Universidad de la República | **Probabilidad y Estadística**
Facultad de Ingeniería - IMERL | **Segundo semestre 2015**

Segundo parcial
Martes 1 de diciembre de 2015

Número de Parcial	Cédula	Nombre y Apellido

Ejercicio 1. Se sabe que en una cierta localidad el 5% no tiene computadora en su casa, el 45% tiene una computadora y el 50% tiene dos computadoras. Las computadoras pueden ser laptops o PCs, y la probabilidad de que una computadora sea una laptop es $p = 0.3$. Se elige una persona al azar de dicha comunidad y se definen las siguientes variables aleatorias:

- X = número de computadoras de la persona.
- Y = número de laptops de la persona.

1. Indicar $\mathcal{R}_X$ recorrido de la variable X y su función de probabilidad, esto es $P(X = k)$ con $k \in \mathcal{R}_X$.
2. Sabiendo que $X = k$, hallar la distribución de Y.
3. Hallar la función de probabilidad conjunta de X e Y. Sugerencia: dar el resultado en forma de cuadro.
4. Hallar la función de probabilidad marginal de Y.
5. Calcular $\mathrm{E}(X)$ y $\mathrm{E}(Y)$.
6. ¿Son X e Y independientes? Hallar $\mathrm{Cov}(X, Y)$.
7. Hallar la probabilidad de que el número de laptops de la persona supere al número de PCs.

Ejercicio 1. Solución

1. $\mathcal{R}_X = \{0, 1, 2\}$ donde $P(X = 0) = 0.05$, $P(X = 1) = 0.45$ y $P(X = 2) = 0.5$.
2. Sabiendo que $X = k$, la distribución de $Y \sim \mathrm{Bin}(k, p)$ es decir binomial de parámetros k y $p = 0.3$. Observar que si $k = 0$, entonces $P(Y = 0) = 1$.

   Más en detalle, tenemos que:

   - Si $X = 1$, entonces $P(Y = 0) = 0.7$ y $P(Y = 1) = 0.3$.
   - Si $X = 2$, entonces $P(Y = 0) = 0.7^2 = 0.49$, $P(Y = 1) = 2 \times 0.3 \times 0.7 = 0.42$ y $P(Y = 2) = 0.3^2 = 0.09$.
3. Observar que $P(X = i, Y = j) = P(Y = j|X = i)P(X = i)$ donde $P(X = i)$ con $i \in \{0, 1, 2\}$ fue hallada en la parte 1 y $P(Y = j|X = i) = P(Z = j)$ con $Z_i \sim \mathrm{Bin}(i, 0.3)$, esto es $P(Y = j|X = i) = C_j^i p^j (1 - p)^{j-1}$ para $0 \leq j \leq i$. Por lo tanto tenemos que:

$$P(X = 0, Y = 0) = P(Y = 0|X = 0)P(X = 0) = 1 \times 0.05.$$

$$P(X=1,Y=0)=P(Z_1=0)P(X=1)=0.7\times 0.45=0.315,$$
$$P(X=1,Y=1)=P(Z_1=1)P(X=1)=0.3\times 0.45=0.135.$$

$$P(X=2,Y=0)=P(Z_2=0)P(X=2)=0.7^2\times 0.5=0.245,$$
$$P(X=2,Y=1)=P(Z_2=1)P(X=2)=2\times 0.3\times 0.7\times 0.5=0.21,$$
$$P(X=2,Y=2)=P(Z_)=2)P(X=2)=0.3^2\times 0.5=0.045.$$

Podemos resumir los resultados en el siguiente cuadro:

$Y\|X$	0	1	2	p_Y
0	0.05	0.315	0.245	0.61
1	0	0.135	0.21	0.345
2	0	0	0.045	0.045
p_X	0.05	0.45	0.5	1

4. La función de probabilidad marginal de Y la calculamos sumando por filas en el cuadro anterior, de donde se obtiene que $P(Y=0)=0.61$, $P(Y=1)=0.345$ y $P(Y=2)=0.045$. Observar que la función de probabilidad marginal de X se obtiene sumando por columnas en el cuadro anterior y se recuperan los valores hallados en la parte 1 del ejercicio.

