PROBABILIDAD Y ESTADÍSTICA
EXAMEN JULIO 2004

DATOS DEL ESTUDIANTE

Número de Parcial	Nombre	Cédula

- La duración total del examen es de **3 horas y media**.
- El puntaje mínimo para salvar el examen es de **50 puntos**.
- **Publicación de resultados:** Viernes 6 de agosto - 18:00hs.
- **Muestra de exámenes:** Lunes 9 de agosto - 18:00hs.

Problema 1 (25 puntos)

La Batalla Naval. Considere el siguiente tablero en el que hay un barco representado por los casilleros marcados con una **X**.

	1	2	3	4	5
A					
B			**X**		
C		**X**	**X**	**X**	
D					
E					

El jugador atacante realiza dos disparos (esto es, elige dos casilleros distintos de manera sucesiva) con la siguiente estrategia: Al primer disparo lo elige al azar. Si acierta (esto es, si da en un casillero marcado con una **X**) entonces al segundo disparo lo elige al azar entre los cuatro casilleros que limitan con el primero. Si no acierta el primer disparo entonces al segundo lo elige al azar entre todos los casilleros restantes.

a) **(5 puntos)** Calcule la probabilidad de que acierte el primer disparo.

b) **(7 puntos)** Calcule la probabilidad de que acierte el segundo disparo dado que acertó el primero.

c) **(5 puntos)** Calcule la probabilidad de que acierte el segundo disparo dado que no acertó el primero.

d) **(8 puntos)** Calcule la probabilidad de que haya acertado el primer disparo dado que acertó el segundo.

Problema 2 (25 puntos)

Considere un dado cargado con la siguiente distribución de probabilidad:

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

donde $0<\varepsilon<\frac{1}{6}$. Un apostador tira el dado:
Si sale 1 ó 2 pierde una cantidad $2a$.
Si sale 3 pierde a.
Si sale 4 gana a.
Si sale 5 ó 6 gana $2a$.
Aquí a representa una cantidad de dinero fija.

a) **(10 puntos)** Si la variable G denota la ganancia del jugador en una apuesta, calcule $\mathbf{E}(G)$ y $\mathbf{var}(G)$. ¿Es conveniente participar en este juego?

b) **(15 puntos)** Suponga ahora que el apostador tira el dado 50 veces de manera independiente, obteniendo las ganancias $G_1, G_2, \ldots, G_{50}$. Suponga además que $\varepsilon=\frac{1}{10}$. Estime la probabilidad de que la ganancia total, esto es $S=\sum_{i=1}^{50} G_i$, sea positiva.

Problema 3 (50 puntos)

1) La intensidad de una señal electrónica es modelada por una variable aleatoria X absolutamente continua con densidad de probabilidad

$$f_X(t)=\begin{cases}\lambda+(1-\lambda)(1+\beta)t^{\beta} & \text{si } t\in[0,1]\\ 0 & \text{en caso contrario,}\end{cases} \tag{1}$$

donde $0\leq\lambda\leq 1$ y $\beta>0$.

a) **(8 puntos)** Verificar que $f_X(t)$ es una función de densidad y calcular la función de distribución $F_X(t)$.

b) **(12 puntos)** Suponga que $X_1, X_2, \ldots, X_n$ son variables *i.i.d* con densidad (1) y $\beta=2$, y que el promedio empírico es $\overline{X_n}=0.7$. Construya un estimador consistente para el parámetro λ.

2) Considere la siguiente muestra de mediciones de señales electrónicas:

0.04	0.84	0.60	0.96	0.62	0.58	0.90	0.67	0.36	0.79

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) **(12 puntos)** ¿Es razonable suponer que la muestra es aleatoria? (realice dos test).

b) **(18 puntos)** Implemente el test de ajuste de Kolmogorov-Smirnov para concluir si es razonable o no suponer que la distribución de la muestra corresponde a la densidad dada en (1) con parámetros $\lambda=\frac{1}{3}$ y $\beta=1$.

PROBABILIDAD Y ESTADÍSTICA
EXAMEN JULIO 2004: SOLUCIÓN

Problema 1

Consideremos los sucesos

$$A_i = \{\text{el jugador acierta el disparo } i\}, \;\; i = 1, 2.$$

a) La probabilidad de que acierte el primer disparo es (casos favorables/casos posibles):

$$\mathbf{P}(A_1) = \frac{4}{25}.$$

b) La probabilidad de que acierte el segundo disparo dado que acertó el primero es:

$$\mathbf{P}(A_2|A_1) = \frac{\mathbf{P}(A_2 \cap A_1)}{\mathbf{P}(A_1)}.$$

Ahora, analizando por separado los casilleros $C2$, $C3$, $C4$ y $B3$ (que corresponden al suceso A_1) tenemos, en un abuso de notación:

$$\begin{aligned}
\mathbf{P}(A_2 \cap A_1) &= \mathbf{P}(A_2 \cap C2) + \mathbf{P}(A_2 \cap C3) + \mathbf{P}(A_2 \cap C4) + \mathbf{P}(A_2 \cap B3) \\
&= \mathbf{P}(A_2|C2)\mathbf{P}(C2) + \mathbf{P}(A_2|C3)\mathbf{P}(C3) + \mathbf{P}(A_2|C4)\mathbf{P}(C4) + \mathbf{P}(A_2|B3)\mathbf{P}(B3) \\
&= \frac{1}{4} \times \frac{1}{25} + \frac{3}{4} \times \frac{1}{25} + \frac{1}{4} \times \frac{1}{25} + \frac{1}{4} \times \frac{1}{25} = \frac{6}{100}.
\end{aligned}$$

