Estadística Descriptiva e Introducción a la Probabilidad

Relación de Ejercicios 5

Autores, por orden alfabético:
Shao Jie Hu Chen
Adrián Jaén Fuentes
Aarón Jerónimo Fernández
Noura Lachhab Bouhmadi
Laura Lázaro Soraluce

Doble Grado en Ingeniería Informática y Matemáticas

Lista de Problemas

Problema 1

Apartado 1

Para determinar el valor de k usaremos la siguiente característica de la función masa de probabilidad:

$$\sum_{i=1}^{20} P(X=i) = 1$$

Teniendo esto en cuenta:

$$\sum_{i=1}^{20} P(X=i) = \sum_{i=1}^{20} ki = k\sum_{i=1}^{20} i = k \cdot \frac{20(20+1)}{2} = 210k \Rightarrow k = \frac{1}{210}$$

Luego, la función de distribución de la variable aleatoria X sería la siguiente:

$$F(x) = \begin{cases} 0 & si \quad x < 1 \\ \sum_{i=1}^{x} \dfrac{i}{210} & si \quad 1 \leq x \leq 19 \\ 1 & si \quad x \geq 20 \end{cases}$$

Ahora calculamos las probabilidades que nos piden con la función de distribución calculada:

$$P(X=4) = F(4) - F(4^-) = F(4) - F(3) = \frac{10}{210} - \frac{6}{210} = \frac{4}{210}$$

$$P(X<4) = P(X \leq 4) - p(X=4) = F(4^-) = f(3) = \frac{6}{210}$$

$$P(3 \leq X \leq 10) = P(X \leq 10) - P(X < 3) = F(10) - F(3^-) = F(10) - F(2) =$$
$$= \frac{55}{210} - \frac{3}{210} = \frac{52}{210}$$

$$P(3 < X \leq 10) = P(X \leq 10) - P(X \leq 3) = F(10) - F(3) = \frac{55}{210} - \frac{6}{210} = \frac{49}{210}$$

$$P(3 < X < 10) = P(X < 10) - P(X \leq 3) = F(10^-) - F(3) =$$
$$= F(9) - F(3) = \frac{45}{210} - \frac{6}{210} = \frac{39}{210}$$

Apartado 2

La ganancia viene dada de la siguiente forma:

$$Ganancia = \begin{cases} 20 & si \quad x < 4 \\ 24 & si \quad x = 4 \\ -1 & si \quad x > 4 \end{cases}$$

Calculemos las probabilidades de cada situación:

$$P(x < 4) = \frac{6}{210}$$
$$P(X = 4) = \frac{4}{210}$$
$$P(X > 4) = 1 - P(X \leq 4) = 1 - F(4) = 1 - \frac{10}{210} = \frac{20}{21}$$

Para ver si este juego le es favorable al jugador calculemos la esperanza matemática (media) de una nueva variable Y que tome estos valores. Si la esperanza matemática es positiva, el juego será favorable, si es negativa, no.Definimos la variable Y con sus valores y probabilidades:

$$P(Y = 20) = P(x < 4) = \frac{6}{210} P(Y = 24) = P(X = 4) = \frac{4}{210} P(Y = -1) = P(X > 4) = \frac{20}{21}$$

Ahora calculamos su esperanza matemática:

$$E[Y] = \sum_i y_i \cdot P(Y = y_i) = \frac{120}{210} + \frac{96}{210} - \frac{20}{21} = \frac{8}{105} \approx 0{,}076$$

Al ser $E[Y] > 0$ vemos que el juego es favorable, aunque la ganancia será muy pequeña.

