

# Probabilidad y Estadística

$$P(X \leq x) = \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i}$$

$$\mu_X = np$$

$$\sigma_X^2 = np(1-p)$$

Variables Aleatorias Discretas  
y algunas

Distribuciones de Probabilidad

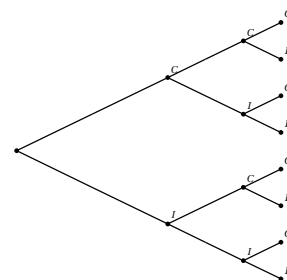
# Índice

<b>1. Variable aleatoria</b>	<b>3</b>
1.1. Ejemplos . . . . .	4
<b>2. Función de probabilidad puntual</b>	<b>6</b>
2.1. Ejemplos . . . . .	7
<b>3. Función de probabilidad acumulada</b>	<b>8</b>
3.1. Ejemplos y observaciones . . . . .	8
3.2. Propuesta . . . . .	9
<b>4. Esperanza matemática de una variable aleatoria discreta</b>	<b>10</b>
4.1. Ejemplos . . . . .	11
4.2. Propuesta . . . . .	12
<b>5. Variancia de una variable aleatoria discreta</b>	<b>13</b>
5.1. Ejemplo . . . . .	13
5.2. Propuesta . . . . .	14
<b>6. Distribución binomial</b>	<b>14</b>
6.1. Esperanza matemática y variancia . . . . .	17
6.1.1. Ejemplo . . . . .	17
6.2. Propuesta . . . . .	17
<b>7. Distribución de Poisson</b>	<b>18</b>
7.1. Esperanza matemática y variancia . . . . .	19
7.2. La distribución de Poisson como caso límite de la distribución Binomial . . . . .	19
7.3. Ejemplos . . . . .	21
<b>8. Distribución geométrica</b>	<b>21</b>
8.1. Esperanza matemática y variancia . . . . .	23
8.2. Período de retorno . . . . .	23
<b>9. Distribución binomial negativa</b>	<b>23</b>
9.1. Ejemplo . . . . .	24
<b>10. Distribución hipergeométrica</b>	<b>25</b>
10.1. Ejemplos . . . . .	25
<b>11. Miscelánea de problemas</b>	<b>26</b>
<b>12. Soluciones</b>	<b>30</b>

# 1. Variable aleatoria

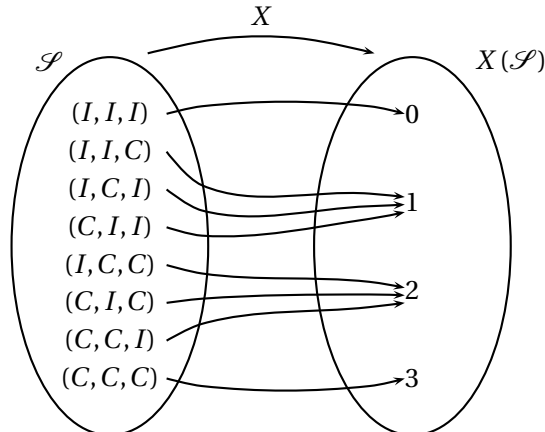
Un estudiante está interesado en conocer la probabilidad de aprobar un examen que consta de 3 preguntas de opción múltiple, contestando cada pregunta al azar. El examen se aprueba con al menos dos respuestas correctas. Por otra parte, cada pregunta consta de cuatro opciones, una sola de las cuales es correcta. El estudiante escribe todas las formas posibles en que puede responder de manera correcta (C) o incorrecta (I) las tres preguntas, usando un diagrama de árbol como el que se muestra.

Luego escribe el conjunto  $\mathcal{S}$  formado por todos los posibles resultados (espacio muestral asociado a la experiencia de observar la secuencia correcto-incorrecto en una serie de tres preguntas)



$$\mathcal{S} = \{(I, I, I), (I, I, C), (I, C, I), (C, I, I), (I, C, C), (C, I, C), (C, C, I), (C, C, C)\}.$$

La terna ordenada  $(C, I, C)$ , por ejemplo, representa haber contestado bien a las preguntas 1 y 3, y respondido incorrectamente la pregunta 2, en cambio la terna  $(C, C, I)$  representa haber contestado bien a las preguntas 1 y 2, y respondido incorrectamente la pregunta 3. En ambos casos, se respondieron bien a 2 preguntas. En total, hay tres formas diferentes de contestar bien a dos preguntas:  $(I, C, C)$ ,  $(C, I, C)$ ,  $(C, C, I)$ . El estudiante encuentra útil asociar a cada elemento de  $\mathcal{S}$ , la cantidad de respuestas correctas que representa.



Surge así la función  $X$  que a cada elemento de  $\mathcal{S}$  le hace corresponder la cantidad de respuestas correctas que representa.

Calcular  $P(\{(I, I, I)\})$ , por ejemplo, equivale a calcular  $P(X = 0)$ . Entonces resulta

- $P(X = 0) = P(\{(I, I, I)\}) = P(I_1 \cap I_2 \cap I_3) = P(I_1)P(I_2)P(I_3) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}$ 
donde  $I_i$  denota el suceso la  $i$ -ésima respuesta es incorrecta. Así mismo hemos considerado que los sucesos  $I_1$ ,  $I_2$  e  $I_3$  son independientes.

$$\begin{aligned} \blacksquare P(X = 1) &= P(\{(I, C, I), (C, I, I), (I, I, C)\}) = P[(I_1 \cap C_2 \cap I_3) \cup (C_1 \cap I_2 \cap I_3) \cup (I_1 \cap I_2 \cap C_3)] = \\ &= P(I_1 \cap C_2 \cap I_3) + P(C_1 \cap I_2 \cap I_3) + P(I_1 \cap I_2 \cap C_3) = P(I_1) \cdot P(C_2) \cdot P(I_3) + P(C_1) \cdot P(I_2) \cdot P(I_3) + \\ &= P(I_1) \cdot P(I_2) \cdot P(C_3) = 3 \cdot \left(\frac{3}{4}\right)^2 \cdot \frac{1}{4} = \frac{27}{64}. \end{aligned}$$

La suma  $P(X = 0) + P(X = 1)$  determina la probabilidad de no aprobar el examen. ¿Por qué?

Le proponemos calcular  $P(X = 2) + P(X = 3)$ . ¿Qué representa esta suma?

El estudiante, para resolver el problema, asoció a cada elemento del espacio muestral un número. Como veremos más adelante, funciones de este tipo, que se conocen como variables aleatorias, facilitarán la resolución de problemas de probabilidad.

Una variable aleatoria es una función que asigna a cada elemento del espacio muestral un número real.

La función  $X$ , definida por el estudiante, es una variable aleatoria. A la imagen de  $\mathcal{S}$  por  $X$ , que simbolizamos  $\mathcal{R}_X$ , la llamamos recorrido de la variable aleatoria  $X$ , y está formado por todos los valores que asume la variable aleatoria  $X$ . En el ejemplo,  $\mathcal{R}_X = \{0, 1, 2, 3\}$ . El conjunto  $\mathcal{R}_X$  es el espacio muestral asociado a la experiencia de observar la cantidad de respuestas correctas sobre un total de tres preguntas. El conjunto  $\mathcal{R}_X$  es finito y por lo tanto, decimos que la variable aleatoria  $X$  es discreta.

Una variable aleatoria es discreta si su recorrido es un conjunto finito o infinito numerable (susceptible de ser contado). Si el recorrido es un intervalo real, la variable aleatoria se dice continua.

## 1.1. Ejemplos

1. Una fábrica produce lámparas de bajo consumo que para su distribución son envasadas en cajas de 25 unidades cada una. El encargado del control de calidad de la producción, selecciona una caja al azar y registra la cantidad de lámparas que, al probarlas, no encienden. Sea

$X_1$ : ‘cantidad de lámparas que no encienden en una caja de 25 lámparas’.

En este caso

$$\mathcal{R}_{X_1} = \{0, 1, 2, \dots, 25\},$$

y por lo tanto la variable aleatoria  $X_1$  es discreta.

2. En una cola de un supermercado se registra la cantidad de artículos que compra un cliente elegido al azar. Sea

$X_2$ : ‘cantidad de artículos que compra un cliente elegido al azar’.

En este caso

$$\mathcal{R}_{X_2} = \mathbb{N} = \{1, 2, \dots\},$$

y por lo tanto la variable aleatoria  $X_2$  es discreta.

3. Se registra, en un día elegido al azar, el número de servicios demandados en un cuerpo de bomberos. Sea

$X_3$ : ‘cantidad de llamados registrados en un día elegido al azar’.

En este caso

$$\mathcal{R}_{X_3} = \mathbb{N}_0 = \{0, 1, 2, \dots\},$$

y por lo tanto la variable aleatoria  $X_3$  es discreta.

4. Se registra la cantidad de artículos que produce una máquina hasta que se descompone. Sea

$X_4$ : ‘cantidad de artículos que produce una máquina elegida al azar hasta que se descompone’.

En este caso

$$\mathcal{R}_{X_4} = \mathbb{N}_0 = \{0, 1, 2, \dots\},$$

y por lo tanto la variable aleatoria  $X_4$  es discreta.

5. Se observa el tiempo que transcurre hasta que una componente de una máquina falla. Sea

$X_5$ : ‘tiempo que transcurre hasta que una componente elegida al azar falla’.

En este caso

$$\mathcal{R}_{X_5} = \mathbb{R}_0^+ = [0, +\infty),$$

y por lo tanto la variable aleatoria  $X_5$  es continua.

6. Se fracciona cemento en bolsas de 50 kg. Sea

$X_6$ : ‘peso de una bolsa de cemento elegida al azar’.

En este caso

$$\mathcal{R}_{X_6} = \mathbb{R}_0^+ = [0, +\infty),$$

y por lo tanto la variable aleatoria  $X_6$  es continua.

