UNIVERSIDAD DE LA REPÚBLICA - Facultad de Ingeniería

Primer Parcial de PROBABILIDAD Y ESTADÍSTICA 2002 - 18 de mayo del 2002

N^o de PARCIAL

____

APELLIDOS *NOMBRE* *CEDULA DE IDENTIDAD*

Nombre de su docente de Teórico (si no asiste a teórico, escriba NO; si asiste regularmente a dos teóricos, escriba los dos nombres): _

Nombre de su docente de Práctico (si no asiste a práctico, escriba NO; si asiste regularmente a dos prácticos, escriba los dos nombres): _

Total de puntos: 40.
Duración de la prueba: 4 horas.
Puede emplear su material de estudio (libros, apuntes, calculadora, etc.)
Se ruega no plantear preguntas que no sean meras aclaraciones de letra.

Buena suerte !!

Problema 1 (13 puntos)

Considere dos bolilleros, A y B. El bolillero A tiene un total de 12 bolillas, de las cuales 10 son rojas y 2 azules. El bolillero B tiene un total de 10 bolillas, de las cuales 4 son rojas y 6 azules. Se realiza un sorteo para decidir de qué bolillero se van a extraer las bolillas. En este sorteo se tiene que

$$\mathbf{P}(\text{salga sorteado el bolillero } A) = \tfrac{1}{5}$$

$$\mathbf{P}(\text{salga sorteado el bolillero } B) = \tfrac{4}{5}$$

Una vez elegido un bolillero se extrae al azar una bolilla del bolillero seleccionado.

(a) (**5 puntos**) Calcular la probabilidad de sacar una bolilla roja.

(b) (**4 puntos**) Si se extrae una bolilla y es roja, ¿cuál es la probabilidad de que haya sido seleccionado en el sorteo el bolillero A?

(c) (**4 puntos**) Se modifica ahora el experimento: se realiza el sorteo como antes, pero se extraen 4 bolillas en lugar de una del bolillero sorteado con la condición adicional de que si el bolillero seleccionado es el A las extracciones son **con reposición**, en tanto que si el bolillero seleccionado es el B las extracciones son **sin reposición**. En esta nueva situación, calcular la probabilidad de sacar exactamente k bolillas rojas (indicar para que valores de k tiene sentido el cálculo anterior)

(**Sugerencia:** Considerar la distribución del número de bolillas rojas dentro de las 4 seleccionadas dado que el sorteo favoreció al bolillero A y hacer lo mismo para el bolillero B.)

Problema 2 (14 puntos)

Decimos que la variable aleatoria X tiene distribución de Maidana (y lo denotamos $X \sim \mathcal{M}$) si X es absolutamente continua con densidad

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{si } x \in [-1,1]; \\ 0 & \text{si no.} \end{cases}$$

Asumiendo $X \sim \mathcal{M}$,

(a) (**2 puntos**) Calcular la función de distribución de X.

(b) (**2 puntos**) Hallar a tal que $\mathbf{P}(X \leq a) = \mathbf{P}(X > a) = \frac{1}{2}$.

(c) (**3 puntos**) Calcular $\mathbf{E}(X)$, $\mathbf{E}(X^2)$ y $\mathbf{E}(X^4)$.

Considere variables aleatorias X_1, X_2 y X_3 *i.i.d.* $\sim \mathcal{M}$. Se definen las variables $U = X_1X_2$ y $V = X_2X_3$,

(d) i. (**2 puntos**) Calcular $\mathbf{E}(U)$, $\mathbf{E}(V)$, $\mathbf{E}(UV)$ y $\mathbf{Cov}(U,V)$.
   ii. (**2 puntos**) Calcular $\mathbf{E}\left(U^2\right)$, $\mathbf{E}(V^2)$ y $\mathbf{E}((UV)^2)$

(e) (**3 puntos**) ¿Son U y V independientes?. Justifique su respuesta.

