ASIGNATURA: ESTADÍSTICA Y PROCESOS ESTOCÁSTICOS
Examen Final (Primavera 2020/21) **FECHA:** 3 de Junio de 2021 **Duración:** 3 horas

SOLUCIONES

Ejercicio 1 *(1 punto) Para una clave, se necesita escoger un número de 6 dígitos (a elegir entre* $0, 1, 2, \ldots, 9$*), con la condición de que no puede haber ningún dígito repetido.*

a) ¿Cuántas claves pueden formarse en total?

b) Si elegimos una clave al azar entre todas las posibles, ¿cuál es la probabilidad de que contenga los dígitos 2 *y* 3*?*

c) Si elegimos una clave al azar entre todas las posibles, ¿cuál es la probabilidad de que contenga la subcadena de dígitos '89'?

Solución.

a) Se trata de variaciones (sin repetición) de diez elementos tomados de seis en seis. En total, hay

$$V_{10,6} = \frac{10!}{4!} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151200$$

claves posibles.

b) Utilizaremos la Regla de Laplace. Por tanto, debemos contar primero el número de claves que contienen los dígitos 2 *y* 3*. Elegimos primero en qué dos posiciones de la clave colocaremos el* 2 *y el* 3*, y en qué orden los colocamos: en total hay* $2 \cdot C_{6,2}$ *posibilidades. Para las cuatro posiciones restantes, debemos elegir qué dígitos (entre los ocho disponibles) colocamos: en total hay* $V_{8,4}$ *posibilidades. Por lo tanto, la probabilidad que se pide es*

$$\frac{\textit{casos favorables}}{\textit{casos totales}} = \frac{2C_{6,2} \cdot V_{8,4}}{V_{10,6}} = \frac{(2 \cdot 15) \cdot (8 \cdot 7 \cdot 6 \cdot 5)}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5} = \frac{1}{3}.$$

c) Utilizamos de nuevo la Regla de Laplace. Ahora, debemos tener en cuenta en qué posiciones podemos colocar la subcadena de dígitos '89': basta observar que el dígito '8' puede ir en cualquier posición excepto en la última, es decir, tenemos 5 *posibilidades. Una vez fijada esa posición, en las cuatro posiciones restantes debemos elegir qué dígitos (entre los ocho disponibles) colocamos: en total hay* $V_{8,4}$ *posibilidades. En consecuencia, la probabilidad pedida es*

$$\frac{\textit{casos favorables}}{\textit{casos totales}} = \frac{5 \cdot V_{8,4}}{V_{10,6}} = \frac{5 \cdot (8 \cdot 7 \cdot 6 \cdot 5)}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5} = \frac{1}{18}.$$

Ejercicio 2 *(1 punto) En una caja hay cinco chips, entre los cuales puede haber alguno defectuoso. Si* D *denota el número de chips defectuosos en la caja, se sabe que*

$$P(D = i) = \frac{5 - i}{15}, \qquad i = 0, 1, 2, 3, 4, 5.$$

a) Se extrae de la caja un chip al azar. Calcula la probabilidad de que el chip extraído sea defectuoso.

b) Se extrae un chip de la caja y se comprueba que es defectuoso. Calcula la probabilidad de que inicialmente hubiera exactamente 2 chips defectuosos en la caja.

Solución.

a) Sea E el suceso "el chip extraído al azar es defectuoso". Para calcular la probabilidad $P(E)$, aplicamos el Teorema de la Probabilidad Total (los sucesos $(D = i)$ son incompatibles dos a dos, y su unión es el suceso seguro):

$$P(E) = \sum_{i=0}^{5} P(E/D = i)P(D = i).$$

Ahora bien, claramente $P(E/D = 0) = 0$ y $P(D = 5) = 0$, con lo cual basta sumar ente $i = 1$ e $i = 4$. Para $i = 1, 2, 3, 4$, sabemos que $P(D = i) = (5 - i)/15$, y $P(E/D = i)$ se calcula usando la Regla de Laplace. Tenemos entonces

