BIOESTADÍSTICA
EXAMEN 11 DE MARZO DE 2006

DATOS DEL ESTUDIANTE

Nombre	Cédula

- **La duración del examen es 3 horas.**
- **El puntaje mínimo para aprobar es 50 puntos.**

Problema 1 (40 + 5 puntos)

Se considera un sitio en la cadena de ADN de un virus y se estudia su evolución en el tiempo. Se supone que en cada ronda de replicación, la probabilidad de que la base que ocupa el sitio en cuestión cambie a otra base determinada es 1/6, para cualquiera de la otras tres bases. Esto implica que la probabilidad de que la base no cambie en una ronda de replicación cualquiera es 1/2. Se considera el proceso aleatorio $X_0, X_1, X_2, \ldots, X_n, \ldots$, en donde X_i indica la base presente en el sitio en el instante i.

a) **(10 puntos)** Calcule $p_{AA}^{(2)} = \mathbf{P}(X_2 = A | X_0 = A)$.

b) **(10 puntos)** Calcule $p_{GA}^{(2)} = \mathbf{P}(X_2 = A | X_0 = G)$ y muestre que se cumple

$$p_{GA}^{(2)} = p_{CA}^{(2)} = p_{TA}^{(2)}.$$

Suponga ahora que el sitio se elige de manera aleatoria, de modo que

$$\mathbf{P}(X_0 = A) = \mathbf{P}(X_0 = T) = \mathbf{P}(X_0 = G) = \mathbf{P}(X_0 = C) = \frac{1}{4}.$$

c) **(10 puntos)** Calcule $\mathbf{P}(X_2 = A)$.

d) **(10 puntos)** Calcule $\mathbf{P}(X_0 = G | X_2 = A)$.

e) **Parte extra: (5 puntos)** Calcule $p_{AA}^{(3)} = \mathbf{P}(X_3 = A | X_0 = A)$.

Problema 2 (20 puntos)

Cada luciérnaga tiene un modo peculiar de centellear; un destello corto de luz es seguido por un período de reposo. Los siguientes datos corresponden a los períodos de reposo entre centelleos (medidos en segundos) para una muestra de 11 luciérnagas:

4.05	3.95	3.74	3.33	3.94	4.04	3.73	3.75	3.88	3.50	3.59

Se asume que los datos corresponden a una **distribución normal** de valor esperado μ. Realice una prueba de hipótesis al nivel 90% para decidir entre las siguientes hipótesis:

$$\begin{cases} H_0: & \mu = 4 \\ H_1: & \mu \neq 4 \end{cases}$$

Problema 3 (40 puntos)

Se realiza un estudio sobre los tiempo de reacción (en milisegundos) de un grupo de boxeadores del Palermo Boxing Club. Para cada boxeador se registraron dos tiempos: el tiempo de **reacción visual** (muestra A) es el tiempo de respuesta a una señal luminosa; el tiempo de **reacción auditiva** (muestra B) es el tiempo de respuesta al chasquido de un interruptor eléctrico.

Boxeador	Reacción visual (A)	Reacción auditiva (B)
1	159	201
2	183	218
3	190	185
4	186	169
5	177	191
6	176	165
7	197	171
8	157	193
9	181	211
10	191	187

Nota: En las pruebas de hipótesis utilice el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a 0.10

a) **(10 puntos)** Para cada muestra realice una prueba de hipótesis para decidir si es razonable suponer que los datos son independientes e idénticamente distribuidos.

b) **(18 puntos)** Implemente la prueba de Lilliefors en la Muestra A para decidir si es razonable suponer que esa muestra ajusta a una distribución exponencial.

c) **(12 puntos)** Realice una prueba de hipótesis para decidir si es razonable suponer que la Muestra A y la Muestra B son independientes entre sí.