Problema 3 (13 puntos)

Una pieza de una máquina se verifica al final de cada hora de producción y se cambia por una nueva en caso de encontrarse rota. El tiempo de vida en horas de la pieza se puede modelar con una variable aleatoria T con distribución exponencial de parámetro λ ($T \sim Exp(\lambda)$), por lo tanto el tiempo en horas que transcurre hasta el recambio de la pieza se puede modelar con una variable aleatoria $X = [T] + 1$, donde $[T]$ es la parte entera de T (esto es, $[T] = n$ si y sólo si $n \leq T < n+1$).

(a) i. (**3 puntos**) Hallar la función de probabilidad de la variable aleatoria X
   ii. (**1 punto**) Deducir que $X \sim Geo(p)$ (distribución geométrica de parámetro p) con $p = 1 - e^{-\lambda}$

(b) Denotaremos (de aquí y hasta el final del ejercicio) por X_1 el tiempo (en horas) de recambio de la primer pieza, X_2 el tiempo (en horas) de recambio de la segunda pieza y así sucesivamente, de modo que el tiempo (en horas) Y_n en que se realiza el n-ésimo recambio es $Y_n = X_1 + X_2 + \ldots + X_n$, con $X_1, X_2, \ldots, X_n$ i.i.d.

   (i) (**2 puntos**) Hallar la distribución de Y_n . ¿Cuáles son su valor esperado y su varianza?

   (ii) (**2 puntos**) Calcular aproximadamente la probabilidad de que se tengan que cambiar 100 piezas antes de las 240 horas de producción si el tiempo de vida de las piezas es $T \sim Exp(\ln 2)$.
   (**Sugerencia:** usar el teorema central del límite.)

(c) A partir de los tiempos (en horas) $X_1, X_2, \ldots, X_n$ en los que se realiza el recambio de las piezas se desea estimar el parámetro λ del tiempo de vida de dichas piezas.

   (i) (**1 punto**) Calcular λ en función de μ siendo $\mu = \mathbf{E}(X_1)$.

   (ii) (**1 punto**) ¿Cómo estimaría μ a partir de las observaciones de los tiempos de recambio de las piezas?

   (iii) (**3 puntos**) Construir un estimador consistente para λ en función de las observaciones $X_1, X_2, \ldots, X_n$ de los tiempos de recambio de las piezas.

UNIVERSIDAD DE LA REPÚBLICA - Facultad de Ingeniería

Primer Parcial de PROBABILIDAD Y ESTADÍSTICA 2002 - 18 de mayo del 2002

SOLUCIONES

Problema 1

(a) Consideremos los sucesos

$$A = \text{"el bollillero sorteado es el } A\text{"}$$
$$B = \text{"el bollillero sorteado es el } B\text{"}$$
$$R = \text{"la bollila extraída es roja"}$$

Luego

$$\begin{aligned}\mathbf{P}(R) &= \mathbf{P}(R \cap A) + \mathbf{P}(R \cap B) = \mathbf{P}(R/A)\,\mathbf{P}(A) + \mathbf{P}(R/B)\,\mathbf{P}(B) \\ &= \left(\frac{10}{12}\right)\left(\frac{1}{5}\right) + \left(\frac{4}{10}\right)\left(\frac{4}{5}\right) = \boxed{\frac{73}{150} = 0.486\,67}\end{aligned}$$

(b)

$$\mathbf{P}(A/R) = \frac{\mathbf{P}(A \cap R)}{\mathbf{P}(R)} = \frac{\mathbf{P}(R/A)\,\mathbf{P}(A)}{\mathbf{P}(R)} = \frac{\left(\frac{10}{12}\right)\left(\frac{1}{5}\right)}{\frac{73}{150}} = \boxed{\frac{25}{73} = 0.342\,47}$$

(c) Consideremos el suceso

$$R_k = \text{"hay exactamente } k \text{ bolillas rojas entre las 4 bollilas extraidas"} \qquad \text{para } k = 0, 1, 2, 3, 4$$