Problema 2

Apartado 1

Tenemos que la variable X puede tomar 3 valores: 0, 1, 2 (No sale ninguna bola blanca, sale una o salen dos). Por tanto:

$$P[X = 0] = \frac{2}{10} \cdot \frac{1}{9} = \frac{1}{45} \quad P[X = 1] = \frac{8}{10} \cdot \frac{2}{9} = \frac{16}{45} \quad P[X = 2] = \frac{8}{10} \cdot \frac{7}{9} = \frac{28}{45}$$

Ahora podemos definir la función de distribución:

$$F(X) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{45} & \text{si } 0 <= x < 1 \\ \frac{17}{45} & \text{si } 1 <= x < 2 \\ 1 & \text{si } x >= 1 \end{cases}$$

Apartado 2

- Media: Es la esperanza matemática, es decir, el centro de gravedad de la distribución de probabilidad y se calcula como

$$E[X] = \sum_i x_i \cdot P[X = x_i] = \frac{16}{45} + 2 \cdot \frac{28}{45} = \frac{8}{5}$$

- Mediana: Es el valor de la variable que deja por debajo la el 50 % de la probabilidad. En nuestro caso, mirando a la función de distribución tenemos que

$$\mathrm{Me} = 2$$

- Moda: Es el valor de x_i con mayor probabilidad, por lo que se calculará como el máximo de los $P[X = x_i]$. Por tanto

$$\mathrm{Mo} = 2$$

Apartado 3

Para el recorrido intercuartílico calculemos los percentiles 75 y 25:

$$P_{75} = \min\{\ x_i\ /\ P[X < x_i] = F[x_i] > 0{,}75\} = x_2 = 2$$

$$P_{25} = \min\{\ x_i\ /\ P[X < x_i] = F[x_i] > 0{,}25\} = x_1 = 1$$

Por tanto el recorrido intercuartílico será $Q_3 - Q_1 = 2 - 1 = 1$, que indica la longitud del intervalo en el que se encuentra el 50 % de los datos.

Problema 3

Sabemos que se trata de una variable aleatoria discreta.
Para probar que la función masa de probabilidad está bien definida, comprobamos que cumple:

Apartado 1

1. $P(X = x_i) \geq 0\ \forall i \in N \Rightarrow P[X = x] = \frac{1}{2^x}$ que siempre va a ser positiva pues nunca se anula; y 2^x está siempre en el intervalo $(0, +\infty)$, por lo que la fracción nunca va a ser negativa.
2. $\sum_{i=1}^{\infty} P[X = x_i] = \sum_{i=1}^{\infty} \frac{1}{2^x} = \frac{1}{2} + \frac{1}{2} + \frac{1}{8} + ...$ Esta es una serie geométrica de razón $\frac{1}{2} \Rightarrow \frac{1}{2} < 1 \Rightarrow$ la serie converge. Su suma viene dada por $\sum_i^n 2^{-n} = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$

Apartado 2

$P(4 \leq x \leq 10) = \sum_{i=4}^{10} p_i = \sum_{i=4}^{10} 2^{-x_i} = \frac{1}{2^4} + \frac{1}{2^5} + \frac{1}{2^6} + \frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \frac{1}{2^{10}} = \frac{127}{1024} \simeq 0'124$

Apartado 3

$Q_1 : P(X \leq x_i) = \frac{25}{100} = 0'25 \Rightarrow \sum_{i=1}^{x_i} 2^{-x_i} = 0'25 < 0'5 = 2^{-1} \Rightarrow Q_1 = 1$
$Q_2 : P(X \leq x_i) = \frac{50}{100} = 0'5 \Rightarrow 2^{-1} = 0'5 \Rightarrow Q_2 = [1, 2)$
$Q_3 : P(X \leq x_i) = \frac{75}{100=0'75} \Rightarrow \sum_{i=1}^{2} 2^{-x_i} = 0'5 + 0'25 = 0'75 \Rightarrow Q_3 = [2, 3)$
M_o: valor máximo. Cuanto mayor es x_i, menor es $\frac{1}{2_i^x}$. Por lo tanto, el valor máximo va a ser el primero, cuando $x_i = 1 \Rightarrow Mo = 1$

