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

EXAMEN
Sábado 19 de diciembre de 2015

Número de Parcial	Cédula	Nombre y Apellido

Ejercicio 1. El tiempo de vida en años de las baterías fabricadas por una determinada empresa A, puede ser modelado por una variable aleatoria T. Se asume que T tiene distribución exponencial de parámetro λ.

1. Si la empresa asegura que sus baterías duran en promedio 5 años. ¿Cuál es el valor de λ? Justificar.

En lo que sigue se asume que $\lambda = 0.2$.

2. (a) Se sabe que una batería funciona desde hace 4 años. ¿Cuál es la probabilidad de que funcione al menos seis años?
   (b) Se considera que la batería es defectuosa si su tiempo de vida es menor a 6 años . Si en un componente se instalan 5 baterías de forma independiente, ¿cuál es la probabilidad de que al menos una de ellas sea defectuosa?
   (c) ¿ Cuál es el número mínimo de baterías que deben instalarse de modo que dicha probabilidad sea mayor que 0.95?
3. Existe otra empresa que fabrica baterías (que llamaremos B) que asegura que la probabilidad de que sus baterías sean defectuosas es igual a 0.1. Se compran 35 baterías de la empresa A y 15 de la empresa B. De estas 50 se eligen 2 sin reposición. ¿Cuál es la probabilidad de que ambas sean defectuosas?

Ejercicio 1. Solución

1. Como las bateras duran en promedio 5 aos, tenemos que $E(T) = 5$. Como T es exponencial, sabemos que $E(T) = \frac{1}{\lambda}$. Por lo tanto $\lambda = 1/5 = 0.2$.
2. (a) Usando la propiedad de pérdida de memoria de la exponencial (o la definición de probabilidad condicional) se tiene $P(T > 6|T > 4) = P(T > 2) = e^{-2/5} \approx 0.67$.
   (b) Sea X el número de baterías defectuosas entre las 5 baterías instadas. Por ser independientes se tiene que $X \sim \text{Bin}(5, p)$ donde $p = P($ batería defectuosa$) = P(T \leq 6) = 1 - e^{-6/5} \approx 0.7$. La probabilidad entonces de que al menos una sea defectuosa es
   $$P(X \geq 1) = 1 - P(X < 1) = P(X = 0) = 1 - 0.3^5 \approx 0.998$$
   (c) Sea n el número de baterías instaladas. Con el mismo razonamiento que en la parte anterior se obtiene que la probabilidad de que al menos una batería funcione es $1 - 0.3^n$. Se busca n tal que dicha probabilidad sea mayor que 0.95. Esto es:
   $$1 - 0.3^n > 0.95 \Rightarrow 0.3^n < 0.05 \Rightarrow n > \frac{\log(0.05)}{\log(0.3)} \Rightarrow n > 2.4.$$
   Por lo tanto el mínimo n es 3.

3.

$$P(\text{2 sean defectuosas}) = P(\text{2 sean defectuosos, las dos de A}) +$$
$$P(\text{2 sean defectuosas, los dos de B}) + P(\text{2 sean defectuosos, uno de A y uno de B})$$

donde:

$$P(\text{2 sean defectuosos, los dos de A}) = P(\text{2 sean defectuosos}|\text{ los dos de A})P(\text{los dos de A})$$
$$= 0.3^2 \times \frac{35}{50} \times \frac{34}{49} = 0.044,$$

$$P(\text{2 sean defectuosos, los dos de B}) = P(\text{2 sean defectuosos}|\text{ los dos de B})P(\text{los dos de B})$$
$$= 0.1^2 \times \frac{15}{50} \times \frac{14}{49} = 0.03,$$

$$P(\text{2 sean defectuosos, uno de cada cada uno}) = P(\text{2 sean defectuosos}|\text{uno de A y uno de B})P(\text{uno de A y uno de B}$$
$$= 0.3 \times 0.1 \times 2 \cdot \frac{35}{50} \times \frac{15}{50} = 0.013.$$

Por lo tanto: $P(\text{2 sean defectuosos}) = 0.044 + 0.03 + 0.013 = 0.087$

Ejercicio 2. Se considera la función $f : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} ax^{-(a+1)} & \text{si } x \geq 1 \\ 0 & \text{en otro caso} \end{cases}$$

donde $a \in \mathbb{R}$ es tal que $a > 2$.

1. Verificar que f es la función densidad para alguna variable aleatoria X.
2. Hallar la función de distribución de X.
3. (a) Hallar $\mathrm{E}(X)$, $\mathrm{Var}(X)$ y m_X, esperanza, varianza y mediana de X.
   (b) Sea $X_1, X_2, \ldots, X_n$ una muestra independiente e idénticamente distribuida de variables aleatorias con la densidad definida antes. Hallar el estimador por momentos de a. Probar que es consistente.
4. Dado $c > 0$, se define la variable aleatoria $Y = cX$:
   (a) Hallar la función de distribución de Y.
   (b) Hallar $\mathrm{E}(Y)$ y $\mathrm{Var}(Y)$.
5. (a) Probar que la variable $T = \log(X)$ tiene distribución exponencial de parámetro a.
   (b) Sea $X_1, X_2, \ldots, X_n$ una muestra independiente e idénticamente distribuida de variables aleatorias con la densidad definida antes. Utilizando la parte anterior construir un nuevo estimador de a diferente del hallado en la parte 3.b.

