Departamento de Matemática Aplicada a las T.I.C.

ASIGNATURA: ESTADÍSTICA Y PROCESOS ESTOCÁSTICOS

Examen Final (Otoño 2019/20) **FECHA:** 17 de Enero de 2020 **Duración:** 3 horas

SOLUCIONES

Ejercicio 1 *(1,2 puntos) Una urna contiene 6 bolas blancas y 12 bolas rojas. Se lanza un dado equilibrado y se eliminan de la urna tantas bolas rojas como los puntos obtenidos en el resultado de la tirada. A continuación, se extrae una bola al azar.*

1. *Halla la probabilidad de que la bola extraída sea blanca.*

2. *Si la bola extraída resulta ser blanca, halla la probabilidad de que el resultado de la tirada haya sido 2 o 4.*

Solución: *Sean los sucesos:*

$D_i =$ *El resultado de la tirada del dado son i puntos,* $i = 1, \ldots, 6$; $B =$ *La bola extraída es blanca*

1. *Para hallar la probabilidad de que la bola extraída sea blanca utilizamos el teorema de la probabilidad total, utilizando el sistema completo de sucesos* $\{D_1, \ldots, D_6\}$*, ya que la unión de todos ellos es un suceso seguro ("obtener entre 1 y 6 puntos al lanzar el dado"), no puede darse conjuntamente dos de esos resultados diferentes (*$D_i \cap D_j = \emptyset$*, si* $i \neq j$*) y cada suceso tiene probabilidad no nula (*$P(D_i) = \frac{1}{3}$*, si* $i = 1, \ldots, 6$*). Usando también que las probabilidades, en caso de sucesos equiprobables, son los casos favorables entre casos posibles*

$$P(B) = P(D_1)P(B/D_1) + \ldots + P(D_6)P(B/D_6) = \sum_{i=1}^{6} P(D_i)P(B/D_i) = \frac{1}{6}\sum_{i=1}^{6} P(B/D_i) =$$

$$= \frac{1}{6}\left(\frac{6}{17} + \frac{6}{16} + \frac{6}{15} + \frac{6}{14} + \frac{6}{13} + \frac{6}{12}\right) = \frac{51939}{123760} \approx 0.4197$$

2. *Sabiendo que la bola extraída ha sido blanca, la probabilidad de que el resultado haya sido* $D_2 \cup D_4$ *es* $P\Big((D_2 \cup D_4)/B\Big)$*, y utilizando la definición de la probabilidad condicionada, propiedades de sucesos y probabilidades y que* $D_2 \cap D_4 = \emptyset$

$$P\Big((D_2 \cup D_4)/B\Big) = \frac{P[(D_2 \cup D_4) \cap B]}{P(B)} = \frac{P[(D_2 \cap B) \cup (D_4 \cap B)]}{P(B)} =$$

$$= \frac{P(D_2 \cap B) + P(D_4 \cap B) - P(D_2 \cap B \cap D_4 \cup B)}{P(B)} = \frac{P(D_2)P(B/D_2) + P(D_4)P(B/D_4) - P(\emptyset)}{P(B)} =$$

$$= \frac{\frac{1}{6}\frac{6}{16} + \frac{1}{6}\frac{6}{14} - 0}{P(B)} = \frac{\frac{1}{16} + \frac{1}{14}}{\frac{51939}{123760}} = \frac{5525}{17313} \approx 0.3191$$

Ejercicio 2 *(1,3 puntos) Sea* T *la variable aleatoria "tiempo hasta que falla una componente" (medido en horas).* T *sigue una distribución exponencial de media* $1/2$ *hora.*

1. *Halla la probabilidad de que una componente siga funcionando al cabo de una hora.*

2. *Se instalan 20 de esas componentes (se supone que las componentes fallan con independencia unas de otras). ¿Cuál es la probabilidad de que al menos 19 de esas componentes sigan funcionando al cabo de una hora?*

Solución:

1. *T tiene distribución Exp(λ) donde el parámetro $\lambda = 2$ ya que $E[T] = 1/2 = 1/\lambda$. Entonces la probabilidad p de que una componente funcione al cabo de una hora es*

$$p = P(T > 1) = \int_1^\infty 2e^{-2t}dt = [-e^{-2t}]_{t=\infty} - [-e^{-2t}]_{t=1} = e^{-2}$$