Apartado 4

Función generatriz de momentos: $M_x(t) = \sum_{x=1}^{\infty} e^{tx} * P[X = x_i] = \sum_{x=1}^{\infty} e^{tx} * \frac{1}{2^x} = \sum_{x=1}^{\infty} (\frac{e^t}{2})^x = \frac{\frac{e^t}{2}}{1-\frac{e^t}{2}}$.
Esto es una serie geométrica que converge si $\frac{e^t}{2} < 1 \Rightarrow e^t < 2 \Rightarrow t < log(2)$. Por lo tanto, converge en un entorno de cero.
Esperanza: $E[X] = M_x^{'}(t) = \frac{(\frac{e^t}{2})^{'}(1-\frac{e^t}{2})-(\frac{e^t}{2})(1-\frac{e^t}{2})^{'}}{(1-\frac{e^t}{2})^2} = \frac{\frac{e^t}{2}}{1-e^t+\frac{e^t}{4}}$. Sustituimos t=0: $\frac{\frac{1}{2}}{\frac{1}{4}} = 2 = E[X]$
Desviación típica: $E[X^2] = M_x^{''}(t) = \frac{\frac{e^t}{2}-\frac{e^{3t}}{8}}{1-2e^t+\frac{3e^{2t}}{2}-\frac{e^{3t}}{2}+\frac{e^{4t}}{16}}$. Sustituimos t=0: $\frac{\frac{6}{16}}{\frac{1}{16}} = 6$
Varianza: $Var[X] = E[X^2] - E[X]^2 = 6 - 2^2 = 2 \Rightarrow \sigma_x = \sqrt{2} \simeq 1'414$

Problema 4

Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} k_1(x+1) & 0 \leq x \leq 4 \\ \\ k_2 x^2 & 4 < x \leq 6 \end{cases}$$

Sabiendo que $P(0 \leq X \leq 4) = 2/3$, determinar $k_1,\ k_2$, y deducir su función de distribución.

Nótese que la variable aleatoria es continua y toma valores entre 0 y 6. En el desarrollo de este ejercicio, se presupondrá que la función de masa dada es correcta.

En primer lugar, calculemos la función de distribución de la variable aleatoria. Aplicamos la siguiente expresión: $F(x) = \int_{-\infty}^{x} f(t)dt$.

Si $0 \leq x \leq 4$, tenemos que:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{x} k_1(t+1)dt = k_1(\frac{x^2}{2} + x)$$

Si $4 < x \leq 6$, tenemos que:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} 0dt + \int_{0}^{4} k_1(t+1)dt + \int_{4}^{x} k_2 t^2 dt \\ &= k_1(\frac{4^2}{2} + 4) + k_2(\frac{x^3}{3} - \frac{4^3}{3}) \end{aligned}$$

Sabemos que $P(0 \leq X \leq 4) = F(4) - F(0^-)$. Como la variable es continua, tenemos que $F(0^-) = F(0)$. Deducimos entonces que:

$$P(0 \leq X \leq 4) = F(4) - F(0^-) = F(4) - F(0) = k_1(\frac{4^2}{2} + 4) = 2/3 \Rightarrow k_1 = \frac{1}{18}$$

Por otra parte, para que se verifique que se trate, en efecto, de una función de distribución, se ha de verificar que:

$$F(+\infty) = k_2(\frac{6^3}{3} - \frac{4^3}{3}) + \frac{2}{3} = 1 \Rightarrow k_2 = \frac{1}{152}$$

Por tanto, la función de distribución viene dado por la siguiente expresión:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{18}(\frac{x^2}{2} + x) & 0 \leq x \leq 4 \\ \frac{x^3}{456} + \frac{10}{19} & 4 < x \leq 6 \\ 1 & x > 6 \end{cases}$$

Problema 5

Apartado 1

Para determinar el valor de k usaremos la siguiente característica de la función de densidad:

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

Teniendo esto en cuenta:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{1}^{10} \frac{k}{x^2}\,dx = \frac{k}{-x}\Big|_1^{10} = \frac{-k}{10} + k = \frac{9k}{10} = 1 \Rightarrow k = \frac{10}{9}$$

