Probabilidad y Estadística
Examen, 1 de febrero de 2012

Datos del estudiante

Nro. de examen	Nombre y apellido	Cédula

- **La duración del examen es de 4 horas.**
- **Publicación de resultados: lunes 13 de febrero, 20:00 hs.**
- **Muestra de exámenes: martes 14 de febrero, 17:00 hs.**

Ejercicio 1
Una persona que sale de vacaciones llega tarde a la terminal y corre para tomar un ómnibus, por lo que elige el ómnibus para el destino que quiere al azar. Para el mismo destino salen 8 ómnibus, 5 son directos y 3 son semidirectos. El tiempo (en horas) que demoran en llegar a destino se modela mediante una variable exponencial de parámetro 1/3 para los directos y de parámetro 1/5 para los semidirectos.

1. Calcular la probabilidad de demorar más de 3 horas y media (3,5 horas) si toma un directo y si toma un semidirecto.
2. Calcular la probabilidad de demorar más de 3 horas y media (3,5 horas) eligiendo el ómnibus al azar como se dijo antes.
3. En el destino otra persona espera al pasajero, y observa que demora más de 3 horas y media, pero no sabe qué ómnibus tomó. Calcular la probabilidad de que haya tomado un ómnibus semidirecto.
4. Si elige el ómnibus al azar, hallar la densidad del tiempo de duración del viaje y su valor esperado.

Ejercicio 2
La entropía de una variable aleatoria discreta X con probabilidad puntual $p(x)$ se mide en bits, y se define de la siguiente manera:

$$H(X) = \sum_{x \in R_X} -p(x) \log_2 p(x).$$

1. Sea $X \sim Ber(p)$, es decir que la probabilidad puntual es $p(1) = p$ y $p(0) = 1-p$, $p \in (0, 1)$. Calcular la entropía $H(p)$ y mostrar que es máxima, y vale un bit, cuando $p = \frac{1}{2}$.
2. Calcular la entropía $H(n)$ de una variable aleatoria uniforme en $\{1, 2, \ldots, n\}$, donde la probabilidad puntual es $p(i) = 1/n$ para $i = 1, \ldots, n$.

3. Para interpretar el resultado anterior se propone el siguiente juego: dos personas juegan a un juego en el que una elige un número al azar de manera uniforme en el conjunto $\{1, 2, ..., n\}$ donde $n = 2^m$. La otra debe deducir, luego de una serie de preguntas el número elegido. Supongamos que la persona que debe deducir el número usa la siguiente estrategia: divide el conjunto posible en dos (digamos mitad superior y mitad inferior) y le pregunta a la otra persona en qué mitad se encuentra el número elegido. En la segunda instancia divide la mitad en la que se encuentra el número en dos mitades y pregunta a la otra persona en qué mitad se encuentra el número y así sucesivamente. Siguiendo las respuestas de la primera persona, va descartando en cada instancia conjuntos (cada vez más pequeños) del conjunto inicial hasta deducir el número. Sea r la cantidad necesaria de preguntas hasta llegar al número. Escribir r en función de n y relacionarlo con $H(n)$.

4. Dos jugadores juegan N veces. El primer jugador elige cada vez, aleatoriamente, un número X en el conjunto $\{1, 2, 3, 4\}$ con probabilidades respectivas p_1, p_2, p_3, p_4 (desconocidas). Conociendo los números elegidos por el primer jugador $X_1, \ldots, X_N$, construya un estimador de $H(X)$, que sea consistente.

Ejercicio 3

Sea la muestra correspondiente a los tiempos de vida de diez componentes eléctricos de cierto tipo y marca:

0.22	0.26	0.12	0.18	0.23	0.87	0.55	0.32	0.43	0.09

1. Realizar dos pruebas de hipótesis para estudiar la aleatoriedad de la muestra.

2. Realizar una prueba de Lilliefors para ver si es razonable afirmar que los datos son exponenciales.

3. Se dispone de una segunda muestra independiente de la anterior:

1.74	6.20	0.24	0.67	1.55	2.27	1.85	0.90	0.89	2.94

   Realice una prueba (Kolmogorov-Smirnov para dos muestras) para decidir si las dos muestras tienen la misma distribución.