$$\begin{aligned} P(E) &= P(E/D = 1)P(D = 1) + P(E/D = 2)P(D = 2) + P(E/D = 3)P(D = 3) + P(E/D = 4)P(D = 4) \\ &= \frac{1}{5} \cdot \frac{4}{15} + \frac{2}{5} \cdot \frac{3}{15} + \frac{3}{5} \cdot \frac{2}{15} + \frac{4}{5} \cdot \frac{1}{15} = \frac{20}{75} = \frac{4}{15}. \end{aligned}$$

b) Se pide calcular la probabilidad condicionada $P(D = 2/E)$. Aplicamos en este caso la Fórmula de Bayes:

$$P(D = 2/E) = \frac{P(E/D = 2)P(D = 2)}{P(E)} = \frac{\frac{2}{5} \cdot \frac{3}{15}}{4/15} = \frac{3}{10}.$$

Ejercicio 3 *(1,5 puntos) El tiempo de vida T de ciertos componentes electrónicos (en años) sigue una distribución exponencial* $\mathrm{Exp}(1/2)$ *de parámetro* $1/2$*. Calcula:*

a) La probabilidad de que uno de esos componentes electrónicos, elegido al azar, dure más de tres años.

b) La probabilidad de que al menos 2 componentes entre 10 elegidos al azar duren más de tres años.

c) La probabilidad de que, inspeccionando componentes secuencialmente, el primero que tenga una vida útil de más de tres años sea el tercero.

Solución.

a) Se pide calcular $P(T > 3) = 1 - P(T \leq 3) = 1 - F_T(3)$, donde F_T denota la función de distribución de T. Puesto que la función de densidad de la exponencial de parámetro $1/2$ es

$$f(t) = \begin{cases} \frac{1}{2}e^{-t/2} & si\ t \geq 0, \\ 0 & si\ t < 0, \end{cases}$$

se tiene

$$P(T > 3) = 1 - \frac{1}{2}\int_0^3 e^{-t/2}dt = 1 - \frac{1}{2}\left[(-2)e^{-t/2}\right]_{t=0}^{t=3} = 1 + \left(e^{-3/2} - 1\right) = e^{-3/2} \approx 0.223.$$

b) *Elegimos 10 componentes electrónicos al azar, y llamamos X a la variable aleatoria que cuenta el número de componentes entre esos 10 cuyo tiempo de vida es de más de tres años. Entonces, X sigue una distribución binomial de parámetros $N = 10$, $p = P(T > 3) \approx 0.223$. Se pide calcular*

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{10}{0} \cdot (1-p)^{10} - \binom{10}{1} p \cdot (1-p)^9$$
$$\approx 1 - 0.777^{10} - 10 \cdot 0.223 \cdot 0.777^9 \approx 0.690.$$

c) *Supongamos que se inspeccionan componentes electrónicos secuencialmente. Consideramos ahora la variable aleatoria Y que cuenta el número de inspecciones hasta encontrar un componente electrónico con tiempo de vida mayor a tres años. Entonces, Y sigue una distribución geométrica de parámetro $p = P(T > 3) \approx 0.223$. Por lo tanto,*

$$P(Y = 3) = p \cdot (1-p)^2 \approx 0.223 \cdot 0.777^2 \approx 0.135.$$

Ejercicio 4 *(1,2 puntos) Sea (X, Y) el vector aleatorio con función de densidad conjunta*

$$f(x,y) = \begin{cases} 2x & \textit{si } 0 < x < 1, 0 < y < 1, \\ 0 & \textit{en otro caso.} \end{cases}$$

a) *Calcula las funciones de densidad marginales f_X, f_Y. ¿Son X e Y variables independientes? Justifica tu respuesta.*

b) *Calcula $E[X]$, $E[Y]$ y $\operatorname{Cov}(X,Y)$.*

c) *Calcula $P(Y < X^2)$.*

Solución.

a) *Tratamos primero la variable aleatoria X. Si $x \notin (0,1)$, claramente $f_X(x) = 0$. En cambio, para $x \in (0,1)$ se tiene*

$$f_X(x) = \int_0^1 2x dy = 2x.$$

Por lo tanto, X tiene función de densidad marginal

$$f_X(x) = \begin{cases} 2x & \textit{si } 0 < x < 1, \\ 0 & \textit{en otro caso.} \end{cases}$$

