Guía de Ejercitación 5. Variables aleatorias multidimensionales

Esta guía contiene tres ejes temáticos

- Parte I. Variables aleatorias bidimensionales caso discreto. Se dejará como opcional el tratamiento del caso continuo. Páginas 1 a 3.
- Parte II. Propiedad reproductiva de la distribución normal. Páginas 4 a 6
- Parte III. Distribuciones muestrales. Teorema Central del Límite. Páginas 7 a 13

Cuenta con material adicional. Páginas 14 a 17.

Guía Ejercitación 5. Parte I. Variables aleatorias bidimensionales caso discreto

Ejercicio 1. La calidad de un producto se determina de acuerdo al número de defectos que contiene. Sea X la variable aleatoria que representa el número de defectos por unidad. Estos productos son sometidos a revisión. Sea Y la variable aleatoria que representa el número de defectos por unidad detectados por el sistema de revisión. La distribución conjunta de probabilidades $p_{X,Y}$ se muestra en la tabla.

		y		
		0	1	2
	0	0,8148	0,0168	0,0084
x	1	0,0010	0,0980	0,0010
	2	0,0006	0,0012	0,0582

- **a)** Calcular $P(X > Y)$, $P(X = Y)$ y $P(X < Y)$. Expresar en forma literal qué representan cada uno de los eventos cuyas probabilidades se han calculado.
- **b)** Encontrar la distribución marginal de X y la correspondiente a Y.
- **c)** Determinar los valores esperados de X y de Y, y las dispersiones de dichas variables.
- **d)** Evaluar las probabilidades condicionales definidas por:
$$P(Y \leq X/X = 1) \text{ y } P(Y \leq X/X \leq 1).$$
- **e)** Hallar el valor esperado de $X - Y$.
- **f)** ¿Son las variables X y Y independientes?
- **g)** ¿Cómo se calcularía la varianza de $X - Y$? ¿Daría igual que la resta o la suma de las varianzas de X y de Y?
- **h)** Hallar la matriz de covarianzas y el coeficiente de correlación de (X, Y).

Observación: la matriz de covarianzas de dos variables aleatorias X y Y es una matriz simétrica definida por

$$\begin{pmatrix} Cov(X,Y) & Cov(X,Y) \\ Cov(Y,X) & Cov(Y,Y) \end{pmatrix} = \begin{pmatrix} V(X) & Cov(X,Y) \\ Cov(Y,X) & V(Y) \end{pmatrix},$$

,

donde $Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$. El coeficiente de correlación está definido por

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Ejercicio 2. Dada la variable aleatoria bidimensional definida por el conjunto de pares

$$\{(X,Y)\} = \{(-2,4); (-1,1); (1,1); (2,4)\}$$

tales que todos ellos tienen igual probabilidad $1/4$,

- **a)** obtener las distribuciones marginales de X y de Y.
- **b)** Calcular $E(X)$, $E(Y)$, $E(XY)$.

c) Verificar que X y Y no son variables aleatorias independientes (existe entre ellas una relación funcional sencilla).

d) Para la variable $W = X + Y$, hallar la tabla de distribución de probabilidad, el valor esperado y la varianza.

Ejercicio 3. El temario de un examen puede contener una, dos o tres preguntas con igual probabilidad (el que diseña el examen no tiene preferencia por ninguna de las tres modalidades). La probabilidad que se tiene de contestar mal una pregunta es $1/4$, independientemente de cuantas preguntas haya y de cualquier otro factor que incida en las respuestas. Sea X la variable aleatoria que indica el número de preguntas que contiene el examen y sea Y la variable aleatoria que corresponde al número de preguntas que se contestan mal en el examen.

a) Esquematizar un diagrama de árbol que identifique la situación.

b) Hallar la probabilidad conjunta $p_{X,Y}(x, y)$. Presentar los resultados en una tabla.

c) ¿Son X y Y variables independientes?

d) ¿Cuál es la probabilidad de que se conteste todo el examen en forma correcta, esto es que todas las preguntas que contenga el examen se contesten correctamente?

e) Si el criterio de aprobación del examen es el de contestar al menos una pregunta bien, ¿cuál es la probabilidad de aprobar el examen?

f) Si el examen es de tres preguntas, ¿cuál es la probabilidad de que tenga exactamente dos respuestas erróneas? Expresar la pregunta en función de valores adecuados de las variables X y Y, con el formato apropiado.

g) Si el examen tiene exactamente dos respuestas erróneas, ¿cuál es la probabilidad de que el examen haya sido de tres preguntas? Expresar la pregunta en función de valores adecuados de las variables X y Y, con el formato apropiado.

h) Sea W la variable aleatoria correspondiente a la nota del examen que se asigna con el siguiente criterio: el número de respuestas correctas multiplicado por 10 y dividido por el número de preguntas totales en el examen. ¿Cuál es el valor esperado de W? ¿Cuál es su dispersión?