2. *La variable aleatoria $N \equiv$ número de componentes que funcionan al cabo de una hora, sigue una distribución binomial $n = 20$ y $p = e^{-2}$. Entonces, la probabilidad de que al menos 19 de esas componentes funcionen al cabo de 1 hora es*

$$P(N = 19) + P(N = 20) = 20(e^{-2})^{19}(1 - e^{-2})^1 + (e^{-2})^{20} = 20e^{-18} - 20e^{-20} + e^{-40} \approx 0$$

Ejercicio 3 *(1,5 puntos) Un dispositivo electrónico ha sido diseñado para que, una vez activado y de forma aleatoria, encienda y apague las luces de una casa exactamente una vez a la hora. Sea X la variable aleatoria que recoge el instante de la hora en que el dispositivo enciende las luces e Y la variable aleatoria del momento de la hora en que las apaga.*

La función de densidad conjunta de la variable aleatoria bidimensional $(X,\ Y)$ es

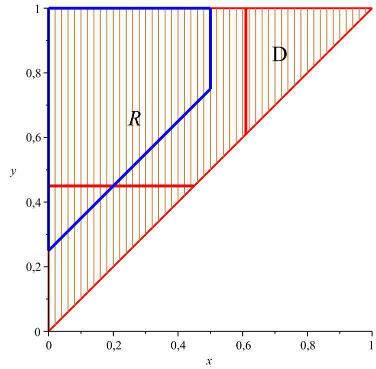
$$f(x,\ y) = \begin{cases} 8\,x\,y & si\ 0 < x < y < 1 \\ 0 & en\ el\ resto \end{cases}$$

1. *Halla la función de densidad marginal de la variable aleatoria Y.*

2. *Calcula $E[X]$.*

3. *Calcula la probabilidad de que las luces se enciendan en la primera media hora y permanezcan encendidas al menos durante un cuarto de hora.*

Solución: *En el siguiente gráfico representamos la zona rayada, el triángulo D de vértices $(0, 0)$, $(1, 1)$ y $(0, 1)$, que es la región donde la función de densidad no es nula y las dos líneas rojas representan donde se debe integrar para hallar las densidades marginales. La zona R encerrada por rectas en azul del siguiente gráfico se utilizará más adelante.*



1. *La función de densidad marginal de Y fuera del intervalo $(0,\ 1)$ es cero y, si $y \in (0,\ 1)$, es*

$$f_Y(y) = \int_0^y 8xy\ dx = 4yx^2\Big|_0^y = 4\ y^3$$

$$\text{Así, } f_Y(y) = \begin{cases} 4\ y^3 & si\ 0 < y < 1 \\ 0 & en\ el\ resto \end{cases}$$

2. $E[X] = \displaystyle\int_0^1 \int_x^1 x \cdot (8xy)\ dydx$

$$E[X] = 8\int_0^1 x^2 \int_x^1 y\ dydx = 8\int_0^1 x^2 \left[\frac{y^2}{2}\right]_x^1 dx = 4\int_0^1 x^2(1-x^2)\ dx =$$
$$= 4\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_0^1 = 4\frac{5-3}{15} = \approx 0.53333$$

Otra forma sería hallar la densidad marginal de X, que es cero fuera del intervalo $(0,\ 1)$ y, si $x \in (0,\ 1)$, es $f_X(x) = \displaystyle\int_x^1 8xy\ dy = 4xy^2\Big|_x^1 = 4x(1-x^2)$. Entonces,

$$E[X] = 4\int_0^1 x \cdot x(1-x^2)\ dx = 4\left(\frac{x^3}{3} - \frac{x^5}{5}\right)\Big|_0^1 = 4\frac{5-3}{15} = \approx 0.53333$$