Y finalmente tenemos que $k = \dfrac{10}{9}$. Para calcular la función de distribución debemos calcular la integral de la función de densidad, como ya la habíamos calculado antes, ahora solo tenemos que definirla. La definimos:

$$F(x) = \begin{cases} 0 & si & x < 1 \\ \frac{10/9}{-x}\Big|_1^x & si & 1 \leq x \leq 10 \\ 1 & si & x > 10 \end{cases}$$

Apartado 2

Para calcular esta probabilidad usaremos la función de distribución:

$$P(2 \leq x \leq 5) = F(5) - F(2) = \frac{10/9}{-x}\Big|_1^5 - \frac{10/9}{-x}\Big|_1^2 = \frac{8}{9} - \frac{5}{9} = \frac{3}{9} = \frac{1}{3}$$

Apartado 3

En este apartado nos piden que calculemos el percentil 50 y el percentil 95, para calcularlos debemos resolver la ecuación $P(X \leq x_i) = \dfrac{r}{100}$ donde r es el percentil que queremos calcular. Resolvemos estas ecuaciones:

-P_{50}:

$$P(X \leq x_i) = 0{,}5; \quad F(x_i) = 0{,}5; \quad \frac{10/9}{-x}\Big|_1^{x_i} = 0{,}5; \quad \frac{10/9}{-x_i} + \frac{10}{9} = 0{,}5;$$
$$10 - 10x_i = -4{,}5x_i; \quad x_i = \frac{20}{11} = 1{,}88$$

-P_{95}:

$$P(X \leq x_i) = 0{,}95; \quad F(x_i) = 0{,}95; \quad \frac{10/9}{-x}\Big|_1^{x_i} = 0{,}95; \quad \frac{10/9}{-x_i} + \frac{10}{9} = 0{,}95;$$
$$10 - 10x_i = -8{,}55x_i; \quad x_i = \frac{200}{29} \approx 6{,}897$$

Problema 6

Apartado 1

$$P(1'5 < x \leq 2) = \int_{1'5}^{2} f(x)\,dx = \int_{1'5}^{2} \frac{2x-1}{10}\,dx = \frac{1}{10}[\int 2x\,dx - \int dx] = \frac{2}{10} - 0'075 = 0'125$$
$$P(2'5 < x \leq 3'5) = 0$$
$$P(4'5 < x \leq 5'5) = \int_{4'5}^{5'5} 0'4\,dx = 2'2 - 1'8 = 0'4$$
$$P(1'2 < x \leq 5'2) = \int_{1'2}^{2} \frac{2x-1}{10}\,dx + \int_{4}^{5'2} 0'4\,dx = 0'2 - 0'024 + 2'08 - 1'6 = 0'656$$

Apartado 2

$$E[X] = \int_{-\infty}^{\infty} e^{xt} * f(x)\,dx = \int_{1}^{2} \frac{2x-1}{10} * e^{xt}\,dx + \int_{4}^{6} 0'4\,dx = \frac{67}{60}$$
$$M_k = \int_{-\infty}^{\infty} x^k * f(x)\,dx = \int_{1}^{2} \frac{2x-1}{10} * x^k\,dx + \int_{4}^{6} 0'4 * x^k\,dx$$

Apartado 3

$$M_x(t) = E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt} * f(x)\,dx = \int_{1}^{2} \frac{2x-1}{10} * e^{xt}\,dx + \int_{4}^{6} 0'4 * e^{xt}\,dx =$$
$$\frac{e^t((3t-2)e^t - t + 2)}{10t^2} + \frac{2(e^{2t}-1)e^{4t}}{5t}$$

Problema 7

Calcular las funciones masa de probabilidad de las variables $Y = X + 2$ y $Z = X^2$, siendo X una variable aleatoria con distribución:

$$P(X=-2) = \frac{1}{5}, \quad P(X=-1) = \frac{1}{10}, \quad P(X=0) = \frac{1}{5}, \quad P(X=1) = \frac{2}{5}, \quad P(X=2) = \frac{1}{10}.$$

¿Cómo afecta el cambio de X a Y en el coeficiente de variación?