Ejercicio 2. Solución

1. Se observa que $f(x) \geq 0 \ \forall x \in \mathbb{R}$.
   Además:
   $$\int_{-\infty}^{\infty} f(x)dx = \int_{1}^{\infty} ax^{-(a+1)}dt = -x^{-a}|_{1}^{\infty} = 1$$
   Por lo tanto, es una densidad para alguna v.a. X.

2. $$F_X(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0dt = 0 \text{ si } x < 1$$
   $$F_X(x) = \int_{-\infty}^{x} f(t)dt = \int_{1}^{x} at^{-(a+1)}dt = -t^{-a}|_{1}^{x} = 1 - x^{-a} \text{ si } x \geq 1$$
   Por tanto:
   $$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - x^{-a} & \text{si } x \geq 1 \end{cases}$$

3. (a) $$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{1}^{\infty} ax^{-a}dx = \left.\frac{ax^{-a+1}}{-a+1}\right|_{1}^{\infty} = \frac{a}{a-1}$$
   $$\text{Var}(X) = E(X^2) - E(X)^2 = E(X^2) - \left(\frac{a}{a-1}\right)^2,$$
   donde $E(X^2) = \int\limits_{1}^{\infty} ax^{-a+1}dx = \frac{a}{a-2}$
   Entonces:
   $$\text{Var}(X) = \frac{a}{a-2} - \left(\frac{a}{a-1}\right)^2 = \frac{a}{(a-2)(a-1)^2}$$
   Finalmente, buscamos m_X tal que $F_X(m_X) = \frac{1}{2}$. Por tanto $1 - m_X^{-a} = \frac{1}{2}$. Entonces, despejando, $m_X = \sqrt[a]{2}$.

   (b) Utilizando el método de los momentos planteamos $\bar{X_n} = \frac{\hat{a}}{\hat{a}-1}$. Despejando, $\hat{a} = \frac{\bar{X_n}}{\bar{X_n}-1}$, que es consistente pues $f(x) = \frac{x}{x-1}$ es continua si $x \neq 1$.

4. (a) $F_Y(y) = P(Y \leq y) = P(cX \leq y) = P(X \leq \frac{y}{c}) = F_X(\frac{y}{c})$
   Por tanto:
   $$F_Y(y) = \begin{cases} 0 & \text{si } y < c \\ 1 - (\frac{y}{c})^{-a} & \text{si } y \geq c \end{cases}$$

   (b) Utilizando las propiedades del valor esperado, se tien que:
   $$E(Y) = E(cX) = cE(X) = \frac{ca}{a-1}$$
   $$\text{Var}(Y) = \text{Var}(cX) = c^2\text{Var}(X) = \frac{c^2a}{(a-2)(a-1)^2}$$

5. (a) $F_T(t) = P(T \leq t) = P(\log X \leq t) = P(X \leq e^t) = F_X(e^t)$.
   Por tanto:
$$F_T(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - e^{-at} & \text{si } t \geq 0 \end{cases}$$
   Entonces $T \sim exp(a)$.

   (b) Se define $T_i = \log(X_i)$, entonces $T_1, T_2, \ldots, T_n$ una muestra iid. Utilizando la ley fuerte de los grandes números se tiene que: $\bar{T_n} = \frac{1}{n} \sum_{i=1}^{i=n} T_i = \frac{1}{n} \sum_{i=1}^{i=n} \log(X_i)$ converge casi seguramente a $\frac{1}{\hat{a}}$ por ser $T \sim exp(a)$.

   Así, un nuevo estimador es $\hat{a} = \frac{n}{\sum_{i=1}^{i=n} \log(X_i)} = \frac{1}{\bar{T_n}}$.

Ejercicio 3. En una red de datos de gran dimensión se definen las siguientes categorías para cada enlace en base a la carga (C) de dicho enlace.

- Baja si $\mathrm{C} \leq 0.4$,
- Media si $0.4 < \mathrm{C} \leq 0.7$,
- Alta si $\mathrm{C} > 0.7$.

Se asume que la carga C de un enlace sigue una distribución Gaussiana con parámetros $\mu = 0.5$ y $\sigma^2 = 0.01$.