3. *Las luces se enciendan antes de transcurrir media hora desde la activación del dispositivo, por lo que $X < 1/2$, y estén encendidas al menos un cuarto de hora, $Y > X + 1/4$, es la zona R encerrada por rectas en azul del gráfico anterior. Con lo que,*

$$P(X < 1/2,\ Y > X + 1/4) = 8\int_0^{1/2} \int_{x+0{,}25}^1 x \cdot y\ dydx = 8\int_0^{1/2} x \left[\frac{y^2}{2}\right]_{x+0{,}25}^1 dx =$$
$$= 4\int_0^{1/2} x\left(\frac{15}{16} - x^2 - \frac{x}{2}\right)\ dx = 4\left(\frac{15}{16}\frac{x^2}{2} - \frac{x^4}{4} - \frac{1}{2}\frac{x^3}{3}\right)\Big|_0^{1/2} = \frac{31}{96} \approx 0.3229$$

Ejercicio 4 *(1 punto) El tiempo T de espera en minutos en la parada de un cierto autobús se supone que sigue una distribución uniforme en el intervalo $(0, 12)$. Un usuario apunta el tiempo que espera el autobús cada día de un mes (de 30 días). Utilizando el teorema central del límite, calcula aproximadamente la probabilidad de que la media de los 30 tiempos de espera sea mayor de 7 minutos.*

Solución: *Sean $T_1, \ldots, T_{30}$ las variables aleatorias que modelan el tiempo de espera de cada uno de los 30 días del mes. Dichas variables son independientes entre sí y siguen una distribución $U(0, 12)$. El problema nos pide hallar la probabilidad de que $\dfrac{T_1 + \ldots + T_{30}}{30} > 7$.*

Como $E[T_i] = 6$ y $\mathrm{Var}(T_i) = \frac{12^2}{12} = 12$, el teorema central del límite dice que

$$T_1 + \ldots + T_{30} \sim N(\mu = 30 \cdot 6, \sigma^2 = 30 \cdot 12)$$

aproximadamente. Equivalentemente, $\dfrac{T_1 + \ldots + T_{30}}{30} \sim N\left(\mu = 6, \sigma^2 = \dfrac{12}{30} = 0.4\right)$. Por tanto:

$$P\left(\frac{T_1 + \ldots + T_{30}}{30} > 7\right) = P\left(Z > \frac{7-6}{\sqrt{0.4}} = 1.58\right) = 1 - F_Z(1.58) = 0.057$$

Ejercicio 5 *(1 punto)*

1. *Si el porcentaje de pelirrojos en una población fuera del 40 %, ¿cuál sería la probabilidad de que de 10 personas de esta población solo una fuera pelirroja?*

2. *Para contrastar la hipótesis nula de que el porcentaje de pelirrojos en cierta población es del 40 %, frente a la alternativa de que este porcentaje es distinto del 40 %, se estudia una muestra aleatoria de 10 personas de esta población. Si resulta que solo una de ellas es pelirroja, ¿cuál es el p-valor del test?*

Solución:

1. *Si el porcentaje de pelirrojos fuera del 40 %, la variable aleatoria* $N \equiv$ *número de pelirrojos entre 10, tendría distribución binomial con* $n = 10$ *y* $p = 0,4$*, y la probabilidad de que de 10 personas de esta población solo una sea pelirroja sería*

$$P(N = 1) = 10 \times 0.6^9 \times 0.4 \approx 0.04$$

2. *Si la hipótesis nula es cierta, la variable aleatoria* $N \equiv$ *número de pelirrojos entre 10, tiene distribución binomial con* $n = 10$ *y* $p = 0,4$*. Entonces el p-valor es*

$$\textit{p-valor} = 2 \times P(N \leq 1) = 2 \times \left[P(N = 0) + P(N = 1)\right] = 2 \times \left[0.6^{10} + 10 \times 0.6^9 \times 0.4\right] \approx 0,0927$$

Ejercicio 6 *(1,5 puntos) El diámetro en milímetros de ciertos tornillos sigue una distribución normal con desviación típica* $0,04$ *mm. Se mide el diámetro de 25 tornillos. La media de estas 25 medidas es* $5,01$ *mm.*

1. *Obtén un intervalo de confianza para el diámetro medio de los tornillos con un 92 % de nivel de confianza.*

2. *Obtén el p-valor del contraste cuya hipótesis nula es que el diámetro medio de los tornillos es* 5 *mm siendo la alternativa que el diámetro medio no es 5 mm. ¿Se acepta la hipótesis nula con un nivel de significación* $\alpha = 0,1$ $(10\,\%)$*? ¿y con un nivel de significación* $\alpha = 0,3$ $(30\,\%)$*?*