5.
$$\mathrm{E}(X)=0\times P(X=0)+1\times P(X=1)+2\times P(X=2)=0.45+2\times 0.5=1.45,$$
$$\mathrm{E}(Y)=0\times P(Y=0)+1\times P(Y=1)+2\times P(Y=2)=0.345+2\times 0.045=0.435.$$

6. Es claro que X e Y no son independientes (ver parte 2). También es fácil deducirlo del cuadro de la función de probabilidad conjunta de la parte 3, pues hay valores de la conjunta que son nulos para valores no nulos de las marginales.

   Para hallar $\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)$, es necesario calcular $\mathrm{E}(XY)$. Del cuadro de la conjunta podemos ver que $P(XY=0)=0.61$, $P(XY=1)=0.135$, $P(XY=2)=0.21$ y $P(XY=4)=0.045$. Por lo tanto:
$$\begin{aligned}\mathrm{E}(XY) &= 0\times P(XY=0)+1\times P(XY=1)+2\times P(XY=2)+4\times P(XY=4)\\ &= 0.345+2\times 0.21+4\times 0.045=0.735.\end{aligned}$$
   De donde $\mathrm{Cov}(X,Y)=0.735-1.45\times 0.435=0.10425$.

7. La probabilidad de que el número de laptops supere al número de PCs, es $P(Y>X-Y)$. Por lo tanto hay que calcular $P(2Y>X)$:
$$P(2Y>X)=\sum_{i=0}^{2}P(2Y>X,X=i)=P(X=1,Y=1)+P(X=2,Y=2)=0.135+0.045=0.18.$$

Ejercicio 2. Sea $f:\mathbb{R}\to\mathbb{R}$ la función dada por:
$$f(x)=\begin{cases} ax & \text{si } x\in[0,1]\\ b(2-x) & \text{si } x\in(1,2]\\ 0 & \text{otro caso}\end{cases}$$
donde a y b son constantes reales positivas.

1. Hallar el conjunto de valores de a y b para los cuáles f es la densidad de una variable aleatoria X. Sugerencia: graficar f.
2. Definir un estimador consistente de a. ¿Es insesgado? Justifique su respuesta.
3. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de la variable X, es decir variables aleatorias independientes e idénticamente distribuidas con densidad f. A partir de la muestra definir un intervalo de confianza aproximado a nivel α para el parámetro a.
4. (a) Calcular m_X mediana de X, suponiendo que m_X es menor que 1.
   (b) Definir un nuevo estimador de a, diferente al hallado en la parte 2b.
5. Se toma una muestra de la variable aleatoria X y se observa que el $76,3\%$ de los datos son menores que 1. A partir de esta información, dar estimaciones de los parámetros a y b.

Ejercicio 2: Solución

1. Para que f sea una densidad, necesitamos que $\int_{\mathbb{R}} f(x)dx = 1$. Como la función de densidad está definida por dos triángulos, para calcular el área podemos hacer directamente el cálculo de las áreas de dichos triángulos, esto es $\frac{a(1-0)}{2} + \frac{b(2-1)}{2} = 1$. Por lo tanto, $a + b = 2$.
   Además necesitamos que f sea positiva, pero esto es directo pues $a, b > 0$ por hipótesis. Entonces, el conjunto de los valores que pueden tomar a y b es
   $$\{a, b \in \mathbb{R}^+ : \ a + b = 2\}$$
2. $$\mathrm{E}(X) = \int_{\mathbb{R}} xf(x)dx = a\int_0^1 ax^2dx + b\int_1^2 (2x - x^2)dx = \frac{1}{3}(a + 2b)$$
   Usando la condición anterior $a + b = 2$, resulta que $\mathrm{E}(X) = \frac{1}{3}(4 - a)$ y por lo tanto $a = 4 - 3\mathrm{E}(X)$. Usando la ley fuerte de los grandes números, podemos estimar $\mathrm{E}(X)$ con $\bar{X}_n$. Este estimador es un estimador consistente de la esperanza y por lo tanto $\hat{a}_n = 4 - 3\bar{X}_n$ es un estimador consistente de a.
   Para ver si el estimador de a es insesgado, calculamos su esperanza:
   $\mathrm{E}(\hat{a}_n) = \mathrm{E}(4 - 3\bar{X}_n) = 4 - 3\mathrm{E}(\bar{X}_n) = 4 - 3\mathrm{E}(X) = 4 - 3\frac{1}{3}(4 - a) = a$. Por lo tanto $\hat{a}_n$, el estimador por momentos de a, es insesgado.
3. Sabemos que un intervalo de confianza aproximado al nivel α para la $\mathrm{E}(X)$ es:
   $$I_\alpha(\mathrm{E}(X)) = \left[\bar{X}_n - \frac{\sigma_n}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma_n}{\sqrt{n}} z_{\frac{\alpha}{2}}\right].$$
   Usando que $a = 4 - 3\mathrm{E}(X)$ (función decreciente), un intervalo de confianza para a al nivel α es:
   $$I_\alpha(a) = \left[4 - 3\left(\bar{X}_n + \frac{\sigma_n}{\sqrt{n}} z_{\frac{\alpha}{2}}\right), 4 - 3\left(\bar{X}_n - \frac{\sigma_n}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)\right].$$
4. (a) Para $x \leq 1$, por definición tenemos que $F(x) = \int_0^x atdt = \frac{a}{2}x^2$. La mediana m, cumple que $F(m) = \frac{1}{2} \Rightarrow \frac{a}{2}m^2 = \frac{1}{2} \Rightarrow a = \frac{1}{m^2}$.
   (b) Sabemos que $\hat{m}_n$ la mediana empírica de los datos $X_1, X_2, \ldots X_n$ es un estimador de m. Por lo tanto $\overline{a}_n = \frac{1}{\hat{m_n}^2}$ es otro estimador de a.