7. Se registra el consumo de combustible de un colectivo que realiza el recorrido Rosario-Santa Fe. Sea

$X_7$ : ‘consumo de combustible, en litros, para uno de esos viajes elegido al azar’.

En este caso

$$\mathcal{R}_{X_7} = \mathbb{R}_0^+ = [0, +\infty),$$

y por lo tanto la variable aleatoria  $X_7$  es continua.

En ocasiones los alumnos creen que ciertas variables como por ejemplo, el consumo de combustible de un colectivo es discreta. Es cierto que, por limitaciones en el proceso de medición, las observaciones son discretas. Sin embargo desde un punto de vista teórico el consumo de combustible puede asumir cualquier valor real no negativo.

El llenado o vaciado de un tanque de combustible es un ‘proceso continuo’.

8. Ciertas especificaciones requieren que el espesor de las hojas de aluminio para latas estén entre 8 y 11 milésimos de pulgada. Sea

$X_8$ : ‘espesor de una hoja de aluminio para latas elegida al azar’.

En este caso

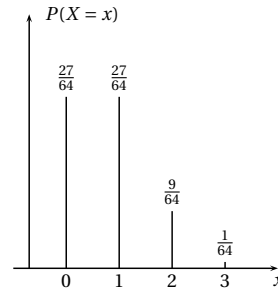
$$\mathcal{R}_{X_8} = \mathbb{R}_0^+ = [0, +\infty),$$

y por lo tanto la variable aleatoria  $X_8$  es continua.

## 2. Función de probabilidad puntual

La siguiente tabla muestra los valores que asume la variable aleatoria  $X$ : ‘cantidad de respuestas correctas en un examen de tres preguntas de opción múltiple’, con sus correspondientes probabilidades; éstas fueron calculadas en la introducción.

$x$	$P(X = x)$
0	$27/64$
1	$27/64$
2	$9/64$
3	$1/64$



Esta correspondencia es una función que llamamos función de probabilidad puntual.

Sea  $X$  una variable aleatoria y  $\mathcal{R}_X$  su recorrido. La función que a cada valor de la variable le hace corresponder la probabilidad de que la variable  $X$  asuma ese valor, se denomina función de probabilidad puntual asociada a la variable aleatoria  $X$ . En símbolos

$$\begin{aligned} p_X : \mathcal{R}_X &\rightarrow [0, 1] \\ x &\mapsto p_X(x) = P(X = x) \end{aligned}$$

El símbolo  $P(X = x)$  se lee ‘probabilidad de que la variable aleatoria  $X$  asuma el valor  $x$ ’.

Si  $p_X$  es una función de probabilidad, entonces cumple que

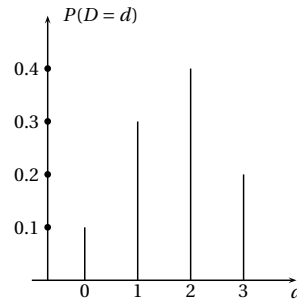
1.  $p_X(x) \geq 0, \forall x \in \mathcal{R}_X$ .
2.  $\sum_{x \in \mathcal{R}_X} p_X(x) = 1$ .

A los pares ordenados  $(x, p_X(x))$  con  $x \in \mathcal{R}_X$ , como a su representación gráfica, la llamamos indistintamente distribución de probabilidad de la variable aleatoria  $X$ .

## 2.1. Ejemplos

1. La siguiente tabla y gráfica corresponden a la distribución de probabilidad de la variable aleatoria  $D$ : 'demanda diaria de un artículo en un negocio'.

$d$	$P(D = d)$
0	0,1
1	0,3
2	0,4
3	0,2



¿Cómo cree que se determinaron las probabilidades de la tabla?

La probabilidad de que se demanden en un día elegido al azar a lo sumo 2 artículos es

$$P(D = 0) + P(D = 1) + P(D = 2) = 0,1 + 0,3 + 0,4 = 0,8.$$

2. La probabilidad de que una variable aleatoria  $Y$  asuma valores en el conjunto  $\{0, 1, 2\}$  está dada por la fórmula  $P(Y = y) = y/3$ , con  $y = 0, 1, 2$ . Esta función es una función de probabilidad puntual porque

a) cada probabilidad es no negativa:

$$p_Y(0) = P(Y = 0) = 0 \geq 0,$$

$$p_Y(1) = P(Y = 1) = \frac{1}{3} \geq 0,$$

$$p_Y(2) = P(Y = 2) = \frac{2}{3} \geq 0;$$

b) la suma de las probabilidades con que la variable aleatoria  $Y$  asume cada valor de su recorrido es 1:

$$\sum_{y \in \mathcal{R}_Y} p_Y(y) = p_Y(0) + p_Y(1) + p_Y(2) = 0 + \frac{1}{3} + \frac{2}{3} = 1.$$

### 3. Función de probabilidad acumulada

En algunas ocasiones resulta necesario calcular la probabilidad de que una variable aleatoria asuma un valor menor o igual a un cierto valor dado. Esto motiva la siguiente definición.

Sea  $X$  una variable aleatoria discreta con recorrido  $\mathcal{R}_X$  y función de probabilidad  $p_X$ . La función  $F_X$  tal que

$$\begin{aligned} F_X: \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto F_X(x) = P(X \leq x), \end{aligned}$$

la denominamos función de distribución acumulada de la variable aleatoria  $X$ .

Es decir,  $F_X(x)$  representa la probabilidad de que la variable aleatoria  $X$  asuma un valor menor o igual que  $x$ , para cada  $x \in \mathcal{R}_X$ . Observe que

$$P(X \leq x) = \sum_{\substack{t \leq x \\ t \in \mathcal{R}_X}} p_X(t).$$

(Suma de las probabilidades de aquellos valores que asume la variable aleatoria y que son menores o igual que  $x$ )

#### 3.1. Ejemplos y observaciones

1. La distribución de probabilidad de la variable aleatoria  $Y$ : ‘cantidad de imperfecciones que se encuentran en un cable de 40 metros que viene en rollos de ancho uniforme’ está dada en la siguiente tabla. A partir de esta distribución calculamos algunos valores de la función de distribución acumulada.

$y$	$P(Y = y)$
0	0,41
1	0,37
2	0,16
3	0,05
4	0,01

- $F_Y(-1) = P(Y \leq -1) = 0$ , ¿por qué?

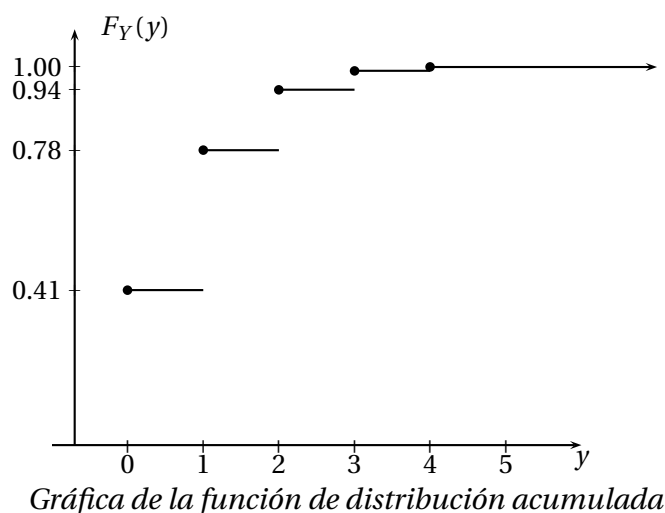
- $F_Y(1) = P(Y \leq 1) = p_Y(0) + p_Y(1) = 0,41 + 0,37 = 0,78$ .

- $F_Y(1,3) = P(Y \leq 1,3) = p_Y(0) + p_Y(1) = 0,41 + 0,37 = 0,78$ .

- $F_Y(4) = P(Y \leq 4) = p_Y(0) + p_Y(1) + p_Y(2) + p_Y(3) + p_Y(4) = 0,41 + 0,37 + 0,16 + 0,05 + 0,01 = 1,00$ .

- $F_Y(y) = 1,00$  cualquiera sea  $y \geq 4$ .





Observe que  $F_Y(3) - F_Y(2) = P(Y = 3)$ . ¿Por qué?

2. Cualquiera sea la variable aleatoria  $U$ , su función de probabilidad acumulada  $F_U$  verifica

- a)  $0 \leq F_U(u) \leq 1$ , para  $u \in \mathbb{R}$ .
- b)  $F_U$  es no decreciente en  $\mathbb{R}$ .
- c)  $F_U$  es discontinua en cada punto  $u$  donde  $P(U = u) > 0$ .

### 3.2. Propuesta

1. Considere la variable aleatoria  $X$ : 'cantidad de cerámicos defectuosos en una caja de 100 unidades', con la siguiente distribución de probabilidad:

$x$	$P(X = x)$
0	$t$
1	0,05
2	0,03
3	0,02
4 o más	0

- a) Calcule el valor de  $t$ .
- b) Explique cómo cree que se obtuvieron las probabilidades de la tabla.
- c) ¿Cuál es la probabilidad de que al elegir una caja al azar, la misma contenga a lo sumo un cerámico defectuoso?
- d) ¿Cuál es la probabilidad de que al elegir una caja al azar, la misma contenga por lo menos un cerámico defectuoso?
- e) Grafique la función de distribución acumulada para la variable aleatoria  $X$ .

2. La función de distribución acumulada para la variable aleatoria  $W$ : ‘cantidad de veces por semana que un proceso de producción necesita recalibrarse’ verifica

$w$	$F_W(w)$
0	0,15
1	0,55
2	0,85
3	0,98
4	1,00

- a) ¿Cuál es la probabilidad de que en una semana elegida al azar, el proceso requiera recalibrarse 1 vez?
- b) ¿Cuál es la probabilidad de que en una semana elegida al azar, el proceso requiera recalibrarse a lo sumo 2 veces?
- c) ¿Cuál es la probabilidad de que en una semana elegida al azar, el proceso requiera recalibrarse por lo menos 5 veces?
- d) Grafique la función de probabilidad puntual asociada a la variable aleatoria  $W$ .