Para Y, procedemos de la misma manera. Cuando $y \notin (0,1)$, se tiene $f_Y(y) = 0$, mientras que para $y \in (0,1)$ comprobamos que

$$f_Y(y) = \int_0^1 2x dx = 1.$$

Es decir, la función de densidad marginal para Y es

$$f_Y(y) = \begin{cases} 1 & \textit{si } 0 < y < 1, \\ 0 & \textit{en otro caso.} \end{cases}$$

Se comprueba fácilmente del enunciado y del apartado anterior que $f_X(x) \cdot f_Y(y) = f(x,y)$. Por lo tanto, X e Y son independientes.

b) *Por un lado, tenemos*

$$E[X] = \int_0^1 x f_X(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}, \quad E[Y] = \int_0^1 y f_Y(y) dy = \int_0^1 y dy = \frac{1}{2}.$$

Por otro lado, como X e Y son independientes, se sigue que $\operatorname{Cov}(X,Y) = 0$. *De forma alternativa, como*

$$E[XY] = \int_0^1 \int_0^1 2x^2 y dx dy = \frac{1}{3},$$

podemos calcular directamente la covarianza entre X e Y:

$$\operatorname{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{2}{3} \cdot \frac{1}{2} = 0.$$

c) *Para calcular $P(Y < X^2)$ debemos integrar la función de densidad conjunta $f(x,y)$ en la subregión del cuadrado $(0,1) \times (0,1) \subset \mathbb{R}^2$ por debajo del arco de parábola determinado por la ecuación $y = x^2$. Es decir,*

$$P(Y < X^2) = \int_0^1 \left(\int_0^{x^2} 2x dy \right) dx = \int_0^1 2x^3 dx = 2 \cdot \frac{1}{4} = \frac{1}{2}.$$

Ejercicio 5 *(1,3 puntos) Una ONG tiene un sistema de donaciones online en su página web que permite seleccionar tres cantidades para aportar a la ONG: 10, 20 y 50 euros. Basándonos en el histórico de donaciones, la probabilidad de que un donante al azar seleccione cada una de esas cantidades es 0.2, 0.7 y 0.1, respectivamente. Para una determinada causa, la ONG lanza una campaña con el objetivo de recaudar 6.000 euros mediante su sistema online. Calcula, de manera aproximada utilizando el Teorema del Límite Central, la probabilidad de que esa recaudación se supere con 300 donaciones, suponiendo independencia entre las cantidades aportadas en cada donación.*

Solución. *Sea X la variable que mide el importe de una donación al azar en el sistema online de la ONG. Se trata de una variable aleatoria discreta cuya función de probabilidad viene dada por*

$$P(X = 10) = 0.2, \quad P(X = 20) = 0.7, \quad P(X = 50) = 0.1.$$

La esperanza de X es por lo tanto

$$\mu = E[X] = 10 \cdot 0.2 + 20 \cdot 0.7 + 50 \cdot 0.1 = 21.$$

Por otro lado, $E[X^2] = 10^2 \cdot 0.2 + 20^2 \cdot 0.7 + 50^2 \cdot 0.1 = 550$, *con lo cual la varianza de X es*

$$\operatorname{Var}(X) = E[X^2] - E[X]^2 = 550 - 21^2 = 550 - 441 = 109,$$

y la desviación típica de X es por lo tanto

$$\sigma = +\sqrt{\operatorname{Var}(X)} = +\sqrt{109} \approx 10.44.$$

Sean ahora $X_1, X_2, \ldots, X_{300}$ las variables aleatorias que miden el importe de 300 donaciones en el sistema online de la ONG. Todas ellas siguen la misma distribución que X, y son independientes entre ellas. Aplicando el Teorema del Límite Central, se deduce que su suma

$$S := X_1 + X_2 + \cdots + X_{300} \overset{\text{aprox}}{\sim} N(6300, 180.83)$$

sigue (aproximadamente) una distribución normal de media $6300 = 300 \cdot 21$ *y desviación típica* $180.83 \approx \sqrt{300} \cdot 10.44$.