Nota: Para todas las pruebas de hipótesis considerar $\alpha = 0,10$. Escriba claramente, en cada caso, el resultado de la prueba efectuada, indicando el valor de los estadísticos y los respectivos p-valores.

Resumen de la solución del examen, 1 de febrero de 2012

Ejercicio 1

Sea X la duración del viaje y D el suceso "el ómnibus es directo".

1. $P(X > 3,5|D) = e^{-\frac{3,5}{3}} = 0,311$, $P(X > 3,5|D^c) = e^{-\frac{3,5}{5}} = 0,497$.

2. $P(X > 3,5) = P(X > 3,5|D)P(D) + P(X > 3,5|D^c)P(D^c) = e^{-\frac{3,5}{3}}\frac{5}{8} + e^{-\frac{3,5}{5}}\frac{3}{8} = 0,381$.

3. $P(D^c|X > 3,5) = \dfrac{P(X > 3,5|D^c)P(D^c)}{P(X > 3,5)} = 0,4889$.

4. La función de distribución de X es $F_X(t) = P(X \le t) = P(X \le t|D)P(D) + P(X \le t|D^c)P(D^c)$ de donde $F_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ (1-e^{-t/3})\dfrac{5}{8} + (1-e^{-t/5})\dfrac{3}{8} & \text{si } t > 0 \end{cases}$, y la densidad es $f_X(t) = \begin{cases} 0 & \text{si } t < 0 \\ \dfrac{5}{24}e^{-t/3} + \dfrac{3}{24}e^{-t/5} & \text{si } t > 0 \end{cases}$. $E(X) = \int_0^{+\infty} t\left(\dfrac{5}{24}e^{-t/3} + \dfrac{3}{24}e^{-t/5}\right) dt = \dfrac{15}{4}$.

Ejercicio 2

1. La entropía de una variable aleatoria $X \sim Ber(p)$ es $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$. Derivando tenemos que:
$$H'(p) = -\log_2 p - \log_2 e + \log_2(1-p) + \log_2 e = \log_2\left(\frac{1-p}{p}\right).$$
Luego la entropía es máxima cuando $H'(p) = 0$, es decir con $p = \frac{1}{2}$. Sustituyendo: $H(\frac{1}{2}) = 1$ bit.

2. Sustituyendo en la definición con $p_i = \frac{1}{n}$ tenemos que $H(n) = \log_2 n$.

3. Consideremos primeramente que $n = 2^m$. Preguntemos si $x > 2^{m-1}$. Sea la respuesta afirmativa o negativa, en el segundo intento tenemos ahora la mitad de opciones. Reiteramos el proceso de preguntar en la mitad del intervalo incógnita. Luego de la i-ésima pregunta tendremos 2^{m-i} números posibles para x. Luego, con $i = m$ preguntas hemos logrado deducir el número x. Puesto que $n = 2^m$, descubrimos x en un total de $m = H(n)$ preguntas, o menos.

4. Las frecuencias
$$p_{iN} = \frac{\#\{1 \le j \le N : X_j = i\}}{N}$$
para $i = 1, \ldots, 4$ son estimadores consistentes casi seguramente (convergen casi seguramente, por la ley fuerte de los grandes números) a las probabilidades p_i. Al ser la función
$$g(x_1, x_2, x_3, x_4) = -x_1\log_2 x_1 - x_2\log_2 x_2 - x_3\log_2 x_3 - x_4\log_2 x_4$$
una función continua, tenemos que
$$H_N(X) = -p_{1N}\log_2 p_{1N} - p_{2N}\log_2 p_{2N} - p_{3N}\log_2 p_{3N} - p_{4N}\log_2 p_{4N}$$
es un estimador consistente de $H(X)$.

Ejercicio 3

1. Realizamos los tests de rachas y de correlación de rangos de Spearman. El número de rachas es 6 y el p-valor es 0,7573, $R_S = 0,212$y el p-valor es 0,28, por lo que ambos tests aceptan la hipótesis de aleatoriedad.

2. El estadístico del test de Lilliefors es 0,241 y el p-valor es mayor que 0,2, por lo que aceptamos la hipótesis de que los datos tengan distribución exponencial.

3. El estadístico del test de Kolmogorov-Smirnov para comparación de dos muestras es 0,8 y el p-valor es menor que 0,01, por lo que rechazamos que ambas muestras tengan la misma distribución.