3. Considere la variable  $R$ : ‘número de bacterias por  $\text{cm}^3$  en una muestra de agua contaminada’, cuya función de probabilidad es

$$P(R = r) = p_R(r) = \frac{2^r e^{-2}}{r!}, \quad r = 0, 1, 2, \dots$$

- a) ¿Cuál es la probabilidad de que no haya bacterias en una muestra de  $1 \text{ cm}^3$ ?
- b) ¿Cuál es la probabilidad de que haya a lo sumo 2 bacterias en una muestra de  $1 \text{ cm}^3$ ?
- c) ¿Cuál es la probabilidad de que haya por lo menos 3 bacterias en una muestra de  $1 \text{ cm}^3$ ?
- d) Explique por qué la siguiente afirmación es falsa.

$$F_R(6) - F_R(2) = P(2 \leq R \leq 6).$$

## 4. Esperanza matemática de una variable aleatoria discreta

En su momento hemos expresado la media aritmética (calculada sobre los datos de una muestra) de la siguiente forma

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i n_i = \sum_{i=1}^k x_i \frac{n_i}{n} = \sum_{i=1}^k x_i f_r(x_i), \quad (1)$$

donde

- $x_i$  son los diferentes valores de la variable  $X$  en la muestra,
- $n_i$  es la frecuencia absoluta correspondiente al valor  $x_i$  en la muestra,
- $n = \sum_{i=1}^k n_i$ , es decir el tamaño de la muestra,
- $f_r(x_i)$  es la frecuencia relativa del valor  $x_i$  en la muestra.

Hemos visto también que para  $n \rightarrow +\infty$  la frecuencia relativa correspondiente al valor  $x_i$  tiende a la probabilidad de ocurrencia de ese valor. Es decir

$$f_r(x_i) \xrightarrow{n \rightarrow +\infty} p_X(x_i), \quad (2)$$

donde  $p_X(x_i) = P(X = x_i)$ . De (1) y (2)

$$\bar{x} = \sum_{i=1}^k x_i f_r(x_i) \xrightarrow{n \rightarrow +\infty} \mu_X = \sum_{i=1}^k x_i p_X(x_i).$$

¿Cuál es el significado del valor de  $\mu_X$  generado a partir de este *límite especial*? Se trata de la media aritmética calculada *en la población*. Esta población está formada por todas las observaciones posibles de la variable aleatoria  $X$ . El valor  $\mu_X$  se genera a partir de la media aritmética  $\bar{x}$  para  $n$  tendiendo a infinito.

Sea  $X$  una variable aleatoria discreta con recorrido  $\mathcal{R}_X$  y  $p_X$  su función de probabilidad asociada. Definimos esperanza matemática de la variable aleatoria  $X$  o media poblacional de la variable aleatoria  $X$  y la notamos indistintamente  $E(X)$  o  $\mu_X$  al número

$$E(X) = \mu_X = \sum_{x \in \mathcal{R}_X} x p_X(x).$$

Si  $\mathcal{R}_X$  es infinito numerable, la *suma* de la definición es una serie numérica. En tal caso, la esperanza de la variable existe si la serie es absolutamente convergente.

## 4.1. Ejemplos

1. Sea  $X$ : ‘número que se obtiene al lanzar un dado equilibrado’. Su recorrido es  $\mathcal{R}_X = \{1, 2, 3, 4, 5, 6\}$  y su función de probabilidad puntual es  $p_X(x) = \frac{1}{6}$ , cualquiera sea  $x \in \mathcal{R}_X$ . Entonces

$$\mu_X = \sum_{k=1}^6 x p_X(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5.$$

Si realizamos un *gran* número de tiradas de un dado equilibrado en condiciones uniformes y calculamos la media aritmética de los valores observados, obtendremos, *en general*, una media aritmética próxima a 3,5.

2. La longitud de un mensaje (medido en cantidad de caracteres) que llega a un centro es una variable aleatoria  $Y$  con la siguiente distribución de probabilidad.

$y$	$P(Y = y)$
6	0,15
7	0,20
8	0,40
9	0,25

Entonces

$$\mu_Y = 6 \cdot 0,15 + 7 \cdot 0,20 + 8 \cdot 0,40 + 9 \cdot 0,25 = 7,75 \text{ caracteres.}$$

Si registramos la longitud de un *gran* número de mensajes recibidos en un centro, la media aritmética de esos valores, *en general*, será próxima a 7,75 caracteres.

Observaciones:

1. En ninguno de los ejemplos el valor de la esperanza matemática corresponde a un valor observable, es decir no forma parte del recorrido de la variable aleatoria. ¿Esto es siempre así? Explique.
2. El equivalente mecánico de la esperanza matemática de una variable aleatoria es el centro de gravedad de una distribución de masas.

## 4.2. Propuesta

1. En relación al ejercicio 2 de la página 9 le pedimos que calcule e interprete la esperanza matemática.
2. En la unidad Introducción a la Probabilidad, se tienen los datos correspondientes a 400 tiradas de un dado equilibrado. La media aritmética correspondiente al tercer grupo de 100 observaciones es 3,5000 mientras que la media aritmética correspondiente al grupo de 400 observaciones es 3,5125. La media aritmética calculada sobre una muestra de tamaño 100 se encuentra más próxima a la esperanza matemática que la calculada sobre una muestra de mayor tamaño. Encuentre una explicación de este hecho.
3. Sea  $X$  la variable aleatoria ‘número que se obtiene al arrojar un dado equilibrado de color azul’ e  $Y$  la variable aleatoria ‘número que se obtiene al arrojar un dado equilibrado de color rojo’. Consideramos la variable aleatoria  $M$  correspondiente al mayor valor que se obtiene al tirar los dados rojo y azul. Su distribución de probabilidad está dada por la siguiente tabla.

$m$	$P(M = m)$
1	$1/36$
2	$3/36$
3	$5/36$
4	$7/36$
5	$9/36$
6	$11/36$

- a) ¿Cuál es el recorrido de  $M$ ?
- b) Verifique las probabilidades de la tabla.
- c) Calcule  $E(M)$ .
- d) Ubique el valor obtenido en la gráfica de la distribución de probabilidad de  $M$ .
- e) Repita el ejercicio con la variable aleatoria que corresponde al menor valor que se obtiene al tirar los dados rojo y azul.

## 5. Variancia de una variable aleatoria discreta

En su momento hemos expresado la variancia muestral de la siguiente forma.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 n_i = \frac{n}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 \frac{n_i}{n} = \frac{n}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_r(x_i). \quad (3)$$

También hemos visto que para  $n \rightarrow +\infty$  la frecuencia relativa correspondiente al valor  $x_i$  tiende a la probabilidad de ocurrencia de ese valor. Relacionando estas dos condiciones, resulta

$$s^2 = \frac{n}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_r(x_i) \xrightarrow{n \rightarrow +\infty} \sigma_X^2 = \sum_{i=1}^k (x_i - \mu_X)^2 p_X(x_i).$$

¿Cuál es el significado del valor de  $\sigma_X^2$  generado a partir de este *límite especial*? Se trata de la variancia calculada *en la población*. Esta población está formada por todas las observaciones posibles de la variable aleatoria  $X$ . El valor  $\sigma_X^2$  se genera a partir de la variancia muestral  $s^2$  para  $n$  tendiendo a infinito.

Sea  $X$  una variable aleatoria discreta con recorrido  $\mathcal{R}_X$  y  $p_X$  su función de probabilidad asociada. Se define variancia de una variable aleatoria  $X$  y la notamos indistintamente  $V(X)$  o  $\sigma_X^2$  al número

$$V(X) = \sigma_X^2 = \sum_{x \in \mathcal{R}_X} (x - \mu_X)^2 p_X(x).$$

La variancia de una variable aleatoria cuantifica la variación, en la población, de los valores de la variable respecto de su media. La variancia resulta grande cuando la variable asume valores alejados de la media con alta probabilidad.

El valor de  $V(X)$  está expresado en unidades al cuadrado de  $X$ . Esto motiva a considerar  $\sqrt{V(X)} = \sigma_X$  que llamamos desvío estándar de la variable aleatoria  $X$ .

Así como interpretamos la esperanza matemática de una variable aleatoria como el centro de gravedad de una distribución de masas, la variancia puede interpretarse como el momento de inercia de esa distribución de masas.

### 5.1. Ejemplo

Sea  $X$ : ‘número que se obtiene al lanzar un dado equilibrado’. La variancia de  $X$  es

$$\sigma_X^2 = (1-3,5)^2 \frac{1}{6} + (2-3,5)^2 \frac{1}{6} + (3-3,5)^2 \frac{1}{6} + (4-3,5)^2 \frac{1}{6} + (5-3,5)^2 \frac{1}{6} + (6-3,5)^2 \frac{1}{6} = 2,91\bar{6}.$$

La desviación estándar es

$$\sigma_X = \sqrt{2,91\bar{6}} \approx 1,707.$$

## 5.2. Propuesta

1. Ordene de menor a mayor las variancias correspondientes a las variables aleatorias  $X_1$ ,  $X_2$  y  $X_3$  con recorrido común  $\mathcal{R}$  cuyas distribuciones de probabilidad se muestran en los siguientes gráficos, que tienen la misma escala.



2. Verifique que para el ejemplo 2 de la página 11, el desvío estándar es aproximadamente 0,994 caracteres.
3. Argumente por qué la variancia puede interpretarse como la esperanza matemática de una variable aleatoria, comparando las definiciones correspondientes. ¿Cuál es esa variable aleatoria? Complete eligiendo una de las siguientes opciones

$$E(X - E(X)), \quad E((X - E(X))^2), \quad E(X^2 - E(X)).$$

$$V(X) = \text{.....}.$$

## 6. Distribución binomial

Llamamos ensayo de Bernoulli a toda experiencia aleatoria en relación a la cual observamos si ocurre un suceso  $A$ , o si ocurre  $\bar{A}$ . Los siguientes son ejemplos de ensayos de Bernoulli.

