Probabilidad y Estadística
Examen, 14 de diciembre de 2011

Datos del estudiante

Nro. de examen	Nombre y apellido	Cédula

- **La duración del examen es de 4 horas.**
- **Publicación de resultados: miércoles 21 de diciembre, 20:00 hs.**
- **Muestra de exámenes: jueves 22 de diciembre, 12:30 hs.**

Ejercicio 1
Un supermercado sortea 10 autos, uno por día durante 10 días. Los clientes obtienen determinada cantidad de cupones de acuerdo al monto de su compra, cada vez que compran.

1. Suponemos que cada día participan del sorteo 10000 cupones. Un cliente obtiene 10 cupones el primer día. El cliente puede seguir dos estrategias: depositar los 10 cupones todos juntos o guardarlos y depositar un cupón cada día. Indique cuál estrategia es la mejor para tener mayor probabilidad de ganar.

2. Ahora la cantidad de cupones por cliente se modela con una variable aleatoria con distribución de Poisson de parámetro 2. Consideramos un día donde concurrieron 4000 clientes. Calcule aproximadamente la probabilidad de que la cantidad de cupones sea superior a 8250.

3. Si la cantidad de clientes por día sigue una distribución de Poisson de parámetro 4000, calcule aproximadamente la probabilidad de que el mismo día concurran 4000 clientes y se sorteen menos de 8000 cupones. (Se sigue asumiendo que la cantidad de cupones por persona tiene distribución de Poisson de parámetro 2.)

4. Se desea estimar λ, la cantidad media de cupones por cliente. Asumiendo que la cantidad de cupones por cliente tiene distribución de Poisson de parámetro λ, construya un intervalo de confianza aproximado al 90 %, si en un día hay 4500 clientes y 9000 cupones.

Ejercicio 2
En una parada de ómnibus un pasajero llega entre las 0:00 a.m y la 1:00 a.m. en un tiempo T fijo (no aleatorio, $0 \leq T \leq a$ y $0 < a < 1$). Hay dos ómnibus que le sirven, uno pasa a las 0:00 y el otro a la 1:00. Ambos tienen un retraso aleatorio, representado por variables X e Y, independientes entre si. X e Y tienen distribución uniforme en el intervalo $[0, a]$. Sea Z la variable que indica el tiempo que el pasajero espera en la parada.

1. Calcule la función de distribución y pruebe que la densidad de Z es:

$$f_Z(z) = \begin{cases} \frac{1}{a} & \text{si } 0 < z < a - T \\ \\ \frac{T}{a^2} & \text{si } 1 - T < z < a + 1 - T \\ \\ 0 & \text{en otro caso} \end{cases}.$$

   Grafique.

2. Calcule el valor esperado de Z. Halle T tal que minimice la esperanza del tiempo de espera del ómnibus.

3. Se sabe ahora que se llega a la parada siempre a las 0:12 ($T = 1/5$). Estime a sabiendo que el promedio de n observaciones $Z_1, Z_2, \ldots, Z_n$ independientes de los tiempos de espera en la parada, para n grande, es media hora.

Ejercicio 3
Sean X, Y variables aleatorias independientes con distribución $N(0,1)$. Sea la variable aleatoria $R = \sqrt{X^2 + Y^2}$.

1. Probar que la función de distribución de R es $F_R(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - e^{-t^2/2} & \text{si } t \geq 0 \end{cases}$.

2. Se tiene la siguiente muestra correspondiente a las distancias al origen de 10 puntos elegidos al azar en el plano.

0,9	1,85	0,3	2,50	1,92	1,95	1,76	0,92	0,68	0,31

   a) Indique si se puede suponer que la muestra es aleatoria (realice 2 pruebas de hipótesis).

   b) Indique si posee la misma distribución que la variable R definida anteriormente (realice una prueba de hipótesis).

   c) Indique si se puede suponer que la siguiente muestra, aleatoria e independiente de la anterior, tiene la misma distribución (realice una prueba de hipótesis).

0,79	1,68	0,5	1,18	0,87	0,5	0,76	0,2	2,22	1,69

   En todas las pruebas de hipótesis indique el p-valor y utilice el criterio de decisión p-valor$> 0,1$.

Resolución del examen de Probabilidad y Estadística

Ejercicio 1

1) Llamemos A_1 al suceso depositar todos los cupones el día 1, A_2 al suceso depositar un cupón por día durante 10 días y G al suceso ganar (donde ganar es ganar al menos un auto).

$$P(X_1 \bigcap G) = 10/10000 = 1/1000$$
$$P(X_2 \bigcap G) = 1 \text{ - } P((X_2 \bigcap G)^c)$$

En este caso P$((X_2 \bigcap G)^c)$ es la probabilidad de perder 10 días seguidos depositando 1 cupón al día, como los sorteos son independientes tenemos que

$$P((X_2 \bigcap G)^c) = (9999/10000)^{10} \approx 0{,}99900044.$$

Tenemos así que

$$P((X_1 \bigcap G)^c) = 999/1000 > 0{,}99900044 \approx P((X_2 \bigcap G)^c).$$

Por tanto la mejor estrategia es depositar todos los cupones el primer día.