Luego

$$\mathbf{P}(R_k) = \mathbf{P}(R_k \cap A) + \mathbf{P}(R_k \cap B) =$$

$$= \underbrace{\mathbf{P}(R_k/A)}_{\substack{\mathsf{C}_k^4 \left(\frac{10}{12}\right)^k \left(1 - \frac{10}{12}\right)^{4-k} \\ \text{(muestreo con reposición)}}} \mathbf{P}(A) + \underbrace{\mathbf{P}(R_k/B)}_{\substack{\frac{\mathsf{C}_k^4 \mathsf{C}_{4-k}^6}{\mathsf{C}_4^{10}} \\ \text{(muestreo sin reposición)}}} \mathbf{P}(B) =$$

$\Rightarrow$

$$\boxed{\mathbf{P}(R_k) = \mathsf{C}_k^4 \left(\frac{10}{12}\right)^k \left(1 - \frac{10}{12}\right)^{4-k} \left(\frac{1}{5}\right) + \frac{\mathsf{C}_k^4 \mathsf{C}_{4-k}^6}{\mathsf{C}_4^{10}} \left(\frac{4}{5}\right)} \qquad \text{para } k = 0, 1, 2, 3, 4$$

Problema 2

(a) La función de distribución de X es

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \forall x \in \mathbb{R}$$

Si $x < -1$

$$F(x) = \int_{-\infty}^{x} \underbrace{f(t)}_{=0} dt = 0$$

Si $-1 \leq x < 1$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \underbrace{\int_{-\infty}^{-1} \underbrace{f(t)}_{=0} dt}_{=0} + \int_{-1}^{x} f(t)dt = \int_{-1}^{x} \frac{3}{4}(1-t^2)dt = \frac{3}{4}\left(x - \frac{1}{3}x^3 + \frac{2}{3}\right) = \frac{1}{4}(2-x)(x+1)^2$$

Si $x \geq 1$

$$F(x) = \int_{-\infty}^{x} f(t)dt = \underbrace{\int_{-\infty}^{-1} \underbrace{f(t)}_{=0} dt}_{=0} + \int_{-1}^{1} f(t)dt + \underbrace{\int_{1}^{+\infty} \underbrace{f(t)}_{=0} dt}_{=0} = \int_{-1}^{1} \frac{3}{4}(1-t^2)dt = 1$$

En resumen

$$\boxed{F(x) = \begin{cases} 0 & \text{si } x < -1 \\ \frac{1}{4}(2-x)(x+1)^2 & \text{si } -1 \leq x < 1 \\ 1 & \text{si } x \geq -1 \end{cases}.}$$

(b)

$$\mathbf{P}(X \leq a) = \mathbf{P}(X > a) = \frac{1}{2} \Leftrightarrow \begin{cases} F(a) = \frac{1}{2} \\ -1 \leq a \leq 1 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{4}(2-a)(a+1)^2 = \frac{1}{2} \\ -1 \leq a \leq 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} a\left(a-\sqrt{3}\right)\left(a+\sqrt{3}\right) = 0 \\ -1 \leq a \leq 1 \end{cases} \Leftrightarrow \boxed{a = 0}$$

(c)

$$\mathbf{E}(X) = \int_{-1}^{1} xf(x)dx = \frac{3}{4}\int_{-1}^{1} x(1-x^2)dx = \boxed{0}$$

$$\mathbf{E}(X^2) = \int_{-1}^{1} x^2 f(x)dx = \frac{3}{4}\int_{-1}^{1} x^2(1-x^2)dx = \boxed{\frac{1}{5}}$$

$$\mathbf{E}(X^4) = \int_{-1}^{1} x^4 f(x)dx = \frac{3}{4}\int_{-1}^{1} x^4(1-x^2)dx = \boxed{\frac{3}{35}}$$

(d) i.