1. Observamos si la respuesta dada a una pregunta respondida al azar es o no correcta. En este caso podríamos considerar  $A$ : 'la respuesta es correcta',  $\bar{A}$ : 'la respuesta es incorrecta'.
2. Observamos si el tiempo hasta la falla o duración de una componente elegida al azar de un proceso de producción es o no mayor a 1000 horas. En este caso podríamos considerar  $A$ : 'una componente dura más de 1000 horas',  $\bar{A}$ : 'una componente dura a lo sumo 1000 horas'.
3. Por un canal se transmite un dígito binario y se observa si el número transmitido es 0 o 1. En este caso podríamos considerar  $A$ : 'se transmite un 0',  $\bar{A}$ : 'se transmite 1'.

En relación al ejemplo introductorio, de responder al azar una pregunta de opción múltiple, nos interesó observar si la respuesta dada es o no correcta. También en este caso se trata de un ensayo de Bernoulli. Veamos algunas condiciones que se cumplen al realizar las tres repeticiones del ensayo.

- El suceso  $A$ : 'la respuesta a una pregunta es correcta' tiene una probabilidad que permanece invariante o constante en cada ensayo. Cada pregunta tiene 4 respuestas de las cuales una es la correcta, por lo tanto  $P(A) = 1/4$  en cada ensayo.

- Los ensayos son independientes. La respuesta dada a una pregunta no influye en la respuesta dada a cualquier otra pregunta.

Bajo tales condiciones definimos la variable aleatoria  $X$ : ‘cantidad de respuestas correctas sobre un total de 3’ y se obtuvo la distribución de probabilidad asociada a dicha variable aleatoria  $X$ . Nos proponemos generalizar ese resultado particular tratado en el ejemplo introductorio.

Consideremos un ensayo de Bernoulli en relación al cual se observa si ocurre o no, un determinado suceso  $A$ , cuya probabilidad de ocurrencia es  $p = P(A)$ . Realizamos  $n$  repeticiones independientes del ensayo bajo el supuesto de que  $P(A)$  permanece constante en cada ensayo. En tales condiciones consideramos la variable aleatoria  $Y$ : ‘cantidad de veces que ocurre el suceso  $A$  en las  $n$  repeticiones independientes’. Para la variable aleatoria así definida resulta  $\mathcal{R}_Y = \{0, 1, 2, \dots, n\}$ . ¿Cuál es la distribución de probabilidad de la variable aleatoria  $Y$  así definida? Para contestar la pregunta razonamos de la siguiente manera.

Si al realizar  $n$  repeticiones de un ensayo de Bernoulli el suceso  $A$  ocurre en los primeros  $k$  ensayos y  $\bar{A}$  en los  $n - k$  ensayos restantes, es decir, cuando ocurre la secuencia

$$\underbrace{\underbrace{A A A \cdots A}_{k \text{ veces}} \underbrace{\bar{A} \bar{A} \bar{A} \cdots \bar{A}}_{n-k \text{ veces}}}_{n \text{ veces}}$$

entonces la variable aleatoria  $Y$  asume el valor  $k$ . La recíproca no es cierta. La variable aleatoria puede tomar el valor  $k$ , pero no darse necesariamente esta secuencia. Podría haber ocurrido por ejemplo

$$\underbrace{\underbrace{\bar{A} \bar{A} \bar{A} \cdots \bar{A}}_{n-k \text{ veces}} \underbrace{A A A \cdots A}_{k \text{ veces}}}_{n \text{ veces}}$$

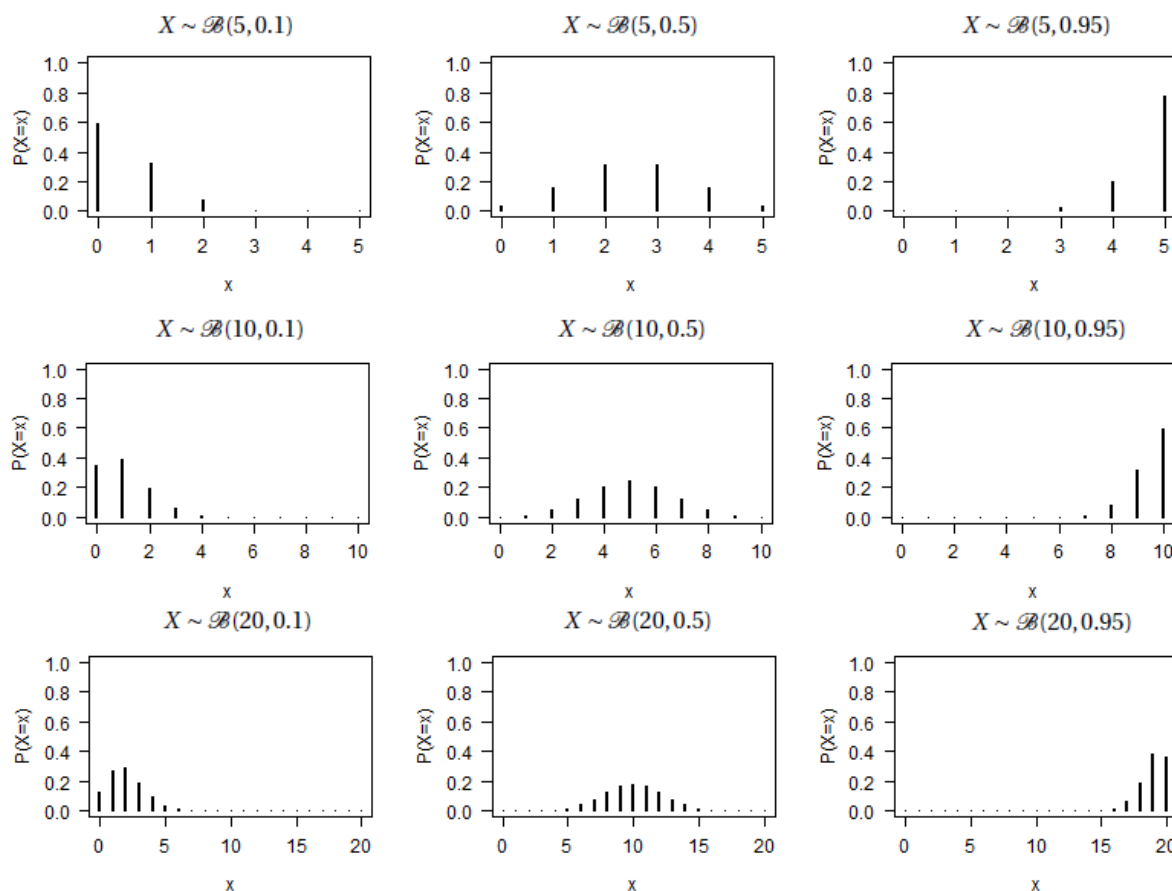
o tal vez haberse intercalado los sucesos  $A$  y  $\bar{A}$ , etc. Cualquiera de las secuencias dadas tiene una probabilidad de ocurrencia igual a  $p^k (1 - p)^{n-k}$ . Para determinar la probabilidad  $P(Y = k)$  necesitamos conocer de cuántas formas diferentes se puede presentar  $k$  veces el suceso  $A$  en  $n$  ensayos independientes de Bernoulli. Usando análisis combinatorio podemos deducir que la respuesta es  $\binom{n}{k}$ . Además las diferentes secuencias son mutuamente excluyentes (no pueden presentarse simultáneamente dos secuencias diferentes).

Sea  $A$  un suceso con probabilidad  $p$ . Entonces, la variable aleatoria  $Y$ : ‘cantidad de veces que se presenta el suceso  $A$  en  $n$  repeticiones independientes’ tiene un comportamiento o distribución binomial con parámetros  $n$  y  $p$ . En símbolos  $Y \sim \mathcal{B}(n, p)$ ; es decir

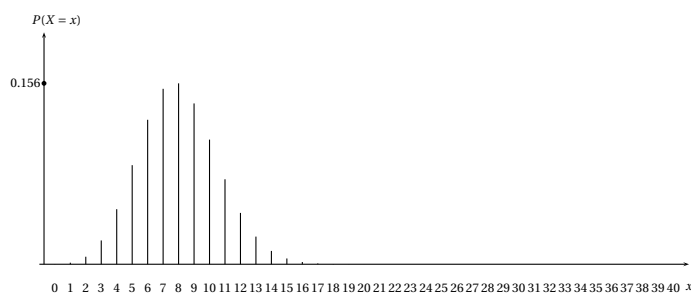
$$p_Y(k) = P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{para } k = 0, 1, 2, \dots, n.$$

Observemos que podemos considerar la variable aleatoria  $Z$ : ‘cantidad de veces que se presenta  $\bar{A}$  en  $n$  repeticiones independientes de un ensayo de Bernoulli’. Obviamente  $Z$  también tiene

distribución binomial pero con parámetros  $n$  y  $(1 - p)$  y verifica que  $Y + Z = n$ . En consecuencia  $P(Y = k) = P(Z = n - k)$ . A continuación se presentan algunas distribuciones binomiales con diferentes valores para  $n$  y  $p$ . ¿Por qué las distribuciones binomiales con  $p = 0,50$  son simétricas?



Las distribuciones binomiales con  $p \neq 0,50$  no son simétricas. Sin embargo, a medida que se aumenta el tamaño de la muestra o número de ensayos,  $n$ , las distribuciones *cambian su forma*. Observe que la distribución binomial con parámetros  $n = 40$  y  $p = 0,2$  es *razonablemente simétrica* a pesar de que  $0,2 < 0,5$ . Más adelante veremos una explicación teórica de esto.