En primer lugar, notemos que la variable aleatoria X es discreta. Basta aplicar el teorema de cambio de variable:

$$P(y_i) = \begin{cases} 1/5 & y_i = 0 \\ 1/10 & y_i = 1 \\ 1/5 & y_i = 2 \\ 2/5 & y_i = 3 \\ 1/10 & y_i = 4 \end{cases}$$

$$P(z_i) = \begin{cases} 1/5 & z_i = 0 \\ 1/2 & z_i = 1 \\ 3/10 & z_i = 4 \end{cases}$$

Notemos que:

$$\begin{aligned} EY &= \sum_i P(Y = y_i) y_i = \sum_i P(Y = y_i)(x_i - a) = \sum_i P(X = x_i)(x_i - a) \\ &= \sum_i P(X = x_i) x_i - a \sum_i P(X = x_i) = \sum_i P(X = x_i) x_i - a \\ &= EX - a \end{aligned}$$

Por tanto, se verifica que:

$$\begin{aligned} E[(Y - EY)^2] &= \sum_i P(Y = y_i)(y_i - EY)^2 = \sum_i P(X = x_i)(x_i - a - EX + a)^2 \\ &= \sum_i P(X = x_i)(x_i - EX)^2 = E[(X - EX)^2] \end{aligned}$$

Deducimos que el coeficiente de variación no ha de variar.

Por otra parte, sabemos que $C.V.(X) = \frac{E[(X-EX)^2]}{EX}$ y $C.V.(Y) = \frac{E[(Y-EY)^2]}{EY}$. Dividiendo ambas expresiones, tenemos que $C.V.(X) = \frac{EY}{EX} C.V.(Y)$. Operamos:

	$p(x_i)$	$p(x_i)x_i$	$(x_i - EX)^2$	$p(x_i)(x_i - EX)^2$
-2	0,2	-0,4	4,41	0,882
-1	0,1	-0,1	1,21	0,121
0	0,2	0	0,01	0,002
1	0,4	0,4	0,81	0,324
2	0,1	0,2	3,61	0,361
		EX = 0,1		$E[(X - EX)^2]$=1,69

	$p(y_i)$	$p(y_i)y_i$	$(y_i - EY)^2$	$p(y_i)(y_i - EY)^2$
0	0,2	0	4,41	0,882
1	0,1	0,1	1,21	0,121
2	0,2	0,4	0,01	0,002
3	0,4	1,2	0,81	0,324
4	0,1	0,4	3,61	0,361
		$EY = 2,1$		$E[(Y - EY)^2] = 1{,}69$

Deducimos que $C.V.(X) = 21 C.V.(Y)$.

Problema 8

Si X es una variable aleatoria con función de densidad

$$f(x) = \frac{e^{-|x|}}{2}, \quad -\infty < x < \infty,$$

hallar su función de distribución y las probabilidad de cada uno de los siguientes sucesos:

a) $\{|X| \leq 2\}$.

b) $\{|X| \leq 2 \quad \text{ó} \quad X \geq 0\}$.

c) $\{|X| \leq 2 \quad \text{y} \quad X \leq -1\}$.

d) $\{X^3 - X^2 - X - 2 \leq 0\}$.

e) $\{X \text{ es irracional}\}$.

En el desarrollo de este ejercicio se presupondrá que la función de masa de probabilidad proporcionada es correcta y está bien definida (nótese que X es una variable aleatoria continua). Calculamos a partir de ella la función de distribución.