$$\mathbf{E}(U) = \mathbf{E}(X_1 X_2) = \text{ (por la independencia) } \mathbf{E}(X_1)\mathbf{E}(X_2) = \boxed{0}$$

$$\mathbf{E}(V) = \boxed{0} \text{ (idem)}$$

$$\begin{aligned}\mathbf{E}(UV) &= \mathbf{E}\left((X_1X_2)(X_2X_3)\right) = \mathbf{E}\left(X_1 (X_2)^2 X_3\right) = \\ &= \text{(por la independencia) } \mathbf{E}(X_1)\,\mathbf{E}\left((X_2)^2\right)\mathbf{E}(X_3) = \boxed{0}\end{aligned}$$

$$\mathbf{Cov}(U,V) = \mathbf{E}(UV) - \mathbf{E}(U)\mathbf{E}(V) = \boxed{0}$$

ii.

$$\begin{aligned}\mathbf{E}\left(U^2\right) &= \mathbf{E}((X_1X_2)^2) = \mathbf{E}\left((X_1)^2 (X_2)^2\right) = \text{(por la independencia) } \mathbf{E}\left((X_1)^2\right)\mathbf{E}\left((X_2)^2\right) = \\ &= \left(\frac{1}{5}\right)\left(\frac{1}{5}\right) = \boxed{\frac{1}{25}}\end{aligned}$$

$$\mathbf{E}(V^2) = \boxed{\frac{1}{25}} \quad \text{(idem)}$$

$$\begin{aligned}\mathbf{E}((UV)^2) &= \mathbf{E}\left((X_1X_2)^2 (X_2X_3)^2\right) = \mathbf{E}\left((X_1)^2 (X_1)^4 (X_3)^2\right) = \mathbf{E}\left((X_1)^2\right)\mathbf{E}\left((X_2)^4\right)\mathbf{E}\left((X_3)^2\right) = \\ &= \left(\frac{1}{5}\right)\left(\frac{3}{35}\right)\left(\frac{1}{5}\right) = \boxed{\frac{3}{875}}\end{aligned}$$

(e) **No son independientes**, pues si U y V fuesen independientes $\Rightarrow U^2$ y V^2 independientes $\Rightarrow \mathbf{E}((UV)^2) = \mathbf{E}\left(U^2\right)\mathbf{E}(V^2)$. lo cual es falso, pues por la parte anterior

$$\mathbf{E}((UV)^2) = \frac{3}{875} \neq \frac{1}{625} = \left(\frac{1}{25}\right)\left(\frac{1}{25}\right) = \mathbf{E}\left(U^2\right)\mathbf{E}(V^2)$$

Problema 3

(a) i. Como

$$T \sim Exp(\lambda) \Leftrightarrow f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

se tiene que

$$\begin{aligned}\mathbf{P}(X = n) &= \mathbf{P}([T] + 1 = n) = \mathbf{P}([T] = n-1) = \mathbf{P}(n-1 \leq T < n) = \\ &= \int_{n-1}^{n} f_T(t)\,dt = \int_{n-1}^{n} \lambda e^{-\lambda t} dt = \boxed{-e^{-n\lambda} + e^{-(n-1)\lambda}} \qquad \text{con } n \in \mathbb{N}\end{aligned}$$

ii. Para cada $n \in \mathbb{N}$ tenemos que:

$$\begin{aligned}\mathbf{P}(X = n) &= -e^{-n\lambda} + e^{-(n-1)\lambda} = e^{-(n-1)\lambda} - e^{-n\lambda} = e^{-(n-1)\lambda} - e^{-(n-1)\lambda}e^{\lambda} = e^{-(n-1)\lambda}\left(1 - e^{-\lambda}\right) = \\ &= \left(e^{-\lambda}\right)^{n-1}\left(1 - e^{-\lambda}\right) = (1-p)^{n-1}p \qquad \text{donde } p = 1 - e^{-\lambda}\end{aligned}$$

Así $X \sim Geo(p)$ con $p = 1 - e^{-\lambda}$

(b) Sea $Y_n = X_1 + X_2 + \ldots + X_n$ el tiempo en que se realiza el n-ésimo recambio, con $X_1, X_2, \ldots, X_n$ i.i.d con distribución $Geo(p)$.