## 6.1. Esperanza matemática y variancia

Consideremos la variable aleatoria  $X$  con distribución binomial con parámetros  $n$  y  $p$ . A partir de la definición de esperanza matemática de una variable aleatoria discreta, resulta

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \dots = np.$$

A partir de la definición de variancia de una variable aleatoria discreta resulta

$$V(X) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} = \dots = np(1-p).$$

En síntesis

Sea  $X \sim \mathcal{B}(n, p)$ . Entonces

$$E(X) = np, \quad V(X) = np(1-p).$$

### 6.1.1. Ejemplo

La probabilidad de que el vapor se condense en un tubo de aluminio de cubierta delgada a 10 atm. de presión es 0,40. Se prueban 12 tubos de ese tipo elegidos al azar.

Definimos  $X$ : ‘cantidad de tubos que condensan el vapor a 10 atm. de presión, en una muestra de 12 tubos’. Entonces  $X \sim \mathcal{B}(12, 0,40)$ . Según lo visto,

$$E(X) = 12 \times 0,40 = 4,8 \text{ tubos},$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{12 \times 0,40 \times 0,6} \approx 1,697 \text{ tubos}.$$

Es decir, si tomamos sucesivas muestras de 12 tubos con estas características y en cada una de ellas, registramos la cantidad de tubos que condensan vapor, obtendremos en general valores distintos. La media aritmética de estos valores es un número, *en general, cercano* a 4,8 tubos.

El valor 1,697 cuantifica la variación de la cantidad de tubos de aluminio que condensan vapor respecto de la esperanza matemática de 4,8 tubos.

En este problema, el coeficiente de variación es aproximadamente 0.35; es decir, el desvío estándar representa el 35 % de la esperanza matemática.

## 6.2. Propuesta

1. Un sistema está formado por 10 componentes conectadas en paralelo. Las componentes que conforman el sistema provienen de un proceso del cual se conoce que 95 % de las mismas sobreviven el instante  $t$ . Suponga que las componentes funcionan en forma independiente.
  - a) Calcule la probabilidad de que el sistema funcione después del instante  $t$ .

- b) Calcule la probabilidad que por lo menos nueve de las diez componentes sobreviven el instante  $t$ .
2. A un centro llegan dos tipos de mensajes: alta prioridad y prioridad normal. El 40 % de los mensajes son de alta prioridad. Al centro llegan, en forma independiente, seis mensajes. Calcule la probabilidad de que:
- haya la misma cantidad de mensajes de alta prioridad que de baja prioridad;
  - dos de los seis mensajes sean de alta prioridad;
  - los dos primeros mensajes sean de alta prioridad;
  - sólo los dos primeros sean de alta prioridad.
3. En la fabricación de una pieza ocurren en forma independiente dos tipos de fallas: abolladuras y roturas. La probabilidad de que una pieza presente una abolladura es 0,05 y la probabilidad de que presente rotura es 0,08. Halle la probabilidad de que al tomar 8 piezas al azar
- a lo sumo una tenga ambos defectos;
  - menos de 2 tengan algún defecto.

## 7. Distribución de Poisson

Decimos que una variable aleatoria  $X$  tiene una distribución de Poisson con parámetro  $\lambda > 0$  y notamos  $X \sim \mathcal{P}(\lambda)$  cuando

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2, \dots, n, \dots$$

Una variable con distribución de Poisson es discreta porque su recorrido es infinito numerable ( $\mathbb{N}_0$ ). Se puede probar que

$$\sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1.$$

1. La cantidad de llamadas telefónicas que recibe un centro de atención al cliente de una determinada empresa tiene una distribución de Poisson con  $\lambda = 5$  llamadas por minuto. Sea  $Y$ : 'cantidad de llamadas por minuto que recibe el centro de atención al cliente'. La probabilidad de que se reciban 2 llamadas en un minuto es

$$P(Y = 2) = \frac{5^2 e^{-5}}{2!} = 0,084.$$

2. El número de burbujas de aire en una placa de vidrio sigue un proceso de Poisson con  $\lambda = 0,4$  burbujas por cada  $m^2$ . Sea  $X$ : 'cantidad de burbujas de aire por metro cuadrado en una placa de vidrio'. La probabilidad de encontrar una burbuja en un metro cuadrado es

$$P(X = 1) = \frac{0,4^1 e^{-0,4}}{1!} = 0,268.$$

3. El número de bacterias por centímetro cúbico en un líquido es una variable aleatoria  $W$  con distribución de Poisson y  $\lambda = 4$  bacterias por cada  $\text{cm}^3$ . La probabilidad de encontrar al menos una bacteria es

$$P(W \geq 1) = \sum_{k=1}^{+\infty} \frac{4^k e^{-4}}{k!}.$$

A los fines del cálculo resulta más efectivo pensar

$$P(W \geq 1) = 1 - P(W = 0) = 1 - \frac{4^0 e^{-4}}{0!} = 1 - e^{-4} \approx 0,9817.$$

## 7.1. Esperanza matemática y variancia

De la definición de esperanza y variancia de una variable aleatoria discreta, resulta para una variable aleatoria  $X$  con distribución de Poisson con parámetro  $\lambda$

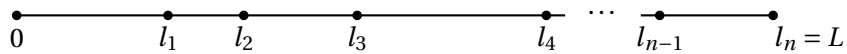
$$E(X) = \sum_{k=0}^{+\infty} k \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{y} \quad V(X) = \sum_{k=0}^{+\infty} (k - \lambda)^2 \frac{\lambda^k e^{-\lambda}}{k!}.$$

Se puede probar que  $E(X) = V(X) = \lambda$ .

## 7.2. La distribución de Poisson como caso límite de la distribución Binomial

Hemos presentado formalmente la distribución de Poisson. En lo que sigue deduciremos esta distribución a partir de la distribución Binomial.

Se tiene un alambre de longitud  $L$ , dispuesto como muestra la figura.



El alambre presenta imperfecciones. Notamos con  $X_l$ : 'cantidad de imperfecciones que presenta el alambre en el intervalo  $[0, l]$ ' y consideramos los siguientes supuestos

**Independencia.** Las variables aleatorias  $X_{l_1}, X_{l_2} - X_{l_1}, X_{l_3} - X_{l_2}, \dots, X_{l_n} - X_{l_{n-1}}$ , que representan la cantidad de imperfecciones en los intervalos disjuntos  $[0, l_1), [l_1, l_2), [l_2, l_3), \dots, [l_{n-1}, l_n)$  respectivamente, son *independientes*, cualesquiera sean  $l_1 < l_2 < \dots < l_n$  en el intervalo  $[0, L]$ . Es decir, que una variable asuma un valor no afecta la probabilidad de que otra variable asuma un valor de su recorrido.

El concepto de variables aleatorias independientes se formalizará más adelante.

**Homogeneidad.** Las variables  $X_l$  y  $X_{s+l} - X_s$  tienen la misma ley o distribución de probabilidad si ambos intervalos están contenidos en el intervalo  $[0, L]$  (los intervalos  $[0, l]$  y  $[s, s+l]$  están contenidos en  $[0, L]$ ). Coloquialmente, la distribución del número de imperfecciones en una sección de alambre depende únicamente de la longitud de la sección y no de su ubicación.

Observemos que la variable  $X_{s+l} - X_s$  corresponde a la variable  $X_l$  *trasladada* desde el origen al punto que se encuentra a una distancia  $s$  del origen.

Dividimos el alambre de longitud  $L$  en  $n$  secciones de igual amplitud (con  $n$  convenientemente grande) y agregamos a las anteriores nuevas hipótesis

- La probabilidad de que en cada sección (de longitud  $L/n$ ) haya una imperfección es proporcional a la longitud de la sección. En particular

$$P\left(X_{\frac{L}{n}} = 1\right) = \alpha \frac{L}{n}.$$

- La probabilidad de que en cada sección (de longitud  $L/n$ ) haya más de una imperfección es aproximadamente 0.

Bajo los supuestos realizados, podemos considerar que estamos en presencia de  $n$  ensayos independientes de Bernoulli (cada ensayo consiste en observar si una sección presenta una o ninguna imperfección) porque en cada sección puede haber cuanto mucho 1 imperfección (es decir, 1 o ninguna imperfección). Además la probabilidad de encontrar 1 imperfección en cada sección es constante (en cada caso vale  $\alpha \frac{L}{n}$ ). Consideramos la variable  $Y$ : ‘cantidad de secciones de longitud  $L/n$  sobre un total de  $n$  secciones, que presentan 1 imperfección’.  $Y \sim \mathcal{B}\left(n, \alpha \frac{L}{n}\right)$ , en consecuencia

$$P(Y = k) = \binom{n}{k} \left(\alpha \frac{L}{n}\right)^k \left(1 - \alpha \frac{L}{n}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Observemos que la probabilidad de que haya  $k$  secciones con una imperfección es aproximadamente igual a la probabilidad de que en un alambre de longitud  $L$  haya  $k$  imperfecciones. Cuando  $n \rightarrow +\infty$  las probabilidades se *igualan*.

$$\begin{aligned} \lim_{n \rightarrow +\infty} P(Y = k) &= \lim_{n \rightarrow +\infty} \binom{n}{k} \left(\alpha \frac{L}{n}\right)^k \left(1 - \alpha \frac{L}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow +\infty} \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \frac{(\alpha L)^k}{k!} \left(1 - \alpha \frac{L}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow +\infty} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n} \frac{(\alpha L)^k}{k!} \left(1 - \alpha \frac{L}{n}\right)^n \left(1 - \alpha \frac{L}{n}\right)^{-k} \\ &= \frac{(\alpha L)^k}{k!} e^{-\alpha L} = P(X_L = k). \end{aligned}$$

El razonamiento anterior es cierto para todo  $k$  entero no negativo. En consecuencia,  $X_L$  tiene una distribución de Poisson con parámetro  $\lambda = \alpha L$ . Como la esperanza de una variable con distribución de Poisson de parámetro  $\lambda$  es igual al propio  $\lambda$ , resulta que  $E(X_L) = \alpha L$  o equivalentemente  $\alpha = \frac{E(X_L)}{L}$ ; es decir,  $\alpha$  representa la media poblacional de imperfecciones por unidad de longitud.