SOLUCIÓN

Problema 1

a)

$$p^{(2)}_{AA} = \frac{1}{2} \times \frac{1}{2} \; + \; 3 \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{3}.$$

b)

$$p^{(2)}_{GA} = 2 \times \frac{1}{2} \times \frac{1}{6} \; + \; 2 \times \frac{1}{6} \times \frac{1}{6} = \frac{2}{9}.$$

c) Condicionando sobre X_0 se obtiene:

$$\begin{aligned}
\mathbf{P}(X_2 = A) &= \mathbf{P}(X_2 = A|X_0 = A)\mathbf{P}(X_0 = A) + \mathbf{P}(X_2 = A|X_0 = G)\mathbf{P}(X_0 = G) \\
&+ \mathbf{P}(X_2 = A|X_0 = C)\mathbf{P}(X_0 = C) + \mathbf{P}(X_2 = A|X_0 = T)\mathbf{P}(X_0 = T) \\
&= \frac{1}{4}\left(p^{(2)}_{AA} + 3\, p^{(2)}_{GA}\right) = \frac{1}{4}.
\end{aligned}$$

d)

$$\mathbf{P}(X_0 = G|X_2 = A) = \frac{\mathbf{P}(X_2 = A|X_0 = G)\; \mathbf{P}(X_0 = G)}{\mathbf{P}(X_2 = A)} = \frac{2}{9}$$

e) Condicionando sobre X_1 y usando los resultados anteriores se obtiene:

$$p^{(3)}_{AA} = \frac{1}{2}\, p^{(2)}_{AA} + 3\, \frac{1}{6}\, p^{(2)}_{GA} = \frac{5}{18}.$$

Problema 2

De la muestra se obtiene $\overline{X_n} = 3.773$ y $s_n = 0.223$, de manera que

$$\sqrt{n}\frac{(\overline{X_n} - 4)}{s_n} = -3.376$$

Por otra parte, de la tabla de Student con 10 grados de libertad:

$$t_{0.05}(10) = 1.812$$

Con esto,

$$\overline{X_n} \leq 4 - \frac{s_n}{\sqrt{n}}\, t_{0.05}(10)$$

y entonces el estadístico cae dentro de la región crítica. Por lo tanto decidimos H_1 al nivel 90%.

Problema 3

a) – **Muestra A:**
Test de Rachas: El número de rachas es $R = 5$ y el p-valor: $\alpha^* = 0.2427$.
Test de correlación de rangos de Spearman: $\sum_{i=1}^{10} (R(X_i) - i)^2 = 136$, de manera que el estadístico es $r_s = 0.175$ y $\alpha^* = 0.316$.
En ambos casos es razonable suponer que la Muestra A es aleatoria.

– **Muestra B:**
Test de Rachas: El número de rachas es $R = 6$ y el p-valor: $\alpha^* = 0.7573$.
Test de correlación de rangos de Spearman: $\sum_{i=1}^{10} (R(Y_i) - i)^2 = 186$, de manera que el estadístico es $r_s = -0.127$ y $\alpha^* = 0.367$.
En ambos casos es razonable suponer que la Muestra B es aleatoria.

b) Para la Muestra A se obtiene $\overline{X_n} = 179.7$ de manera que $\hat{\lambda} = \frac{1}{\overline{X_n}} = 0.0056$.

Para implementar el test de Lilliefors construimos la siguiente tabla, donde (notemos que todos los datos son positivos)

$$F_0(x) = 1 - e^{-\hat{\lambda}x}.$$

X_i	X_i^*	$F_o(X_i^*)$	$\frac{i}{n}$	$\lvert F_o(X_{i+1}^*) - \frac{i}{n}\rvert$	$\lvert F_o(X_i^*) - \frac{i}{n}\rvert$
159	157	0.585	0	**0.585**	0.485
183	159	0.589	0.1	0.489	0.389
190	176	0.627	0.2	0.427	0.327
186	177	0.629	0.3	0.329	0.229
177	181	0.637	0.4	0.237	0.137
176	183	0.641	0.5	0.141	0.041
197	186	0.647	0.6	0.047	0.053
157	190	0.655	0.7	0.0045	0.145
181	191	0.657	0.8	0.143	0.243
191	197	0.668	0.9	0.232	0.332
			1		

De manera que el estadístico es $D = 0.585$ y entonces $\alpha^* < 0.01$. Por lo tanto rechazamos H_0.

c) Implementamos el test de Spearman para independencia de muestras. El estadístico del test es

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

En este caso: $\sum_{i=1}^{10} (R(X_i) - R(Y_i))^2 = 218$ y entonces $r_S = -0.321$ y $\alpha^* = 0.164$. Por lo tanto es razonable suponer que las muestras son independientes entre sí.