2) Sea $X \sim \mathcal{P}(2)$ la variable aleatoria que modela la cantidad de cupones por persona, queremos calcular entonces $P\left(\sum_{i=1}^{4000} X_i > 8250\right)$ donde X_i i.i.d. $\mathcal{P}(2)$. Por tanto por teorema central del límite tenemos que

$$P\left(\sum_{i=1}^{4000} X_i > 8250\right) = P\left(\frac{\sum_{i=1}^{4000} X_i - n\mu}{\sigma\sqrt{n}} > \frac{8250 - n\mu}{\sigma\sqrt{n}}\right) \approx$$
$$P\left(N(0,1) > \frac{8250 - n\mu}{\sigma\sqrt{n}}\right).$$

En este caso $\mu = 2$, $\sigma = \sqrt{2}$ y $n = 4000$

Por tanto $P\left(N(0,1) > \dfrac{250}{\sqrt{4000}\sqrt{2}}\right) \approx 0{,}0026$

3) Notemos como U el suceso de que 4000 personas concurran al supermercado y V al suceso que se sorteen menos de 8000 cupones, queremos calcular $P(U \bigcap V)$. Por probabilidad condicional tenemos que $P(U \bigcap V) = P(V|U)P(U)$, ya que $P(U) \neq 0$. Sea $Y \sim \mathcal{P}(4000)$ la variable aleatoria que modela la cantidad de personas que llegan al supermercado.

$$P(U) = P(Y = 4000) = P(3999 < Y < 4001).$$

A su vez $Y = \sum_{i=1}^{4000} Y_i$ donde $Y_i \sim \mathcal{P}(1)$ independientes, por tanto por teorema central del limite

$$P(U) \approx P(-0{,}016 < N(0,1) < 0{,}016) \approx 0{,}01.$$

Luego $P(U|V)$ se calcula igual que en la parte anterior, es decir cuál es la probabilidad de que se sorteen menos de 8000 cupones dado que fueron 4000 personas, esto nos da $P(N(0,1) < 0)$.

Por tanto $P(U \bigcap V) \approx 0{,}005$

4) Por teorema central del límite tenemos que $P\left(\left|\frac{\overline{X_n} - \lambda}{\sigma\sqrt{n}}\right| > 2z_{0,05}\right) \approx P(|N(0,1)| > 2z_{0,05}) = 0{,}10$

donde $z_{0,05} \approx 1{,}64$

por tanto $\lambda \in \left[\overline{X_n} - \frac{\sigma z_{0,10}}{\sqrt{n}}, \overline{X_n} - \frac{\sigma z_{0,10}}{\sqrt{n}}\right]$

En este caso $n = 4500$, $\overline{X_n} = 2$ y podemos aproximar σ por $\sigma \approx \sqrt{\overline{X_n}}$.

Concluimos que un intervalo de confianza al 90 % es

$$\left[2 - \sqrt{2}\frac{1{,}64}{\sqrt{4500}}, 2 + \sqrt{2}\frac{1{,}64}{\sqrt{4500}}\right] = [1{,}96, 2{,}04]$$

Ejercicio 2

1) Sean $X, Y \sim U[0,a]$ las variables aleatorias que simulan la llegada del primer y segundo bus respectivamente y Z la variable aleatoria que modela el tiempo de espera en la parada del ómnibus, esta claro que $P(Z < 0) = 0$

Dado que el pasajero llego en un tiempo $0 \le T \le a$ la probabilidad de que se tome el primer omnibus es lo mismo a que el primer ómnibus no pase antes del momento T, es decir

$$P(Z \le a - T) = P(X > T) = \frac{a - T}{a}$$

mas en general tenemos que

$$P(Z \le t - T) = P(a - z \ge X \ge T)\ \forall T \le t \le a$$

Es claro que $P(Z \le a - T) = P(Z \le 1 - T)$ es decir que si el primer bus pasa antes de T entonces tiene seguro tiene que esperar hasta la 1.

Veamos ahora $P(Z \ge t - T)$ para $t \ge 1$.

El suceso el pasajero espera más de $t - T$, para $1 \le t \le 1 + a$ es que el primer ómnibus paso antes de T y el segundo no ha pasado antes de t, por probabilidad condicional tenemos entonces que

$$P(Z \ge t - T) = P(Y > t|X < T)P(X < T) = P(Y > t)P(X < T) = \left(\frac{(a + 1 - t)(T)}{a^2}\right)$$

Tenemos así la función de distribución de Z

$$F_Z(z) = \begin{cases} 0 & \text{si } z < 0 \\ z/a & \text{si } z \in [0, a - T] \\ (a - T)/a & \text{si } z \in [a - T, 1 - T] \\ 1 - \left(\frac{(a + 1 - (z + T))(T)}{a^2}\right) & \text{si } z \in [1 - T, 1 - T + a] \\ 1 & \text{si } z > 1 - T + a \end{cases}$$

Tenemos así que F_z es continua, además concluimos que la densidad es como queríamos