Observe que:

$$\begin{aligned} E(Y) &= n \cdot \alpha \cdot \frac{L}{n} = \alpha \cdot L = E(X_L) \\ V(Y) &= n \cdot \alpha \cdot \frac{L}{n} \cdot \left(1 - \alpha \cdot \frac{L}{n}\right) \rightarrow \alpha \cdot L = V(X_L), \text{ cuando } n \rightarrow \infty \end{aligned}$$

Cuando  $n$  es ‘grande’ y  $p$  es ‘pequeña’, las probabilidades binomiales pueden ser aproximadas a través de la distribución Poisson.

### 7.3. Ejemplos

1. La cantidad de bacterias que aparecen en el agua se distribuyen según una ley de Poisson a razón de 2 bacterias/cm<sup>3</sup>.

a) ¿Cuál es la probabilidad de que en una muestra de 3 cm<sup>3</sup> haya por lo menos 1 bacteria?

b) ¿Cuál es la probabilidad de que en una muestra de 5 cm<sup>3</sup> no haya bacterias?

Definimos la variable  $X_\nu$ : 'cantidad de bacterias por cada  $\nu$  cm<sup>3</sup>'. Entonces  $X_3 \sim \mathcal{P}(\lambda \nu = 2 \times 3 = 6)$ . Luego

$$P(X_3 \geq 1) = 1 - P(X_3 = 0) = 1 - \frac{6^0 e^{-6}}{0!} \approx 0,9975.$$

$X_5 \sim \mathcal{P}(\lambda \nu = 2 \times 5 = 10)$ . Luego

$$P(X_5 = 0) = \frac{10^0 e^{-10}}{0!} \approx 4,5 \times 10^{-5}.$$

2. La probabilidad de que una viga falle por compresión 0,02. ¿Cuál es la probabilidad de que en una muestra de 100 vigas fallen a lo sumo 4 vigas?

Llamamos  $X$ : 'cantidad de vigas que fallan por compresión en una muestra aleatoria de 100 vigas' Entonces  $X \sim \mathcal{B}(100, 0,02)$ . Luego

$$P(X \leq 4) = \sum_{i=0}^4 \binom{100}{i} 0,02^i 0,98^{100-i} \approx 0,9491.$$

Si aproximamos el cálculo utilizando una distribución de Poisson, resulta para la misma variable aleatoria pero ahora con una distribución de Poisson de parámetro  $\lambda = E(X) = 100 \times 0,02 = 2$

$$P(X \leq 4) = \sum_{i=0}^4 \frac{2^i e^{-2}}{i!} \approx 0,9473.$$

## 8. Distribución geométrica

Consideremos el experimento de lanzar un dado equilibrado hasta que se presenta por primera vez el número cuatro. Sea la variable aleatoria  $X$ : 'cantidad de lanzamientos realizados hasta que el número cuatro aparece por primera vez'. Por ejemplo, si la secuencia de resultados del lanzamientos del dado es: 3,2,1,3,5,4 entonces la variable aleatoria  $X$  asume el valor 6 porque el número 4 aparece por primera vez en el 6° lanzamiento.

Una variable aleatoria así definida sólo puede asumir valores enteros positivos y por lo tanto se trata de una variable aleatoria discreta.

Nos interesa calcular con qué probabilidad la variable aleatoria  $X$  asume cada valor de su recorrido; es decir, queremos determinar su distribución de probabilidad. Para esto, razonamos de la siguiente manera.

1. La variable aleatoria  $X$  asume el valor 1 siempre y cuando el número 4 aparezca en el primer lanzamiento. Por lo tanto  $P(X = 1) = \frac{1}{6}$ .
2. La variable aleatoria  $X$  asume el valor 2 siempre y cuando en el primer lanzamiento no aparezca el número 4 pero aparezca en el 2° lanzamiento. Por lo tanto  $P(X = 2) = \frac{5}{6} \times \frac{1}{6}$ , donde  $\frac{5}{6}$  es la probabilidad de que no aparezca el 4 en el primer lanzamiento.
3. La variable aleatoria  $X$  asume el valor 3 siempre y cuando no aparezca el número 4 en los primeros dos lanzamientos, y recién lo haga en el tercer lanzamiento. Por lo tanto  $P(X = 3) = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6}$ .
4. ...
5. En general  $X$  asume el valor  $k$  siempre y cuando en los primeros  $k - 1$  lanzamientos no aparezca el número 4 y recién aparezca en el  $k$ -ésimo lanzamiento. Por lo tanto  $P(X = k) = \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6}$ .

Cuando una variable aleatoria  $X$  asume los valores  $1, 2, \dots, k, \dots$  con  $P(X = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$  decimos que  $X$  tiene una distribución geométrica con parámetro  $p = \frac{1}{6}$ .

Veamos las condiciones que se cumplen en las consideraciones realizadas.

1. Se parte de un ensayo de Bernoulli, es decir, de una experiencia aleatoria en relación a la cual se observa si ocurre el suceso  $A$ : 'se obtiene un 4' o si ocurre  $\bar{A}$ : 'no se obtiene un 4'.
2. Se realizan repeticiones independientes del ensayo hasta que el suceso  $A$  se presenta por primera vez.
3. La probabilidad del suceso  $A$  permanece constante en cada ensayo.

En general

1. Se parte de una experiencia aleatoria en relación a la cual interesa observar si ocurre un cierto suceso  $A$ , con  $P(A) = p$  o si ocurre  $\bar{A}$ , con  $P(\bar{A}) = 1 - p$  (ensayo de Bernoulli).
2. Se realizan repeticiones independientes del ensayo hasta que el suceso  $A$  se presenta por primera vez.
3. La probabilidad del suceso  $A$  permanece constante en cada ensayo,

Bajo estas consideraciones se define la variable aleatoria  $X$ : 'cantidad de ensayos que se realizan hasta que el suceso  $A$  se presenta por primera vez'. Calculamos la probabilidad de que la variable aleatoria  $X$  asuma el valor  $k$ ; es decir, que el suceso  $A$  se presente por primera vez en el  $k$ -ésimo ensayo.

$$P(X = k) = P(\underbrace{\bar{A}\bar{A}\bar{A}\cdots\bar{A}}_{k-1 \text{ veces}} A) = P(\bar{A})^{k-1} P(A) = (1 - p)^{k-1} p.$$

Sea  $X$  la variable aleatoria definida como ‘cantidad de ensayos independientes de Bernoulli que se realizan hasta que el suceso  $A$  se presenta por primera vez’, y sea  $P(A) = p$ . Entonces  $X$  tiene una distribución geométrica con parámetro  $p$  y

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots$$

Bajo estas condiciones notamos  $X \sim \mathcal{G}(p)$ .

### 8.1. Esperanza matemática y variancia

Cuando  $X \sim \mathcal{G}(p)$  se puede probar que

$$E(X) = \sum_{k=1}^{+\infty} k p (1 - p)^{k-1} = \frac{1}{p}.$$

En el caso de la variancia, se puede probar que

$$V(X) = \sum_{k=1}^{+\infty} \left(k - \frac{1}{p}\right)^2 p (1 - p)^{k-1} = \frac{1 - p}{p^2}.$$

### 8.2. Período de retorno

Sea el suceso  $A$ : en un año el caudal para una sección específica de un río es de por lo menos  $100 \text{ m}^3/\text{s}$  y consideremos la variable aleatoria  $X$ : ‘cantidad de años que transcurren hasta que el suceso  $A$  ocurre por primera vez, o se repite por primera vez desde la última vez que ocurrió’. La variable  $X$  asume el valor 1 si dentro del 1° año ocurre el suceso  $A$  por lo menos una vez desde la última vez que haya ocurrido. La variable  $X$  asume el valor 2 si dentro del 1° año no ocurre el suceso  $A$  desde la última vez que haya ocurrido y sí lo hace durante el 2° año.

Bajo el supuesto de que los caudales ocurren año tras año de manera independiente entonces la variable aleatoria  $X$  tiene una distribución geométrica con parámetro  $p = P(A)$  y por lo tanto  $E(X) = 1/P(A)$ . Si  $P(A) = 0,02$  entonces  $E(X) = 50$  años. Este valor se conoce por periodo de retorno de un caudal de por lo menos  $100 \text{ m}^3/\text{s}$ . y representa el tiempo medio en años entre la presentación del suceso  $A$ . Esto no implica que no puedan producirse, por ejemplo, en dos años consecutivos, caudales de por lo menos  $100 \text{ m}^3/\text{s}$ , aunque la probabilidad de ocurrencia sea muy baja.

## 9. Distribución binomial negativa

Esta distribución, también conocida como distribución de Pascal, es una generalización de la distribución geométrica. La variable con distribución geométrica se define como ‘cantidad de ensayos independientes que se realizan hasta que el suceso  $A$  se presenta por primera vez’; en el caso de una distribución binomial negativa, la variable se define como ‘cantidad de ensayos que se realizan hasta que el suceso  $A$  se presenta por  $r$ -ésima vez’.

Para deducir la ley de probabilidad de una variable aleatoria  $X$  con distribución binomial negativa, razonamos a partir de este ejemplo.

Sea  $X$ : ‘cantidad de veces que se lanza un dado equilibrado hasta que el número 6 se presenta por cuarta vez’. En primer lugar, destacamos que el menor valor que puede asumir  $X$  es 4 (porque son necesarios como mínimo 4 tiradas para que el número 6 aparezca por cuarta vez.)

¿Cuál es la probabilidad de que en el octavo lanzamiento, el número 6 aparezca por cuarta vez? Para responder, observamos que en los primeros 7 lanzamientos debe aparecer tres veces el número 6, y en el octavo lanzamiento aparecer el 6 por cuarta vez.

