PROBABILIDAD Y ESTADÍSTICA
EXAMEN, 17 de julio de 2010

DATOS DEL ESTUDIANTE

No. de Examen	Nombre y Apellido	Cédula

- **La duración del examen es de 4 horas.**
- **Publicación de resultados: Miércoles 28 de Julio, 20 hs.**
- **Muestra de exámenes: Jueves 29 de Julio, 18 hs.**

	1	2	3	Totales
Pr.1				
Pr.2				
Pr.3				
Total				

Problema 1 (26 puntos)

Un cierto curso de facultad consta de 6 temas diferentes. Al final del curso se realiza un examen que consta de 10 preguntas. Cada pregunta consiste de uno y sólo uno de los 6 temas, elegido al azar, y el tema de cada pregunta es independiente de los temas de las preguntas restantes. Cada pregunta vale 1 punto si está bien contestada y 0 si no. Un estudiante estudió en detalle 4 de los 6 temas y los otros 2 temas en mucho menor profundidad. En cada pregunta el estudiante tiene 80% de chance de responder correctamente si le toca uno de los temas que estudió en detalle, mientras que si le toca alguno de los otros sólo cuenta con un 20% de chance.

1. ¿Cuál es el puntaje esperado del estudiante en la prueba?
2. ¿Cuál es la probabilidad de aprobar el examen si el mínimo de aprobación es de 60%?
3. Si se sabe que el estudiante salvó, ¿cuál es la probabilidad de que todas las preguntas hayan sido de los temas que no estudió en profundidad?

Problema 2 (40 puntos)

Sea X una variable aleatoria con distribución exponencial de parámetro λ. Dado $c > 0$, se define una nueva variable aleatoria $Y = ce^X$. Se dice que Y tiene distribución de Pareto de parámetros λ y c.

1. Calcular E(Y) y Var(Y).

2. (a) Indicar el recorrido de la variable Y.
   (b) Hallar la función de distribución y la densidad de Y.

3. Dada una muestra iid $Y_1, Y_2, \dots, Y_n$ con distribución de Pareto de parámetros λ y c:

   (a) Hallar los estimadores de máxima verosimilitud de λ y c. Sugerencia: observar que la función de verosimilitud es creciente como función de c.
   (b) Hallar un estimador consistente de λ, distinto del anterior, en función de c.
   (c) Hallar un intervalo de confianza aproximado al nivel α para λ, en función de c.

Problema 3 (34 puntos)

Los siguientes datos corresponden a la distancia promedio recorrida por partido en el mundial para cada una de las selecciones indicadas en la tabla.

	Equipo	Distancia
1	Uruguay	109,57
2	Alemania	108,48
3	Argentina	98,79
4	Brasil	100,01
5	Chile	101,78
6	España	105,17
7	Ghana	115,4
8	Países Bajos	102,94
9	Paraguay	110,17
10	Portugal	110,15

1. Estudie la aleatoriedad de los datos. (Basta realizar una prueba de hipótesis.)

2. ¿Puede suponerse que la distribución de los datos es gaussiana? (Realice una prueba de hipótesis de Lilliefors.) En caso afirmativo determine intervalos de confianza al nivel $0,9$ para los parámetros.

3. Se tienen también los datos de promedio de goles a favor por partido para las mismas selecciones. Asumiendo que esta nueva muestra es iid, ¿puede suponerse que la distancia recorrida promedio y el promedio de goles a favor son independientes? (Realice una prueba de hipótesis.)

	Equipo	Goles a favor
1	Uruguay	1,5
2	Alemania	2,17
3	Argentina	2
4	Brasil	1,8
5	Chile	0,75
6	España	1,17
7	Ghana	1
8	Países Bajos	2
9	Paraguay	0,6
10	Portugal	1,75

Datos obtenidos en http://es.fifa.com/ el 09/07/2010.

SOLUCIÓN EXAMEN, 17 de julio de 2010

Problema 1 (26 puntos)

1. Sea X la variable que representa el puntaje obtenido en la prueba (o la cantidad de preguntas bien respondidas). Por la independencia entre preguntas $X \sim \mathrm{Bin}(10; \mathrm{p})$, donde p es la probabilidad de contestar correctamente una pregunta. Para calcular p se consideran:

   A : la pregunta es de uno de los temas que el estudiante estudió en profundidad

   B : el estudiante responde bien la pregunta

   Entonces
   $$p = P(B|A)P(A) + P(B|A^c)P(A^c) = 0,8 \cdot \frac{2}{3} + 0,2\frac{1}{3} = 0,6.$$
   El puntaje esperado es $E(X) = 10p = 6$.

2. La probabilidad de salvar si se aprueba con el 60% es $P(X \geq 6) = 1 - P(X \leq 5)$ y de la tabla de la distribución binomial $P(X \geq 6) = 1 - 0,3669 = 0,6331$.

3. Se considera los sucesos

   C : todas las preguntas son de los temas que no estudió en profundidad ($P(C) = \left(\frac{1}{3}\right)^{10}$),

   D : el estudiante salvó ($D = \{X \geq 6\}$ y de la parte anterior $P(D) = 0,6331$).

   Luego
   $$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{P(D|C)P(C)}{P(D)}.$$
   Condicionado a C (si todas las preguntas son de los temas que el estudiante no estudió en profundidad) el puntaje obtenido es una variable $Z \sim \mathrm{Bin}(10; 0,2)$ y $P(D|C) = P(Z \geq 6) = 1 - P(Z \leq 5) = 1 - 0,9936 = 0,0064$ (de la tabla). Entonces
   $$P(C|D) = \frac{0,0064}{0,6331}\left(\frac{1}{3}\right)^{10} \approx 1,71 \cdot 10^{-7}.$$

Problema 2 (40 puntos)

1. $\mathrm{E}(Y)$ existe si $\lambda > 1$ y en ese caso $\mathrm{E}(Y) = \frac{\lambda c}{\lambda - 1}$; $\mathrm{Var}(Y)$ existe si $\lambda > 2$ y en ese caso $\mathrm{Var}(Y) = \frac{\lambda c^2}{(\lambda-1)^2(\lambda-2)}$.