Solución:

1. *Teniendo en cuenta que el cuantil* $0,96$ *de una N(0,1) es aproximadamente* $1,75$ *se obtiene que el intervalo es*

$$\left(\overline{x} - q_{0,96}\frac{\sigma}{\sqrt{n}}, \overline{x} + q_{0,96}\frac{\sigma}{\sqrt{n}}\right) = \left(5.01 - 1.75\frac{0,04}{\sqrt{25}}\ ,\ 5.01 + 1.75\frac{0,04}{\sqrt{25}}\right) \approx (\,4.996\,,\,5.024\,)$$

2. *Si la hipótesis nula es cierta, el estadístico de contraste*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 5}{0.04/\sqrt{25}}$$

tiene distribución N(0,1). Para la muestra disponible

$$\frac{\overline{x} - 5}{0.04/\sqrt{25}} = \frac{5,01 - 5}{0.04/\sqrt{25}} = 1.25$$

Por tanto el p-valor es

$$\textit{p-valor} = 2 \times P(Z > 1.25) = 2 \times (1 - 0.894) = 0.212$$

En consecuencia, se acepta la hipótesis con un nivel de significación del 0.1 y se rechaza con un nivel de significación del 0.3

Ejercicio 7 *(1,3 puntos) Sea* $\{X(t)\ /t \in \mathbb{R}\}$ *un proceso gaussiano, o normal, estacionario con* $\mu_X(t) = 0$ *y* $R_X(t, t+\tau) = \dfrac{1}{1+\tau^2}$*, la media y la función de autocorrelación del proceso, respectivamente.*

1. *Halla la distribución conjunta de* $X(0)$ *y* $X(0) - 2X(1)$*. ¿Son independientes las variables aleatorias* $X(0)$ *y* $X(0) - 2X(1)$*?*

2. *Halla* $P(X(0) - X(1) > 0)$*.*

Solución:

1. *Al ser el proceso* $X(t)$ *gaussiano la distribución conjunta de las variables aleatorias* $\begin{pmatrix} X(0) \\ X(1) \end{pmatrix}$ *es normal bidimensional.*

   Como $E[X(t)] = \mu_X(t) = \mu_X$*, por ser el proceso estacionario, la media es constante e igual a cero ya que* $\mu_X = 0$*, así el vector de medias de* $\begin{pmatrix} X(0) \\ X(1) \end{pmatrix}$ *es* $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$*.*

   $V[X(t)] = R_X(0) - \mu_X^2 = \dfrac{1}{1+0^2} - 0^2 = 1$ *y* $Cov[X(0),\ X(1)]] = R_X(|1-0|) - \mu_X^2 = \dfrac{1}{1+1^2} = \dfrac{1}{2}$.

   Así que la matriz de varianzas y covarianzas de $\begin{pmatrix} X(0) \\ X(1) \end{pmatrix}$ *es* $\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$*.*

   La variable aleatoria $\begin{pmatrix} X(0) \\ X(0) - 2X(1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} X(0) \\ X(1) \end{pmatrix}$ *es una combinación lineal de variables aleatorias normales, por lo que es una variable aleatoria normal bidimensional. Por ser el operador esperanza lineal y la media del proceso constante e igual cero,*

   $$E[X(0)] = 0 \quad y \quad E[X(0) - 2X(1)] = E[X(0)] - 2E[X(1)] = 0 - 0 + 0 = 0$$

   La matriz de varianzas y covarianzas es

   $$\Sigma = \begin{pmatrix} 1 & 0 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

   La distribución conjunta de las variables aleatorias $X(0)$ *y* $X(0) - 2X(1)$ *es*

   $$\begin{pmatrix} X(0) \\ X(0) - 2X(1) \end{pmatrix} \sim \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \right)$$

   Por tener las v.a. $X(0)$ *y* $X(0) - 2X(1)$ *distribuciones normales y su covarianza ser cero, esas variables aleatorias son independientes.*