La probabilidad de que en 7 lanzamientos aparezca 3 veces el número 6 es

$$\binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4 \text{ ¿por qué?}$$

En consecuencia,

$$P(X = 8) = \binom{7}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} = \binom{7}{3} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^4 \approx 0,013$$

Sea  $X$  la variable aleatoria definida como ‘cantidad de ensayos independientes de Bernoulli que se realizan hasta que el suceso  $A$  se presenta por  $r$ -ésima vez’, y sea  $P(A) = p$ . Entonces  $X$  tiene una distribución binomial negativa con parámetros  $p$  y  $r$ .

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \dots$$

Además,

$$E(X) = \frac{r}{p}, \quad V(X) = \frac{r(1-p)}{p^2}.$$

## 9.1. Ejemplo

La probabilidad de que al comprar un automóvil se lo pida de color negro es 0,40. ¿Cuál es la probabilidad que el octavo cliente en comprar un automóvil sea el tercero en solicitarlo de este color? ¿Cuál es la cantidad media de ventas hasta que tres clientes demanden un automóvil de color negro?

Sea  $Y$ : ‘cantidad de ventas que se realizan hasta que tres clientes demanden un automóvil de color negro’ Entonces  $Y$  tiene una distribución binomial negativa con parámetros  $p = 0,40$  y  $r = 3$ .

$$P(Y = 8) = \binom{7}{2} 0,40^3 0,60^5 \approx 0,10.$$

$$E(Y) = \frac{3}{0,40} = 7,5 \text{ ventas.}$$



## 10. Distribución hipergeométrica

Se tiene un lote de  $N$  artículos de los cuales  $r$  son defectuosos ( $N - r$  no lo son). Se eligen al azar y sin reposición  $n$  artículos del lote ( $n \leq N$ ). Sea  $X$  la cantidad de artículos defectuosos en la muestra de tamaño  $n$ . Entonces  $\mathcal{R}_X = \{0, 1, 2, \dots, \min(n, r)\}$ . En el capítulo correspondiente a la Introducción a la Probabilidad, hemos calculado en base a un razonamiento combinatorio,

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots, \min(n, r).$$

Sea  $p = r/N$ , es decir, la proporción de defectuosos en el lote. Se puede demostrar que  $E(X) = np$  y  $V(X) = np(1-p) \frac{N-n}{N-1}$ .

### 10.1. Ejemplos

De 800 piezas fabricadas por una máquina, 25 resultaron defectuosas. Se eligen al azar 5 piezas del lote de 800 piezas. Calcule la probabilidad de que en la muestra de tamaño 5 haya al menos una defectuosa, suponiendo que

1. el muestreo es sin reposición,
2. el muestreo es con reposición.
3. Si los resultados que obtiene son similares, explique el porqué.

Solución:

1. Sea  $X$ : ‘cantidad de elementos defectuosos en la muestra de tamaño 5 extraída sin reposición’. Entonces  $X$  tiene distribución hipergeométrica.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{775}{5} \binom{25}{0}}{\binom{800}{5}} \approx 0,147.$$

2. Sea  $Y$ : ‘cantidad de elementos defectuosos en la muestra de tamaño 5 extraída con reposición’. Entonces  $Y$  tiene distribución binomial con parámetros  $n=5$ ,  $p=\frac{25}{800}$ .

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{5}{0} \left(\frac{25}{800}\right)^0 \left(\frac{775}{800}\right)^5 \approx 0,146.$$

3. La variable  $X$  *tiende* a comportarse como la variable  $Y$  cuando el tamaño de la población es *grande* respecto al tamaño de la muestra. Esto se debe a que las probabilidades calculadas en el primer caso no difieren sustancialmente de la probabilidad de éxito para la variable  $Y$ . La distribución hipergeométrica se aplica cuando se realiza un muestro sin reposición en una población finita. En cambio, la distribución binomial se aplica cuando se realiza un muestreo con reposición en una población finita o sin reposición en una población infinita.

## 11. Miscelánea de problemas

1. El número de bacterias por  $\text{mm}^3$  de agua en un estanque es una variable aleatoria  $X$  con distribución de Poisson de parámetro  $\lambda = 0,5$ .
  - a) ¿Cuál es la probabilidad de que en  $2 \text{ mm}^3$  haya al menos una bacteria?
  - b) En 40 tubos de ensayo se toman muestras de agua del estanque ( $1 \text{ mm}^3$  de agua en cada tubo). ¿Qué distribución tiene la variable aleatoria  $Y$ : ‘número de tubos de ensayo, entre los 40, que no contienen bacterias’? Justifique
2. Un proceso de producción se detiene para hacer ajustes cuando en una muestra al azar de cinco unidades hay una o más defectuosas.
  - a) Si el proceso produce 2 % de unidades defectuosas, calcule la probabilidad de que sea detenido al inspeccionarse.
  - b) Calcule la probabilidad de que el proceso no sea detenido después de una inspección cuando está produciendo un 10 % de unidades defectuosas.
  - c) En el contexto de la situación planteada emita juicio sobre los valores de las probabilidades calculados en los apartados 2a y 2b.
3. El número de cortes de luz, con una duración superior a una hora, que se produce en un distrito de una ciudad sigue una ley de Poisson a razón de 10 cortes por año.
  - a) Calcule la probabilidad de que en medio año se produzcan a lo sumo dos cortes.
  - b) Una empresa de suministro de energía afirma que ha renovado sus equipos e implementado un nuevo sistema de mantenimiento que les permite un mejor abastecimiento. Después de la renovación de sus equipos se han producido en un año tres cortes. Calcule la probabilidad de que se produzcan a lo sumo tres cortes aún cuando el promedio sigue siendo de 10 cortes por año.
  - c) A la luz del cálculo efectuado en 3b ¿considera que ha mejorado el servicio de abastecimiento? Justifique.
4. Sean  $X$  e  $Y$  dos variables aleatorias discretas tales que  $x < y$  para todo  $x \in \mathcal{R}_X$ , para todo  $y \in \mathcal{R}_Y$ . Analice la veracidad de los siguientes enunciados. Justifique.
  - a)  $V(X) < V(Y)$ .
  - b)  $E(X) < E(Y)$ .
5. Sea  $U$ : ‘número máximo que se obtiene al lanzar dos dados equilibrados’ y  $T$ : ‘número mínimo que se obtiene al lanzar dos dados equilibrados’. Verifique que  $V(U) = V(T)$ .
6. Una empresa compra *una gran cantidad* de bulones a un mismo proveedor. Cada vez que se recibe un envío se realiza un control de calidad por muestreo. Se prueban 80 bulones seleccionados al azar del total. Si se encuentra más de un bulón defectuoso se rechaza el envío; en caso contrario, se lo acepta. Sea  $p$  la probabilidad de producir un bulón defectuoso en el proceso de fabricación. De acuerdo a los estándares de calidad, se desea satisfacer la condición  $p \leq 0,005$ .

- a) Halle la probabilidad de rechazar un lote cuando  $p = 0,004$ .
  - b) Halle la probabilidad de aceptar un lote cuando  $p = 0,05$ .
  - c) Si fuera usted el proveedor, ¿cuál de las probabilidades obtenidas en los puntos anteriores le preocuparía? Justifique su respuesta.
- 7. Una fabricante realiza un control de sus artículos antes de que se embarquen. De cada caja de 25 unidades, se seleccionan al azar y sin reposición 3 artículos. Si se encuentra al menos uno defectuoso, se verifica toda la caja. Si no se encuentran defectuosos, la caja se embarca.
  - a) Calcule la probabilidad de que se embarque una caja que contiene tres defectuosos.
  - b) Calcule la probabilidad de que una caja que contiene sólo un artículo defectuoso se regrese para su revisión total.
  - c) ¿Qué comentario le merecen las probabilidades calculadas en los apartados anteriores?
- 8. El número de defectos en un rollo de alambre puede describirse a través de una distribución de Poisson a razón de 1 defecto cada 10 metros. Un comprador decide adquirir un lote de 10 rollos si al inspeccionar uno cualquiera al azar no encuentra defectos en los primeros 20 metros de alambre. Calcule la probabilidad de aceptar el lote.
- 9. El departamento de selección de personal de una empresa sabe que ante cada convocatoria para cubrir un puesto gerencial, sólo el 20% de los aspirantes cumple con todos los requisitos exigidos. Se hace una convocatoria para cubrir el puesto de un gerente por una reciente jubilación, y se entrevista a los aspirantes uno a uno.
  - a) Calcule la probabilidad de que el cuarto entrevistado sea el primer aspirante que cumpla con todos los requisitos de la convocatoria.
  - b) Determine la media del número de aspirantes a entrevistar hasta encontrar el primero que cumpla con los requisitos pedidos en la convocatoria.
- 10. El cuerpo de bomberos de una ciudad es capaz de atender hasta un máximo de 30 servicios por día. La demanda diaria de servicios sigue una ley de Poisson con una media igual a 25 servicios.
  - a) Calcule la probabilidad de que en un día cualquiera no se puedan atender todos los servicios requeridos.
  - b) Calcule la probabilidad de que en 31 días haya al menos un día en el que no se puedan atender todos los servicios requeridos.
- 11. Una empresa desea comprar un lote de envases de vidrio y quiere asegurarse de que la proporción de envases con la rosca defectuosa no supere 0,01. Para verificar si un lote entregado cumple con el requerimiento se extrae una muestra de 250 envases y se inspeccionan. Sea  $X$  el número de envases inspeccionados con la rosca defectuosa. Se aceptará el lote si  $X \leq c$ , (para algún valor cuidadosamente elegido de  $c$ ) y si  $X > c$ , el lote será devuelto.

- a) ¿Cómo debe elegirse  $c$ , si se quiere tener una probabilidad de a lo sumo 0,1 de aceptar el lote cuando la verdadera proporción de roscas defectuosas es 0,02?
- b) Para ese valor de  $c$ , halle la probabilidad de rechazar un lote con una proporción de roscas defectuosas de 0,005.
12. Dos equipos  $A$  y  $B$  llevan a cabo un trabajo similar. La siguiente tabla muestra la distribución de probabilidad de la variable aleatoria tiempo para la finalización del trabajo, en días, para cada equipo.