Ejercicio 4. Un juego consiste en lanzar dos dados no cargados y la suma de los números que salen determina la ganancia de los participantes. Las variables aleatorias X y Y representan, respectivamente, las ganancias de los jugadores primero y segundo. El primero cobra \$3 si la suma es 4, 5 ó 6 y paga \$2 si la suma es 11 ó 12; para el resto de los posibles resultados, no cobra ni paga. El segundo jugador cobra \$2 para una suma de al menos 8, paga \$3 si obtiene una suma a lo sumo de 5, y no cobra ni paga para el resto de los posibles resultados.

a) Presentar la tabla de probabilidad conjunta.

b) Se definen los siguientes eventos

A: las ganancias del primer jugador son mayores a las del segundo;
B: un jugador tiene una ganancia superior a la del otro;
C: el primer jugador no cobra dinero;
D: el segundo jugador no paga dinero.

Identificarlos en términos de las variables X y Y. Hallar sus probabilidades.

c) Calcular $P(X > 0 \cup Y < 0)$.

d) Sabiendo que el primer jugador no pierde dinero, ¿cuál es la probabilidad de que el segundo jugador tampoco pierda dinero?

e) Dar la interpretación literal a $P(X > 0/Y < 0)$ y calcular su valor.

f) ¿Cuál es el valor esperado de la ganancia de cada jugador?

g) Hallar la covarianza entre ambas variables, $Cov(X, Y)$, y el coeficiente de correlación, $Corr(X, Y)$.

Ejercicio 5. En el desarrollo de un nuevo receptor para la transmisión de información digital, cada bit recibido se califica como aceptable, dudoso o inaceptable, dependiendo de la calidad de la señal recibida, con probabilidades 0,9; 0,08 y 0,02, respectivamente. Considerar los primero cuatro bits transmitidos. Sea X el número de bits aceptables, y Y el número de bits dudosos. La calificación de cada bit es independiente.

a) Construir la tabla de distribución de probabilidades conjunta, $p_{X,Y}(x,y)$.

b) ¿Cuál es la probabilidad de que se reciban exactamente dos bits aceptables y uno dudoso? ¿Cuál para dos bits aceptables?

c) Si se han recibido 3 bits aceptables, ¿cuál es la probabilidad de que no se reciban bits dudosos?

d) ¿Cuál es el valor esperado de bits aceptables recibidos? ¿Cuál el de bits dudosos? ¿Cuál de bits inaceptables?

e) Hallar la covarianza y el coeficiente de correlación. ¿Son las variables X y Y independientes?

Ejercicio 6. Sean X y Y variables aleatorias independientes, idénticamente distribuidas, con distribución geométrica de parámetro p. Demostrar que

$$P(X = i \,/\, X + Y = n) = \frac{1}{n-1}, \quad i = 1,2,\cdots,n-1.$$

Ejercicio 7. Sean X y Y dos variables aleatorias discretas con valores esperados $E(X) = \mu_X$ y $E(Y) = \mu_Y$, varianzas $V(X) = (\sigma_X)^2$ y $V(Y) = (\sigma_Y)^2$, respectivamente, distribución de probabilidad conjunta $p_{X,Y}(x,y)$, covarianza $Cov(X,Y)$ y coeficiente de correlación $Corr(X,Y)$, y sean a y b dos constantes reales.

a) Demostrar que:

$$E(aX + bY) = aE(X) + bE(Y)$$

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X,Y)$$

$$Cov(aX + c, bY + d) = 2abCov(X,Y) \quad \forall c, \forall d \in \mathbf{R}$$

$$Corr(aX + c, bY + d) = Corr(X,Y) \text{ cuando } a \text{ y } b \text{ tienen el mismo signo}$$

b) Demostrar que si, además, X y Y resultan independientes, se verifica que:

$$E(XY) = E(X)E(Y), V(aX + by) = a^2V(X) + b^2V(Y), Cov(X,Y) = 0$$

Observación- Las igualdades dadas en los ítems **a)** y **b)**, valen para el caso que las variables X y Y sean continuas con función de densidad de probabilidad conjunta $f_{X,Y}(x,y)$.