Con esto, y el resultado calculado en la parte anterior, se obtiene: $\mathbf{P}(A_2|A_1) = \frac{3}{8}$.

c) Por la estrategia con que se elige el segundo disparo queda claro que (de nuevo, casos favorables/ casos posibles):

$$\mathbf{P}(A_2|A_1^c) = \frac{4}{24}.$$

d) Usamos la fórmula de Bayes y las probabilidades calculadas arriba:

$$\mathbf{P}(A_1|A_2) = \frac{\mathbf{P}(A_2|A_1)\mathbf{P}(A_1)}{\mathbf{P}(A_2|A_1)\mathbf{P}(A_1) + \mathbf{P}(A_2|A_1^c)\mathbf{P}(A_1^c)} = \frac{3}{10}.$$

Problema 2

a) El valor esperado de la variable G es:

$$\mathbf{E}(G) = -2a2\left(\frac{1}{6}+\varepsilon\right) - a\left(\frac{1}{6}+\varepsilon\right) + a\left(\frac{1}{6}-\varepsilon\right) + 2a2\left(\frac{1}{6}-\varepsilon\right) = -10a\varepsilon.$$

Este resultado nos dice que no es conveniente participar en el juego ya que el valor esperado de la ganancia es negativo.

Calculamos:

$$\mathbf{E}(G^2) = 4a^2 2\left(\frac{1}{6}+\varepsilon\right) + a^2\left(\frac{1}{6}+\varepsilon\right) + a^2\left(\frac{1}{6}-\varepsilon\right) + 4a^2 2\left(\frac{1}{6}-\varepsilon\right) = 3a^2.$$

De manera que: $\mathbf{Var}(G) = \mathbf{E}(G^2) - \mathbf{E}^2(G) = a^2\left(3-100\varepsilon^2\right)$.

b) Para estimar $\mathbf{P}\left(\sum_{i=1}^{50} G_i \geq 0\right)$ usamos la aproximación que nos proporciona el Teorema Central del Límite:

$$\begin{aligned}
\mathbf{P}\left(\sum_{i=1}^{50} G_i \geq 0\right) &= \mathbf{P}\left(\sqrt{50}\frac{\overline{G_{50}} - \mathbf{E}(G)}{\sqrt{\mathbf{Var}(G)}} \geq \sqrt{50}\frac{10a\varepsilon}{a\sqrt{3-100\varepsilon^2}}\right) \\
&\simeq \mathbf{P}\left(N(0,1) \geq \sqrt{50}\frac{10\varepsilon}{\sqrt{3-100\varepsilon^2}}\right).
\end{aligned}$$

Finalmente, usando el valor $\varepsilon = \frac{1}{10}$, resulta

$$\mathbf{P}\left(\sum_{i=1}^{50} G_i \geq 0\right) \simeq \mathbf{P}\left(N(0,1) \geq 5\right) \simeq 0.$$

Problema 3

1) a) Para verificar que la función $f_X(t)$ (que claramente es no negativa) es una función de densidad, calculamos

$$\int_{\mathbf{R}} f_X(t)dt = \int_0^1 [\lambda + (1-\lambda)(1+\beta)t^\beta]dt = \lambda + (1-\lambda)(1+\beta)\frac{1}{(1+\beta)} = 1.$$

La función de distribución es:

$$F_X(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \lambda t + (1-\lambda)t^{1+\beta} & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } 1 \leq t. \end{cases}$$

b) Para construir un estimador consistente para el parámetro λ usamos el método de los momentos: el valor esperado de una variable con densidad (1) es

$$\mathbf{E}(X) = \int_0^1 t[\lambda + (1-\lambda)(1+\beta)t^\beta]dt = \frac{\lambda}{2} + (1-\lambda)\frac{(1+\beta)}{(2+\beta)};$$

de modo que para $\beta = 2$ queda:

$$\mathbf{E}(X) = \frac{\lambda}{2} + (1-\lambda)\frac{3}{4} = \frac{3-\lambda}{4}.$$

El estimador para λ se obtiene, entonces, de resolver la ecuación

$$\overline{X_n} = \frac{3-\lambda}{4} \quad \Rightarrow \quad \hat{\lambda} = 3 - 4\overline{X_n} = 0.2$$

2) 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
0.04	
0.84	1
0.60	0
0.96	1
0.62	0
0.58	0
0.90	1
0.67	0
0.36	0
0.79	1

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	$R(X_i)$
0.04	1	1
0.36	9	2
0.58	6	3
0.60	3	4
0.62	5	5
0.67	8	6
0.79	10	7
0.84	2	8
0.90	7	9
0.96	4	10

Se obtiene $\sum_{i=1}^{10}(R(X_i) - i)^2 = 148$, y por lo tanto el estadístico del test resulta

$$r_s = 1 - 6\frac{\sum_{i=1}^{10}(R(X_i) - i)^2}{10(10^2 - 1)} = 0.103.$$

La tabla del test para $n = 10$ muestra que $\alpha^* = 0.393$. Por lo tanto la muestra pasa el test.

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

b) **Test de Kolmogorov-Smirnov**

Cuando $\lambda = \frac{1}{3}$ y $\beta = 1$, la función de distribución correspondiente a la densidad (1) es

$$F_0\,(t) = \begin{cases} 0 & \text{si } t \leq 0 \\ \frac{1}{3}t + \frac{2}{3}t^2 & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } 1 \leq t. \end{cases}$$

El estadístico del test de Kolmogorov-Smirnov es

$$D = \sup_{t \in \mathbf{R}} |F_0(t) - F_n(t)|,$$

donde $F_n(t)$ es la función de distribución empírica de la muestra. El valor que se obtiene para el estadístico es

$$D = 0.2176.$$

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 F_0.