Si $x < 0$, tenemos que:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{e^t}{2} dt = \frac{e^x}{2}$$

Si $0 \leq x$, tenemos que:

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} \frac{e^t}{2} dt + \int_{0}^{x} \frac{e^{-t}}{2} dt = \frac{1}{2} + \left[\frac{-e^-t}{2} \right]_0^x = 1 - \frac{1}{2e^x}$$

De esta forma, tenemos que la función de distribución viene dado por la siguiente expresión:

$$F(x) = \begin{cases} \frac{e^x}{2} & x < 0 \\ 1 - \frac{1}{2e^x} & x \geq 0 \end{cases}$$

De esta forma, obtenemos las siguientes probabilidades (considerando en todo momento que la variable aleatoria es continua):

1. $P(|X| \leq 2) = P(2) - P(-2^-) = P(2) - P(-2) = 1 - \frac{1}{2e^2} - \frac{1}{2e^2} = 1 - \frac{1}{e^2} \approx 0,8647$
2. $P(|X| \leq 2) \cup X \geq 0) = P(X \geq -2) = 1 - \frac{1}{2e^2} \approx 0,9323$
3. $P(|X| \leq 2 \cap X \leq -1) = P(-2 \leq X \leq -1) = \frac{1}{2e} - \frac{1}{2e^2} \approx 0,1163$
4. $P(X^3 - X^2 - X - 2 \leq 0) = P(X \leq 2)$ (donde para aplicar esta igualdad hemos factorizado mediante el método de Ruffini, dando como resultado que $x^3 - x^2 - x - 2 = (x-2)(x^2+x+1)$, siendo este segundo término siempre positivo y $x^3 - x^2 - x - 2 \leq 0 \iff x - 2 \leq 0 \iff x \leq 0$). Por tanto, $P(X^3 - X^2 - X - 2 \leq 0) = P(X \leq 2) = 1 - \frac{1}{2e^2} \approx 0,9323$.
5. $P(X \in \mathbb{R} \backslash \mathbb{Q}) = 1 - P(X \in \mathbb{Q}) = 1$ (pues el conjunto $\mathbb{Q}$ es numerable, por lo que la suma de todas las probabilidades es 0).

Problema 9

Antes de empezar, tomaremos los datos que nos dan para deducir la desviación típica y la esperanza matemática. Para empezar nos dicen que el coeficiente de variación es 1, lo que implica que:

$$C.V_X = 1 \Rightarrow \frac{\sigma_X}{E[X]} = 1 \Rightarrow \sigma_X = E[X]$$

Además, nos dicen que la variable aleatoria es simétrica con respecto al punto 2. Esto implica que el punto que deja a la misma cantidad de valores a su derecha que a su izquierda es dos, por lo que $E[X] = 2$ y además:

$$E[X] = 2 \Rightarrow \sigma_X = E[X] = 2$$

Ahora, para abordar el ejercicio, usaremos la desigualdad de Chebychev de la siguiente forma:

$$P(|X - E[X]| \geq k\sigma_X) = P(E[X] - k\sigma_X < k < E[X] + k\sigma_X) \geq 1 - \frac{1}{k^2}$$

Apartado 1

Vamos a calcular k para adaptar la desigualdad de Chebychev a esta probabilidad:

$$E[X] + 2k = 12 \Rightarrow k = \frac{12 - 2}{2} = 5$$

Finalmente sustituimos $k = 5$ en la expresión anterior:

$$P(E[X] - k\sigma_X < k < E[X] + k\sigma_X) = P(-8 < x < 12) \geq 1 - \frac{1}{5^2} = 0{,}96$$

Luego, con los datos que tenemos, podemos decir que $P(-8 < x < 12) \geq 0{,}96$.

Apartado 2

Vamos a calcular k para adaptar la desigualdad de Chebychev a esta probabilidad:

$$E[X] + 2k = 10 \Rightarrow k = \frac{10 - 2}{2} = 4$$

Finalmente sustituimos $k = 4$ en la expresión anterior:

$$P(E[X] - k\sigma_X < k < E[X] + k\sigma_X) = P(-6 < x < 10) \geq 1 - \frac{1}{4^2} = 0{,}9375$$

Luego, con los datos que tenemos, podemos decir que $P(-6 < x < 10) \geq 0{,}9375$.