(i) Como $X_1, X_2, \ldots, X_n$ son i.i.d. con distribución *Geo(p)* sabemos (*ver Práctico 3 Ejercicio 7(d)*) que

$$\boxed{Y \text{ tiene distribución binomial negativa } \operatorname{BN}(n,p)}$$

por lo tanto (*ver Práctico 6 Ejercicio 9*)

$$\mathbf{E}(Y_n) = \frac{n}{p} = \boxed{\frac{n}{1-e^{-\lambda}}}$$

$$\mathbf{Var}(Y_n) = n\frac{1-p}{p^2} = n\frac{1-\left(1-e^{-\lambda}\right)}{\left(1-e^{-\lambda}\right)^2} = \boxed{n\frac{e^{-\lambda}}{\left(1-e^{-\lambda}\right)^2}}$$

(ii) Para $n = 100$ piezas de recambio, queremos calcular

$$\begin{aligned}
\mathbf{P}\left(Y_n \leq 240\right) &= \mathbf{P}\left(X_1 + X_2 + \ldots + X_n \leq 240\right) = \\
&= \mathbf{P}\left(\frac{X_1 + X_2 + \ldots + X_n}{n} \leq \frac{240}{n}\right) = \\
&= \mathbf{P}\left(\sqrt{n}\frac{\frac{X_1+X_2+\ldots+X_n}{n} - \mathbf{E}(X_1)}{\sqrt{\mathbf{Var}(X_1)}} \leq \underbrace{\sqrt{n}\frac{\frac{240}{n} - \mathbf{E}(X_1)}{\sqrt{\mathbf{Var}(X_1)}}}_{=a}\right) \qquad \text{donde } X_1 \sim Geo(p)
\end{aligned}$$

Por el teorema central del límite

$$\sqrt{n}\frac{\frac{X_1+X_2+\ldots+X_n}{n} - \mathbf{E}(X_1)}{\sqrt{\mathbf{Var}(X_1)}} \approx Z \qquad \text{con } Z \sim \mathbf{N}(0,1)$$

Así

$$\mathbf{P}\left(Y_n \leq 240\right) \approx \mathbf{P}\left(Z \leq a\right)$$

Como $\lambda = \ln 2$ tenemos que

$$\mathbf{E}(X_1) = \frac{1}{p} = \frac{1}{1-e^{-\ln 2}} = 2 \qquad \text{y} \qquad \mathbf{Var}(X_1) = \frac{1-p}{p^2} = \frac{e^{-\ln 2}}{\left(1-e^{-\ln 2}\right)^2} = 2$$

luego

$$a = \sqrt{n}\frac{\frac{240}{n} - \mathbf{E}(X_1)}{\sqrt{\mathbf{Var}(X_1)}} = \sqrt{100}\frac{\frac{240}{100} - 2}{\sqrt{2}} = 2\sqrt{2}$$

entonces

$$\mathbf{P}\left(Y_n \leq 240\right) \approx \mathbf{P}\left(Z \leq 2\sqrt{2}\right) = \boxed{0.99766107}$$

(también se puede hacer directamente usando los resultados del *ejercicio 5 del Práctico 7*)

(c) (i)

$$\mu = \mathbf{E}(X_1) \Leftrightarrow \mu = \frac{1}{p} \Leftrightarrow \mu = \frac{1}{1-e^{-\lambda}}$$

de donde

$$\boxed{\lambda = -\ln\frac{\mu-1}{\mu} \Leftrightarrow \lambda = -\ln\frac{\mathbf{E}(X_1)-1}{\mathbf{E}(X_1)}}$$

(ii) Por la ley fuerte de los grandes números

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \overset{c.s}{\to} \mu$$

por lo tanto

$$\boxed{\widehat{\mu} = \overline{X}_n}$$

es un estimador consistente para μ

(iii) Por lo anterior, tenemos que

$$-\ln \frac{\overline{X}_n - 1}{\overline{X}_n} \overset{c.s}{\to} -\ln \frac{\mu - 1}{\mu} = \lambda$$

así

$$\boxed{\widehat{\lambda} = -\ln \frac{\overline{X}_n - 1}{\overline{X}_n}}$$

es un estimador consistente para λ (función de las observaciones $X_1, X_2, \ldots, X_n$)