5. Sea $p = P(X \leq 1) = F(1)$, entonces de los datos tenemos que una estimación de p es $\hat{p}_n = 0,763$. Además por definición tenemos que $p = F(1) = \frac{a}{2} \Rightarrow \hat{a}_n = 2\hat{p}_n = 1,526$ es una estimación de a. Usando que $b = 2 - a$, resulta que $\hat{b}_n = 0,474$ es una estimación de b.

Ejercicio 3. La cantidad de líquido (en ml) dispensada por una máquina embotelladora de refrescos puede modelarse como una variable aleatoria $X \sim \mathcal{N}(1000, 10^2)$, es decir normal con parámetros $\mu = 1000$ y $\sigma = 10$. Durante unos días la máquina falla y dispensa menos líquido del debido. En ese período la cantidad (en ml) de líquido por envase se comporta según una variable aleatoria $Y \sim \mathcal{N}(995, 10^2)$. Luego la máquina es reparada y vuelve a dispensar líquido según su distribución habitual X.

Sin embargo, la empresa tarda en detectar la falla y por lo tanto hay varios lotes que no se sabe si fueron envasados durante la falla o no. La empresa no quiere poner a la venta los lotes fallados y encarga al ingeniero de planta decidir cuáles de esos lotes pueden ser puestos a la venta y cuáles deben ser retirados. Se toma uno de estos lotes (de 12 botellas) y se registran la cantidad de líquido de cada botella.

1. Plantear un test de hipótesis parámetrico que permita al ingeniero decidir si el lote es apto para la venta o no. Para esto, indicar:
   (a) hipótesis nula y alternativa,
   (b) región crítica a nivel α.
2. Para el test definido en la parte anterior y dado $\alpha = 0.05$, calcular $\beta = P(\text{error tipo II})$.
3. Para uno de los lotes revisados, se tienen las siguientes cantidades:

995	998	997	1003	1007	999	1002	1009	1010	985	1005	981

   Determinar si el lote puede ser puesto a la venta para los niveles de confianza $\alpha = 0.1$, $\alpha = 0.05$ y $\alpha = 0.01$. Hallar α^*, p-valor del test para esta muestra.
4. (a) Asumiendo que la máquina funciona correctamente, calcular p_0 la probabilidad de que el líquido en la botella sea menor o igual que 995ml.
   (b) A partir de la muestra anterior construir un intervalo de confianza a nivel $\alpha = 0.05$ para p_0.
   (c) En base a la parte anterior ¿le parece razonable asumir que el lote puede ser puesto a la venta? Justifique su respuesta.

