \frac{b}{b-1}$ y que $\frac{b}{b-1}$ es una función decreciente de b, resulta que el valor de b que maximiza la función de verosimilitud es

$$\hat{b} = \max_{1=1,2,\ldots,100} \{X_i\},$$

que también es un estimador consistente para b.

d) Si $a = 1$ entonces $\mathbf{E}(X) = 2\log 2$ y $\mathbf{Var}(X) = 2 - 4\log^2 2 = 0.0782$. Para estimar la probabilidad usamos el Teorema Central del Límite:

$$\begin{aligned}
& \mathbf{P}\left(1.8\ \log 2 \leq \frac{1}{100}\sum_{i=1}^{100} X_i \leq 2.1\ \log 2\right) \\
= \; & \mathbf{P}\left(-\frac{\sqrt{100}}{\sqrt{0.0782}}0.2\ \log 2 \leq \frac{\sqrt{100}}{\sqrt{\mathbf{Var(X)}}}\left[\frac{1}{100}\sum_{i=1}^{100} X_i - \mathbf{E}(X)\right] \leq \frac{\sqrt{100}}{\sqrt{0.0782}}0.1\ \log 2\right) \\
\simeq \; & \mathbf{P}\left(-4.96 \leq N(0,1) \leq 2.48\right) = \Phi(2.48) + \Phi(4.96) - 1 \simeq 0.9934.
\end{aligned}$$

Problema 3

a) Para estudiar si los datos pueden suponerse aleatorios realizamos el *test de rachas de subidas y bajadas* y el *test de Spearman*.

Test de Rachas

Recordamos que se definen las variables auxiliares U_i, que toman valores 0 ó 1, de la siguiente manera: $U_i = 1$ si $X_i \leq X_{i+1}$ y $U_i = 0$ en caso contrario.

X	U
20.05	
1.30	0
2.54	1
1.95	0
9.20	1
4.20	0
1.84	0
1.02	0
5.60	1
1.80	0

Hay 7 rachas, de manera que el estadístico del test es $R = 7$. De la tabla del test se obtiene el p-valor $\alpha^* = 0.4524$. Por lo tanto la muestra pasa el test.

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	$R(X_i)$
1.02	8	1
1.30	2	2
1.80	10	3
1.84	7	4
1.95	4	5
2.54	3	6
4.20	6	7
5.60	9	8
9.20	5	9
20.05	1	10

Se obtiene $\sum_{i=1}^{10}(R(X_i)-i)^2 = 216$, y por lo tanto el estadístico del test resulta

$$r_s = 1 - 6\frac{\sum_{i=1}^{10}(R(X_i)-i)^2}{10(10^2-1)} = -0.30909.$$

La tabla del test para $n = 10$ muestra que $\alpha^* = 0.193$. Por lo tanto la muestra pasa el test.

Es razonable, entonces, suponer que la muestra es aleatoria.

b) Para estudiar si es razonable o no suponer que los datos corresponden a la distribución (2) realizamos el test de ajuste de Kolmogorov-Smirnov.

Test de Kolmogorov-Smirnov

El estadístico del test es

$$D = \sup_{t\in\mathbf{R}} |F(t) - \hat{F}(t)|,$$

donde $\hat{F}(t)$ es la función de distribución empírica de la muestra. En este caso el valor que se obtiene para el estadístico es

$$D = 0.2444.$$

La tabla del test para $n = 10$ nos da $\alpha^* > 0.20$ y por lo tanto es razonable suponer que la muestra ajusta a la distribución (2).

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

El estadístico del test es

$$D = \sup_{t\in\mathbf{R}} |\hat{F}^X(t) - \hat{F}^Y(t)|,$$

donde $\hat{F}^X(t)$ y $\hat{F}^Y(t)$ 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 tiene la misma distribución.