	Días		
	6	7	8
<b>Equipo A</b>	0,15	0,65	0,20
<b>Equipo B</b>	0,08	0,70	0,22

Calcule para cada equipo el tiempo medio y la desviación estándar del tiempo para la finalización del trabajo.

13. El número de fallas de un instrumento de prueba debido a las partículas contaminantes de un producto, es una variable aleatoria de Poisson con media 0,02 fallas por hora. Se desea que la probabilidad de que un instrumento de tal tipo falle en un período de 4 horas sea a lo sumo 0,05.
- a) ¿Satisfacen dichos instrumentos las condiciones deseadas?
- b) Calcule la probabilidad de que se presente al menos una falla en un período de 24 horas.
14. En una línea de producción de placas para circuitos electrónicos se conoce que: el 5 % de las placas no satisfacen la especificación de longitud y el 3 % no satisface la especificación de anchura. Si los cortes para la longitud y el ancho son variables independientes calcule
- a) la proporción de placas que satisfacen ambas especificaciones,
- b) la proporción de placas que no satisfacen al menos una de las dos especificaciones,
- c) la proporción de placas que satisfacen la especificación de longitud entre aquellas que satisfacen ambas especificaciones,
- d) la probabilidad de que en una muestra de 10 placas a lo sumo una de las mismas no cumpla con ambas especificaciones.
15. En un taller de fabricación de una pieza seriada, hay dos tornos automáticos que producen una misma pieza. El torno  $T_1$  produce 97% de unidades buenas mientras que el torno  $T_2$  95%. Un lote se compone de 80% de piezas trabajadas con el torno  $T_1$  y 20% de piezas trabajadas por el torno  $T_2$ . Se forman lotes con 1000 piezas. Un cliente acepta un lote cuando en una muestra de 5 piezas elegidas al azar no encuentra piezas defectuosas.
- a) Calcule la probabilidad de aceptación de un lote.
- b) Calcule el número medio de lotes que se aceptan, cuando se inspeccionan de a 50 lotes.

- c) ¿Realizó algún cálculo donde usó el supuesto de independencia? En caso afirmativo, indique dónde y por qué considera válido realizar ese supuesto.
16. Un lote de 30 piezas contiene 3 que son defectuosas. Un comprador utiliza el siguiente plan de muestreo para la aceptación. Si en una muestra de tamaño 5, sin reposición, encuentra al menos una pieza defectuosa, rechaza el lote; de lo contrario lo acepta.
- a) Calcule la probabilidad de que el lote sea aceptado.
- b) ¿Es válido aproximar la probabilidad del punto anterior a través de la distribución binomial? Explique.
- c) Si se inspeccionan de a 100 de tales lotes, ¿cuál es el número promedio de lotes rechazados?
17. Si  $X$  es una variable aleatoria discreta entonces  $E(X)$  es el valor que tiene mayor probabilidad de ocurrir. ¿Esto es verdadero o falso? Justifique.
18. El departamento de calidad ha evaluado que cierto tipo de anclajes metálicos producidos pueden ser defectuosos debido a las siguientes causas: defectos en la rosca y defectos en las dimensiones. Se ha calculado que el 6% de los anclajes que producen tienen defectos en la rosca, mientras que el 9% tiene defectos en las dimensiones. Sin embargo, el 90% de los anclajes no tiene ningún tipo de defectos.
- a) ¿Cuál es la probabilidad de que un anclaje tenga ambos tipos de defectos?
- b) Los anclajes se empaquetan en cajas de 300 unidades. ¿Cuál es la probabilidad de que una caja tenga más de un 10% de anclajes con algún defecto?
19. Según un informe el número de veces que una planta eléctrica libera cantidades detectables de gases radioactivos, en un período de un mes, es una variable aleatoria con distribución de Poisson a razón de 3 por mes.
- a) En función de los datos informados determine el valor de la esperanza matemática o media de emisiones, en un período de un bimestre.
- b) Calcule la probabilidad de que en un período de un bimestre se detecten por lo menos 12 emisiones.
- c) Se detectaron 12 emisiones durante un período de un bimestre, ¿consideran que existe razones para dudar acerca del valor de la media declarada en el informe? Sugerencia: fundamente su respuesta en base al cálculo realizado en b).
20. Veinte por ciento de los teléfonos de cierto tipo se envían a reparación cuando todavía está vigente su garantía. De éstos, el 60% se puede reparar y el otro 40% debe sustituirse con aparatos nuevos. Si una compañía compra diez de estos teléfonos, ¿cuál es la probabilidad de que exactamente se cambien dos dentro del periodo de garantía?
21. Un lote muy grande de componentes ha llegado a un distribuidor. El distribuidor decide seleccionar al azar 10 componentes y aceptar el lote si el número de componentes defectuosas en la muestra es a lo sumo 2.

- a) ¿Cuál es la probabilidad de que el lote sea aceptado cuando la proporción real de piezas defectuosas es 0,01? ¿0,05?, ¿0,10?, ¿0,20?, ¿0,25?
- b) Sea  $p$  la proporción de piezas defectuosas del lote. La gráfica que representa la probabilidad de aceptar el lote en función de  $p$  se llama *curva característica de operación* para el plan de muestreo de aceptación del lote. Utilice los resultados del apartado anterior para trazar esta curva para  $p \in [0, 1]$ .
- c) Repita los dos primeros apartados cambiando “2” por “1” en el plan de muestreo de aceptación del lote.
- d) Repita los dos primeros apartados cambiando “10” por “15” en el plan de muestreo de aceptación del lote.
- e) ¿Cuál de los tres planes de muestreo aparece más satisfactorio? Justifique.
22. En la serie de campeonato de la NBA, el equipo que gane 4 juegos de 7 será el ganador de la serie. Suponga que el equipo  $A$  tiene una probabilidad de 0,55 de ganarle al equipo  $B$  y que ambos equipos se enfrentarán entre sí en los juegos de campeonato.
- a) ¿Cuál es la probabilidad de que el equipo  $A$  gane la serie en 6 juegos?
- b) ¿Cuál es la probabilidad de que el equipo  $A$  gane la serie?
- c) Si ambos equipos se enfrentan entre sí en una serie regional de play-off y el ganador es quien gana 3 de 5 juegos. ¿Cuál es la probabilidad de que el equipo  $A$  ganará un juego de play-off?
23. Un lote contiene un 2% de artículos defectuosos. ¿Cuál es la probabilidad de que sea necesario extraer de manera aleatoria a lo sumo 20 artículos para encontrar el primer defectuoso?

## 12. Soluciones

Presentamos a continuación algunas respuestas a los problemas de la miscelánea, junto con los comandos de MS-EXCEL o R que permiten reproducirlas.

1. a) 0,632
- En MS-EXCEL =1-POISSON(0;0,5\*2;0)
  - En R `ppois(0,0.5*2,lower.tail=FALSE)`
- b)  $Y \sim \mathcal{B}(n = 40, p = 0,607)$
- En MS-EXCEL =POISSON(0;0,5;0)
  - En R `dpois(0,0.5)`
2. a) 0,096
- En MS-EXCEL =1-DISTR.BINOM(0;5;0,02;0)
  - En R `1-pbinom(0,5,0.02,lower.tail=FALSE)`
- b) 0,590

- En MS-EXCEL =DISTR.BINOM(0;5;0,1;0)
  - En R dbinom(0,5,0.1)
3. a) 0,125
- En MS-EXCEL =POISSON(2;5;1)
  - En R ppois(2,5)
- b) 0,010
- En MS-EXCEL =POISSON(3;10;1)
  - En R ppois(3,10)
4. a) Falso.
- b) Verdadero.
5. 1,97 puntos<sup>2</sup>.
6. a) 0,041
- En MS-EXCEL =1-DISTR.BINOM(1;80;0,004;1)
  - En R pbinom(1,80,0.004,lower.tail=FALSE)
- b) 0,086
- En MS-EXCEL =DISTR.BINOM(1;80;0,05;1)
  - En R pbinom(1,80,0.05)
7. a) 0,670
- En MS-EXCEL=DISTR.HIPERGEOM(3;3;22;25)
  - En R dhyper(3,22,3,3)
- b) 0,120
- En MS-EXCEL=DISTR.HIPERGEOM(2;3;24;25)
  - En R dhyper(2,24,1,3)
9. a) 0,102
- En MS-EXCEL =NEGBINOMDIST(3;1;0,2)
  - En R dgeom(3,0.2)
- b) 5 aspirantes.
10. a) 0,137
- En MS-EXCEL =1-POISSON(30;25;1)
  - En R ppois(30,25,lower.tail=FALSE)
- b) 0,990
- En MS-EXCEL =1-DISTR.BINOM(0;31;1-POISSON(30;25;1);0)
  - En R pbinom(0,31,ppois(30,25,lower.tail=FALSE),lower.tail=FALSE)
13. a) No.

- `En R ppois(0,0.02*4,lower.tail=FALSE)`
- b) 0,381
  - `En R ppois(0,0.02*24,lower.tail=FALSE)`
- 15. a) 0,841
  - `En R dbinom(0,5,.03*.8+.05*.2)`
- 16. a) 0,567
  - `En R dhyper(5,27,3,5)`

b) 0,590

  - `En R dbinom(0,5,0.1)`
- 18. b) 0,452
  - `En R pbinom(30,300,0.1,FALSE)`
- 20. 0,148
  - `En R dbinom(2,10,0.08)`
- 22. a) 0,185
  - `En MS-EXCEL =NEGBINOMDIST(2;4;0,55)`
  - `En R dnbinom(2,4,0.55)`

b) 0,292

  - `En MS-EXCEL =DISTR.BINOM(4;7;0,55;0)`
  - `En R dbinom(4,7,0.55)`

c) 0,337

  - `En MS-EXCEL =DISTR.BINOM(3;5;0,55;0)`
  - `En R dbinom(3,5,0.55)`