Debemos calcular $P(S > 6000)$. *Tipificando, tenemos*

$$P(S > 6000) = P\left(\frac{S - 6300}{180.83} > \frac{-300}{180.83}\right) \approx P(Z > -1.66),$$

donde $Z \sim N(0,1)$. *Utilizando la simetría de la normal estándar y las tablas para su función de distribución, deducimos que*

$$P(S > 6000) = P(Z > -1.66) = P(Z \leq 1.66) \approx 0.952.$$

Por lo tanto, la probabilidad de que con 300 donaciones se alcance la recaudación deseada es aproximadamente 0.952.

Ejercicio 6 *(1 punto) La distribución de probabilidad de una variable aleatoria* X *depende del parámetro* θ. *El estadístico* $T(X_1, X_2, \ldots, X_n; \theta) = \dfrac{\theta\left(\sum_{i=1}^{n} X_i + \sigma_X^2\right)}{n}$, *sigue una distribución* χ^2 *con* $2n$ *grados de libertad, siendo* σ_X^2 *la varianza de la muestra. Halla un intervalo de confianza al 98 % para el parámetro* θ, *sabiendo que para una muestra de tamaño* $n = 10$, *se ha comprobado que* $\sum_{i=1}^{10} x_i = 62$ *y la desviación típica es* 5.

Solución.

$$T(X_1, X_2, \ldots, X_n; \theta) = \frac{\theta\left(\sum_{i=1}^{n} X_i + \sigma_X^2\right)}{n} \sim \chi^2_{2n}$$

En nuestro caso el estadístico T *sigue una distribución* χ^2_{20} *y su valor en la muestra es* $\dfrac{\theta(62+25)}{10}$. *Entonces, para hallar un intervalo de confianza al 98 % consideramos los cuantiles* $q_{0{,}01}$ *y* $q_{0{,}99}$ *de una v.a.* χ^2_{20} *y que, consultando la tabla, son 8,2604 y 37,5663, respectivamente.*

$$P(q_{0{,}01} < \frac{\theta(62+25)}{10} < q_{0{,}99}) = 0,98$$

$$P\left(\frac{10 \cdot q_{0{,}01}}{87} < \theta < \frac{10 \cdot q_{0.99}}{87}\right) = 0,98$$

Por tanto, un intervalo de confianza al 98 % para θ *es*

$$I.C._{0.98}(\theta) = \left(\frac{10 \cdot q_{0{,}01}}{87}, \frac{10 \cdot q_{0{,}99}}{87}\right) = \left(\frac{82,604}{87}, \frac{375,663}{87}\right) = (0,94947,\ 4,31796)$$

Ejercicio 7 *(1 punto) Para estudiar las resistencias, que tienen distribución normal, de una empresa se propone un contraste de hipótesis donde la hipótesis nula es que la media de las resistencias es de 70 ohmios* ($H_0 : \mu = 70$) *frente a la hipótesis alternativa de que la resistencia media es diferente de ese valor. Se eligen 25 resistencias y la cuasi desviación típica resulta ser 3 ohmios,* ($s_1 = 3$). *Con un nivel de significación de* $\alpha = 0,1$ *¿para qué valores de la media de esas resistencias se aceptará la hipótesis nula?*

Solución.

El contraste de hipótesis es

$H_0 : \mu = 70$

$H_1 : \mu \neq 70$

El estadístico del contraste para la media de una población con varianza desconocida es $T = \dfrac{\overline{X} - 70}{s_1/\sqrt{n}}$

Si la hipótesis nula es cierta T *sigue una distribución* t_{n-1}, *nuestro caso* t_{24}.

Los cuantiles $q_{\alpha/2} = q_{0,05}$ *y* $q_{1-\alpha/2} = q_{0,95}$ *de una* t_{24} *son* $\pm 1,7109$, *por tanto la región de aceptación del contraste es el intervalo* $(-1,7109,\ 1,7109)$.