2)

$$E(Z) = \int_{-\infty}^{\infty} (t)(f_Z(t))dt = \int_{0}^{1+a-T} (t)(f_Z(t))\,dt =$$
$$\int_{0}^{a-T} t/a\,dt + \int_{1-T}^{1+a-T} tT/a^2\,dt =$$
$$\frac{(a-T)^2}{2a} + \frac{T[(1-T+a)^2-(1-T)^2]}{2a^2} =$$
$$\frac{a^2-aT+2T-T^2}{2a}$$

Tenemos así que $E(Z)$ en función de T, para minimizar alcanza con derivar en función de T, ver donde se anula y comparar con los extremos, es decir $T=0$ y $T=a$

$$\frac{\partial E(Z)}{\partial T} = \frac{2-a-2T}{2a}$$

se anula cuando $T=(2-a)/2$

Tenemos así que

$$\begin{cases} \text{si } T=0 & E(Z)=a/2 \\ \text{si } T=(2-a)/2 & E(Z)=\dfrac{(5/4)a^2+1-a}{2a} \\ \text{si } T=a & E(Z)=1-a/2 \end{cases}$$

Como $0<a<1$ tenemos $a/2 \le 1-a/2$, ademas

$$\frac{(5/4)a^2+1-a}{2a} > \frac{(5/4)a^2}{2a} > a/2$$

Concluimos así que el T que minimiza el tiempo de espera es $T=0$

3)

$$\overline{Z_n} \approx E(Z) = \frac{a^2-aT+2T-T^2}{2a}$$

tomando $T=1/5$ y $\overline{Z_n}=1/2$ tenemos

$$a = a^2 - a/5 + 2/5 - 1/25$$
$$a^2-(6/5)a+9/25=0$$
$$a = \frac{(6/5)\pm\sqrt{(36/25)-(36/25)}}{2} = 3/5$$

Ejercicio 3

1) Sea $X, Y \sim N(0,1)$ independientes tenemos entonces que

$$P((x,y)\in S)=\int_S f_X(z)f_Y(z)\,dz$$

en particular si $R=\sqrt{X^2+Y^2}$, tenemos asi que

$$P(R<t)=\int_{D(0,t)} f_X(z)f_Y(z)\,dz=\int_{D(0,t)}\frac{e^{-(x^2+y^2)/2}}{2\pi}\,dz=$$
$$=\int_0^t\int_0^{2\pi}\frac{re^{-r^2/2}}{2\pi}\,d\theta\,dr=\int_0^t\frac{re^{-r^2/2}}{2\pi}\,dr=1-e^{t/2}$$

2)
a) El test de Spearman de esta muestra nos deja

i	X_i	$R(i)$	$(R(X_i)-i)^2$
1	0.9	4	9
2	1.85	7	25
3	0.3	1	4
4	2.50	10	36
5	1.92	8	9
6	1.95	9	9
7	1.76	6	1
8	0.92	5	9
9	0.62	3	36
10	0.31	2	64
		Suma	202
		r_S	-0.22
		$p-$valor	0.268

El test de Rachas nos deja

i	X_i	$U(i)$	Fin de Racha
1	0.9		
2	1.85	1	1
3	0.3	0	1
4	2.50	1	1
5	1.92	0	1
6	1.95	1	1
7	1.76	0	
8	0.92	0	
9	0.62	0	
10	0.31	0	1
		R	6
		$p-$valor	0.7573

Por tanto no descartamos la hipotesis de que sean independientes.

b) Realizando el test de Kolmogorov-Smirnov no da

i	X_i^*	$R(X_i^*)$	$\lvert R(X_i^*) - i/n\rvert$	$\lvert R(X_i^*) - (i-1)/n\rvert$
1	0.3	$1-e^{-0,045}$	0.056	
2	0.31	$1-e^{-0,055}$	0.156	0.056
3	0.62	$1-e^{-0,1922}$	0.125	0.0250
4	0.9	$1-e^{-0,405}$	0.067	0.033
5	0.92	$1-e^{-0,406}$	0.167	0.067
6	1.76	$1-e^{-1,549}$	0.188	0.288
7	1.85	$1-e^{-1,711}$	0.119	0.219
8	1.92	$1-e^{-1,843}$	0.041	0.141
9	1.95	$1-e^{-1,901}$	0.049	0.0506
10	2.50	$1-e^{-3,125}$	0.044	0.056

$R(X_0^*) = 1 - e^{-0,045} = 0,362$

$$1 - R(X_{10}^*) = e^{-3,125} = 0,044$$

Por tanto $D_{10} = 0,362$ y el $p-$valor es mayor a 0.2

c)Por el test de comparación de Kolmogorov-Smirnov tenemos, para Y_n la segunda muestra

i	Y_i	Y_i^*
1	0.79	0.2
2	1.68	0.5
3	0.5	0.5
4	1.18	0.76
5	0.87	0.79
6	0.5	0.87
7	0.76	1.18
8	0.2	1.68
9	2.22	1.69
10	1.69	2.22

$D_{10,10} = sup_{t\in\mathbb{R}}\{\lvert F_{10}^X(t) - F_{10}^Y(t)\rvert\}$ En nuestro caso $D_{10,10} = 0,4$ que tiene un $p-$valor mayor a 0.2