2. (a) $R_Y = [c, +\infty]$.

   (b) $F_Y(y) = \begin{cases} 0 & \text{si } y < c \\ 1 - \left(\frac{c}{y}\right)^{\lambda} & \text{si } y > c \end{cases}$, $f_Y(y) = \begin{cases} 0 & \text{si } y < c \\ \frac{\lambda c^{\lambda}}{y^{\lambda+1}} & \text{si } y > c \end{cases}$

3. (a) La función de verosimiliud está dada por $L(\lambda, c) = \prod_{i=1}^{n} \frac{\lambda c^{\lambda}}{Y_i^{\lambda+1}}$ si $Y_i \geq c \ \forall \ i = 1, \ldots, n$. Observando además que es creciente en c se tiene que $\widehat{c} = \min_{1 \leq i \leq n} Y_i$ es el estimador de máxima verosimilitud de c. Tomando logaritmo, derivando respecto de λ y estudiando el signo de la derivada, se obtiene que $\widehat{\lambda} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{Y_i}{\widehat{c}}\right)}$ es el estimador de máxima verosimilitud de λ.

(b) Por la LFGN $\overline{Y}_n$ es un estimador consistente de $\mathrm{E}(Y) = \frac{\lambda c}{\lambda - 1}$. Al despejar λ en función de $\mathrm{E}(Y)$ se tiene $\lambda = \frac{\mathrm{E}(Y)}{\mathrm{E}(Y)-c}$ y como $g(x) = \frac{x}{x-c}$ es una función continua (para $x > c$), entonces $\widetilde{\lambda} = \frac{\overline{Y}_n}{\overline{Y}_n - c}$ es también un estimador consistente de λ

(c) Para construir un intervalo de confianza para λ en función de c se considera un intervalo de confianza aproximado con media y varianza ligadas para el valor esperado de Y

$$[a,b] = \left[\overline{Y}_n - \frac{\sigma(\widetilde{\lambda}) z_{\frac{\alpha}{2}}}{\sqrt{n}}, \overline{Y}_n + \frac{\sigma(\widetilde{\lambda}) z_{\frac{\alpha}{2}}}{\sqrt{n}}\right],$$

donde $\sigma(\widetilde{\lambda})$ es el estimador de la desviación estándar que se obtiene sustituyendo λ en $\sigma(\lambda) = \frac{c}{\lambda-1}\sqrt{\frac{\lambda}{\lambda-2}}$ por el estimador $\widetilde{\lambda} = \frac{\overline{Y}_n}{\overline{Y}_n - c}$. Luego $[a,b]$ es un intervalo de confianza al nivel α para $\mathrm{E}(Y) = \frac{\lambda c}{\lambda-1}$ y $\left[\frac{b}{b-c}, \frac{a}{a-c}\right]$ es un intervalo de confianza para λ.

Problema 3 (34 puntos)

1. Test de rachas: $\begin{cases} H_0 : \text{la muestra es iid} \\ H_1 : \text{hay pocas rachas} \end{cases}$. El estadístico es $R = 5$ y el p-valor $\alpha^* = 0,2427$, entonces se acepta que los datos son iid.

   Test de correlación de rangos de Spearman: $\begin{cases} H_0 : \text{la muestra es iid} \\ H_1 : \text{hay tendencia creciente} \end{cases}$. El estadístico es $R_S = 0,4303$ y el p-valor $\alpha^* \approx 0,109$, entonces se acepta que los datos son iid.

2. Test de Lilliefors para normales:

$$\begin{cases} H_0 : \text{la muestra es gaussiana} \\ H_1 : \text{no } H_0 \end{cases}$$

   El estadístico $D_n = 0,1621$ y el p-valor $\alpha^* > 0,2$, por lo tanto puede asumirse que los datos tienen distribución normal. Los parámetros se estiman mediante intervalos de confianza para media y varianza de datos gaussianos, es decir asumiendo una distribución $N(\mu, \sigma^2)$.

   Intervalo de confianza para μ: $\overline{X}_n = 106,25$, $s_n = 5,34$, $t_{\frac{\alpha}{2}}(n-1) = 1,833$,

$$\left[\overline{X}_n - \frac{s_n t_{\frac{\alpha}{2}}(n-1)}{\sqrt{n}}, \overline{X}_n + \frac{s_n t_{\frac{\alpha}{2}}(n-1)}{\sqrt{n}},\right] = [103,5; 109,34].$$

   Intervalo de confianza para σ^2: $\sigma_n^2 = 28,55$, $\chi^2_{1-\frac{\alpha}{2}}(n-1) = 3,325$, $\chi^2_{\frac{\alpha}{2}}(n-1) = 16,919$,

$$\left[\frac{(n-1)s_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right] = [15,19; 77,29].$$

3. Se realiza un test de correlación de rangos de Spearman para decidir sobre la independencia de ambas muestras:

$$\begin{cases} H_0 : \text{ambas muestras son independientes entre sí} \\ H_1 : \text{no } H_0 \end{cases}$$

   El estadístico es $R_S = -0,4576$ y el p-valor es $\alpha^* \approx 0,096$. Entonces se rechaza la hipótesis nula de que ambas muestras son independientes.