Ejercicio 3: Solución

1. Sabiendo que los datos son normales con media μ y varianza conocida ($\sigma = 10$) un posible test parámetrico a plantear es el siguiente: Sean $X_1, X_2, \dots X_n$ iid con distribución $\mathcal{N}(\mu, 10^2)$, se testea:

$$\begin{cases} H_0: & \mu = \mu_0 = 995 \\ H_1: & \mu = 1000 \end{cases}$$

   Tomamos $\mu = 1000$ como hipótesis alternativa porque de ese modo el error de tipo I (que podemos acotar) es la probabilidad de vender un lote envasado durante la falla.
   Luego, una región crítica para este test (a nivel α) es conocida y está dada por:

$$\mathcal{R}_\alpha = \left\{ \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma} \geq z_\alpha \right\}$$

2. Si $\alpha = 0.05$, entonces la región crítica es

$$\mathcal{R}_{0.05} = \left\{ \frac{\sqrt{n}(\overline{X}_n) - \mu_0)}{\sigma} \geq 1.645 \right\}.$$

Luego

$$\beta = P\left(\frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma} < 1.645 | \mu = 1000 \right)$$

.

Sustituyendo los datos disponibles tenemos entonces

$$\begin{aligned} \beta &= P\left(\frac{\sqrt{12}(\overline{X}_n - \mu_0)}{\sigma} < 1.645 | \mu = 1000 \right) \\ &= P(\overline{X}_n < \frac{10}{\sqrt{12}} 1.645 + 995 | \mu = 1000) \\ &= P\left(\frac{\sqrt{12}(\overline{X}_n - 1000)}{10}) < \left(\frac{10}{\sqrt{12}} 1.645 + 995 - 1000 \right) \frac{\sqrt{12}}{10} | \mu = 1000 \right) \\ &= \Phi(-0.087) = 0.535. \end{aligned}$$

3. Dado que el promedio de los datos es $999,25$, se tiene por la parte 1 que la región crítica al nivel α es

$$\mathcal{R}_\alpha = \left\{ \frac{\sqrt{12}(999,25 - 995)}{10} \geq z_\alpha \right\} = \{1.472 \geq z_\alpha\}$$

.

Por lo tanto:

- si $\alpha = 0.1$, $z_\alpha = 1.28$, con lo cual rechazo H_0.
- si $\alpha = 0.05$, $z_\alpha = 1.64$, con lo cual no rechazo H_0.
- si $\alpha = 0.01$, de lo anterior se deduce que tampoco se rechazará H_0.

De lo anterior sabemos que el p-valor para la muestra $\alpha^* \in [0.05, 0.1)$, más exactamente $\alpha^* = 1\Phi(1.472) = 1 - 0.93 = 0.07$.

4. (a) $P(X \leq 995) = \Phi(\frac{995-1000}{10}) = \Phi(\frac{-1}{2}) = 0.31$.

   (b) Para estimar p_0 se utiliza la proporción de observaciones que son menores o iguales a 995. Así $\hat{p}_0 = \frac{2}{12} = \frac{1}{6}$. Luego busco el menor intervalo $I_1 = [a, b]$ que contenga a $\hat{p}_0$ y tal que $P(\hat{p}_0 \in [a, b]) \geq 0.95$. Esto equivale a buscar ϵ tal que $P(|\hat{p}_0 - p_0| \leq \epsilon) \geq 1 - \alpha$.

   Observar que $\hat{p}_0 = \frac{1}{12} \sum_{i=1}^{12} \mathbf{1}_{\{X_i \leq 995\}} = \frac{S_{12}}{12}$, donde $S_{12} \sim \mathrm{Bin}(12, p_0)$ por ser suma de Bernoullis de parámetro p_0.
   Por lo tanto, buscamos ϵ tal que:

   $$P(12(-\epsilon - 0.31) \leq S_{12} \leq (\epsilon + 0.31)12)) = P(0 \leq S_{12} \leq (\epsilon + 0.31)12)) \geq 1 - \alpha.$$

   De una tabla de la distribución Binomial de parámetros $n = 12$ y $p_0 = 0.31$ obtiene que $(\epsilon + 0.31)12 = 6$ de donde el intervalo buscado es $I_1 = [0, 1/2]$.
   Otra posibilidad es utilizar un intervalo de confianza aproximado, es decir

   $$I_2 = \left[\hat{p}_0 - \frac{z_{\alpha/2}\sqrt{\hat{p}_0(1 - \hat{p}_0)}}{\sqrt{n}}, \hat{p}_0 + \frac{z_{\alpha/2}\sqrt{\hat{p}_0(1 - \hat{p}_0)}}{\sqrt{n}} \right] = [0.05, 0.571]$$

   (c) Como $0,31 \in I_1$, la conclusión sería que el lote puede ser puesto a la venta. La misma conclusión resulta de utilizar el intervalo de confianza aproximado I_2.