1. Hallar la probabilidad de que un enlace pertenezca a cada uno de las categorías definidas antes.
2. Se sabe que la probabilidad de perder datos para un enlace de la categoría Media es de 0.15, mientras que dicha probabilidad se eleva a 0.75 dentro de la categoría Alta. Finalmente, la probabilidad de pérdida de un enlace con carga Baja es de 0.1. Se pide:
   (a) Calcular la probabilidad de que un enlace elegido al azar de la red pierda datos.
   (b) Dado que un enlace tiene pérdida de datos, ¿Cuál es la probabilidad de que pertenezca a la categoría Media?
3. Se decide realizar un cambio de ruteo en la red, de modo de disminuir la carga media de cada enlace. Se realizaron medidas de carga en 10 enlaces y se obtuvieron los siguientes resultados:

0.33	0.51	0.59	0.38	0.62	0.61	0.49	0.53	0.52	0.48

   (a) ¿Puede asumirse que la muestra es independiente e idénticamente distribuida? Realice un test de hipótesis.
   (b) ¿Puede asumirse que los datos siguen una distribución Gaussiana? Realice un test de hipótesis.
   (c) ¿Le parece razonable afirmar con un nivel de confianza de 95% que el nuevo ruteo disminuyó la carga media por enlace de la red? Justifique su respuesta.

Ejercicio 3. Solución

1. - $P(\mathsf{C} < 0.4) = P(\frac{\mathsf{C}-0.5}{0.1} < \frac{0.4-0.5}{0.1}) = P(Z < -1) = \Phi(-1) = 1 - \Phi(1) = 0.16$
   - $P(0.4 \leq \mathsf{C} \leq 0.7) = P(\mathsf{C} \leq 0.7) - P(\mathsf{C} \leq 0.4) = P(Z \leq \frac{0.7-0.5}{0.1}) - 0.16 = P(Z \leq 2) - 0.16 = \Phi(2) - 0.16 = 0.977 - 0.16 = 0.817$
   - $P(\mathsf{C} > 0.7) = 1 - P(\mathsf{C} \leq 0.7) = 1 - 0.977 = 0.023$

2. (a) Se definen los siguientes eventos $D = \{$El enlace elegido al azar pierda datos$\}$, $\mathsf{B} = \{$categoría Baja$\}$, $\mathsf{M} = \{$categoría Media$\}$ y $\mathsf{A} = \{$categoría Alta$\}$. Uilizando la fórmula de la probabildidad total se tiene que

$$\begin{aligned} P(PD) &= P(D|\mathsf{B})P(\mathsf{B}) + P(D|\mathsf{M})P(\mathsf{M}) + P(D|\mathsf{A})P(\mathsf{A}) \\ &= 0.1 \times 0.16 + 0.5 \times 0.817 + 0.75 \times 0.023 = 0.1558 \end{aligned}$$

   (b) $P(\mathsf{M}|D) = \frac{P(D|\mathsf{M})P(\mathsf{M})}{P(D)} = \frac{0.15 \times 0.817}{0.1558} = 0.787$

3. (a) Se realiza test de correlación de Rangos de Spearman. El estadístico es $R_s = 0.17$. Por lo tanto se plantea,

$$\begin{cases} H_0\colon \text{La muestra es aleatoria} \\ H_1\colon \text{Hay tendencia creciente} \end{cases}$$

   De la tabla se obtiene que $0.316 < \alpha^* < 0.328$. Asumiendo un nivel de confianza de 0.10 no se puede rechazar H_0 y por lo tanto es posible asumir que la muestra es aleatoria.
   Si se realiza el test de rachas, se obtiene $R = 6$. Para $n = 10$, se obtiene un p-valor $\alpha^* = 0.7573 > 0.1$. Se concluye entonces que no es posible rechazar que la muestra sea aleatoria.

   (b) $\overline{X}_n = 0.506$ y $s_n^2 = 0.0088 \Rightarrow s_n = 0.094$.
   Se realiza el test de Lilliefors para distribución normal

$$\begin{cases} H_0: F \sim \mathcal{N}(0.506, 0.094) \\ H_1: \text{ no } H_0 \end{cases}$$

   Se obtiene el estadístico $D_n = \max\{0.0307, 0.1151, 0.1897\} = 0.1897$. Mirando en la tabla de Lilliefors se observa que el p-valor α^* es mayor a 0.20 entonces con un nivel de confianza 0.10 no se puede rechazar la hipótesis nula.

   (c) Se realiza un test de hipótesis sobre μ con σ desconocido y $\alpha = 0.05$

$$\begin{cases} H_0\colon \mu \leq \mu_0 \\ H_1\colon \mu > \mu_0 \end{cases}$$

   La región crítica es

$$R_{0.05} = \left\{ \frac{\sqrt{10}(0.506 - 0.5)}{0.094} \geq t_\alpha(n-1) \right\} = \{0.202 \geq 1.83\}$$

   Por lo tanto no se rechaza H_0. Por lo que parece razonable afirmar que con un nivel de confianza del 95% el nuevo ruteo disminuyó la carga media por enlace de la red.