Si $-1,7109 < \dfrac{\overline{X} - 70}{s_1/\sqrt{n}} < 1,7109$, *entonces* $-1,7109 < \dfrac{\overline{X} - 70}{3/5} < 1,7109$. *Y despejando* $\overline{X}$ *obtenemos* $70 - 1,7109\dfrac{3}{5} = 68,97346 < \overline{X} < 70 + 1,7109\dfrac{3}{5} = 71,02654$.

Se aceptará la hipótesis nula, con un nivel de significación $\alpha = 0,1$ *si la media de las 25 resistencias se encuentra en el intervalo* $(68,97346,\ 71,02654)$.

Ejercicio 8 *(1 punto) Un proceso estocástico* $X(t)$ *solo puede tener tres realizaciones* $x_1(t) = 3,\ x_2(t) = 3\cos t,\ x_3(t) = 4\operatorname{sen} t$, *ocurriendo todas con la misma probabilidad.*

a) *Halla la distribución de* $X(0)$, *su media y varianza.*

b) *Halla la distribución de* $X(\pi)$, *su media y varianza.*

c) *¿Es el proceso estacionario en algún sentido? Razona la respuesta.*

Solución.

a) $x_1(0) = 3,\ x_2(0) = 3,\ x_3(0) = 0$. $X(0)$ *toma el valor 0 con probabilidad* $1/3$ *y el valor 3 con probabilidad* $2/3$.
   $E[X(0)] = 0\dfrac{1}{3} + 3\dfrac{2}{3} = 2$.
   $E[X(0)^2] = 3^2\dfrac{2}{3} = 6$. *Entonces,* $V[X(0)] = 6 - 2^2 = 2$

b) $x_1(\pi) = 3,\ x_2(\pi) = -3,\ x_3(\pi) = 0$
   $X(\pi)$ *toma los valores 3, -3 y 0, cada uno con probabilidad* $1/3$.
   $E[X(\pi)] = 3\dfrac{1}{3} + (-3)\dfrac{1}{3} + 0\dfrac{1}{3} = 0$.
   $E[X(\pi)^2] = \dfrac{3^2 + (-3)^2 + 0^2}{3} = 6$. $V[X(\pi)] = 6 - 0 = 6$

c) *El proceso no es E.S.A. porque la media del proceso no es constante. Si no es E.S.A. tampoco es E.S.E.*

Ejercicio 9 *(1 punto) Los clientes llegan a un establecimiento siguiendo un Proceso de Poisson* $N(t)$ *con parámetro* $\lambda = 4$ *clientes por hora. El establecimiento está abierto desde las 10 de la mañana hasta las 8 de la noche.*

a) *¿Cuál es la probabilidad de que exactamente haya llegado un cliente antes de las 11 y un total de seis hasta las 12:30?*

b) *Sabiendo que han llegado 8 clientes hasta las 13 horas ¿cuál es la probabilidad de que lleguen 12 clientes antes de las 15 horas?*

Solución.

a) *El número de clientes entre hasta las 11 es* $N(1)$ *y el número de clientes hasta las 12:30 es* $N(2,5)$. *Entonces,* $P(N(1)=1, N(2,5)=6) = P(N(1)=1, N(2,5)-N(1)=5) =$

$$= P(N(1)=1) \cdot P(N(2,5)-N(1)=5) =$$

$$= e^{-4}\frac{4^1}{1} \cdot e^{-4(2,5-1)}\frac{(4(2,5-1))^5}{5!} \approx 0.0733 \cdot 0.1606 \approx 0.01177$$

b) $P\big(N(5)=12/N(3)=8\big) = \dfrac{P(N(3)=8, N(5)=12)}{P(N(3)=8)} =$

$$= \frac{P[N(3)=8, N(5)-N(3)=4]}{P(N(3)=8)} =$$

$$= \frac{P(N(3)=8)P(N(5)-N(3)=4)}{P(N(3)=8)} = P(N(5)-N(3)=4) =$$

$$= P(N(2)=4) = e^{-8}\,\frac{8^4}{4!} \approx 0.0573$$