2. $\mu_X(0-1) = E[X(0) - X(1)] = E[X(0)] - E[X(1)] = 0 - 0 = 0$ *porque* $\mu_X = 0$*. La variable aleatoria* $X(0) - X(1)$ *es normal, por ser combinación lineal de v.a. normales, por tanto es simétrica respecto de su media, así que*

   $$P(X(0) - X(1) > 0) = P(X(0) - X(1) > \mu_X(0-1)) = \frac{1}{2}$$

Ejercicio 8 *(1,2 puntos) Sea $N(t)$ el número de avisos que llegan en el intervalo $[0,t)$ horas a cierta oficina. Suponemos que $N(t)$ es un proceso de Poisson de tasa $\lambda = 1$ aviso por hora.*

1. *Calcula $P(N(2) \geq 2,\ N(4) \leq 3)$.*
2. *Halla la media del proceso $N(t)$.*
3. *Si $R_N(t_1,t_2)$ es la función de autocorrelación del proceso, calcula $R_N(1,3)$.*

Solución:

1. *Como siempre tiene que ocurrir que $N(2) \leq N(4)$, se verifica*

   $$P(N(2) \geq 2,\ N(4) \leq 3) = P(N(2) = 2,\ N(4) = 2) + P(N(2) = 2,\ N(4) = 3) + P(N(2) = 3,\ N(4) = 3) =$$

   $$= P(N(2) = 2,\ N(4) - N(2) = 0) + P(N(2) = 2,\ N(4) - N(2) = 1) + P(N(2) = 3,\ N(4) - N(2) = 0)$$

   $N(2)$ es el número de clientes que llegan en $[0,2)$ y $N(4) - N(2)$ es el número de clientes que llegan en $[2,4)$, puesto que $[0,2) \cap [2,4) = \emptyset$ y el proceso es de Poisson, entonces las v.a. $N(2)$ y $N(4) - N(2)$ son independientes,

   $$\begin{aligned} P(N(2) \geq 2,\ N(4) \leq 3) =& P(N(2) = 2)P(N(4) - N(2) = 0) + \\ &+ P(N(2) = 2)P(N(4) - N(2) = 1) + P(N(2) = 3)P(N(4) - N(2) = 0) \end{aligned}$$

   Por ser $N(2) \backsim \mathcal{P}(1 \cdot 2)$ y $N(4) - N(2) \backsim \mathcal{P}(1 \cdot (4-2))$

   $$P(N(2) \geq 2,\ N(4) \leq 3) = e^{-2}\frac{2^2}{2!}e^{-2}\frac{2^0}{0!} + e^{-2}\frac{2^2}{2!}e^{-2}\frac{2^1}{1!} + e^{-2}\frac{2^3}{3!}e^{-2}\frac{2^0}{0!} = e^{-4}(2+4+4/3) = e^{-4}\frac{22}{3} \backsimeq 0,1343$$

2. *Puesto que $N(t)$ es una v.a. de Poisson de parámetro $\lambda \cdot t$, su media es el parámetro, en este caso t. Así, $\mu_N(t) = E[N(t)] = t$. A la variable aleatoria $N(3)$ la sumamos y restamos $N(1)$, obteniendo*

   $$E[N(1)N(3)] = E[N(1)(N(1) + N(3) - N(1))] = E[N(1)N(1)] \cdot E[N(1)(N(3) - N(1))]$$

   Usando el mismo razonamiento que en el apartado anterior $N(1)$ es el número de clientes que llegan en $[0,1)$ y $N(3) - N(1)$ es el número de clientes que llegan en $[1,3)$, puesto que los intervalos $[0,1)$ y $[1,3)$ son disjuntos y el proceso es de Poisson las v.a. $N(1)$ y $N(3) - N(1)$ son independientes, por lo que

   $$E[N(1)(N(3) - N(1))] = E[N(1)]E[N(3) - N(1)] = (\lambda \cdot 1) \cdot (\lambda \cdot (3-1)) = 1 \cdot (3-1) = 2$$

   ya que al ser proceso de Poisson la distribución de $N(3) - N(1)$ es de Poisson de parámetro $\lambda \cdot (3-1)$. Concluimos:

   $$\begin{aligned} R_N(1,3) =& E[N(1)N(3)] \\ =& V[N(1)] + E[N(1)]^2 + E[N(1)]E[N(3) - N(1)] = \\ =& 1 + 1^2 + 1 \cdot (3-1) = 4 \end{aligned}$$

   Por tanto, $\quad R_N(1,3) = 4.$