Respuestas Guía de Ejercitación 5. Parte I

Ej.1 **a)** $P(X > Y) = 0{,}0028$; $P(X = Y) = 0{,}9710$; $P(X < Y) = 0{,}0262$.

b)

x	0	1	2
$p(x)$	0,84	0,10	0,06

y	0	1	2
$p(y)$	0,8164	0,1160	0,0676

c) $E(X) = 0{,}22$; $E(Y) = 0{,}2512$; $\sigma_X = 0{,}54$; $\sigma_Y = 0{,}5686$.

d) $P(Y \le X/X = 1) = 0{,}9900$; $P(Y \le X/X \le 1) = 0{,}9721$; **e)** $E(X - Y) = -0{,}0312$.

f) No son independientes.

h) $Cov(X,X) = 0{,}2916$; $Cov(Y,Y) = 0{,}3233$; $Cov(Y,X) = 0{,}2799$;
$Corr(X,Y) = 0{,}9116$.

Ej.2 **a)**

x	-2	-1	1	2
$p_X(x)$	0,25	0,25	0,25	0,25

y	1	4
$p_Y(y)$	0,5	0,5

c) $E(X) = 0$; $E(Y) = 2{,}5$; $E(XY) = 0$; **c)** No son independientes ($Y = X^2$).

d) $E(X+Y) = 2{,}5$; $V(X+Y) = 4{,}75$

$x+y$	0	2	6
$p(x+y)$	0,25	0,50	0,25

Ej.3 **b)**

		y				
		0	1	2	3	$P_X(x)$
x	1	$1/4$	$1/12$	0	0	$1/3$
	2	$3/16$	$1/8$	$1/48$	0	$1/3$
	3	$9/64$	$9/64$	$3/64$	$1/192$	$1/3$
	$P_Y(y)$	$37/64$	$67/192$	$13/192$	$1/192$	

c) Son independientes si $P_{XY}(x,y) = P_X(x) \cdot P_Y(y) \quad \forall (x,y)$.

$P_{XY}(3{,}3) \neq P_X(3) \cdot P_Y(3)$ pues $1/192 \neq 1/192 \cdot 1/3 \Rightarrow$ no son independientes.

d) $P_Y(y=0) = 37/64 \cong 0{,}5781$.

e) A: es el evento que indica que aprobó el examen

$P(A) = P_{XY}(x=1, y=0) + P_{XY}(x=2, y \leq 1) + P_{XY}(x=2, y \leq 2) = \frac{57}{64} \cong 0{,}8906$

f) $P_{Y/X=3}\,(x=3, y=2) = P_{XY}(x=3, y=2)/P_X(x=3) = (3/64)/(1/3) \cong 0{,}1406$

g) $P_{X/Y=2}\,(x=3, y=2) = P_{XY}(x=3, y=2)/P_Y(y=2) = (3/64)/(13/92) \cong 0{,}6923$

h) $W = 10\,(X - Y)\,/\,X$; $R_W = \{0;\ 10/3;\ 5;\ 20/3;\ 10\}$

w	0	$10/3$	5	$20/3$	10
$P(w)$	$7/64$	$3/64$	$1/8$	$9/64$	$37/64$

$E(W) = 7{,}5$; $V(W) = 275/24 \cong 11{,}4583 \Rightarrow \sigma_W \cong 3{,}385$

Ej.4 **a)**

		y			
		-3	0	2	$P_X(x)$
x	-2	0,00 ()	0,00 ()	$3/36$ (11, 12)	$3/36$
	0	$3/36$ (2, 3)	$6/36$ (7)	$12/36$ (8, 9, 10)	$21/36$
	3	$7/36$ (4, 5)	$5/36$ (6)	0,00 ()	$12/36$
	$P_Y(y)$	$10/36$	$11/36$	$15/36$	

b) $P(A) = P(X > Y) = 5/12$; $P(B) = P[(X > Y) \cup (Y > X)] = 5/6$;
$P(C) = P(X \leq 0) = 2/3$; $P(D) = P(Y \geq 0) = 13/18$

c) $5/12$; **d)** $23/33$; **e)** sabiendo que el segundo jugador paga dinero, ¿cuál es la probabilidad de que el primer jugador cobre dinero? $7/10$

f) $E(X) = 5/6$; $E(Y) = 0$; **g)** $Cov(X,Y) = -2{,}083$; $Corr(X,Y) = -0{,}628$

Ej.5 **a)**

		x				
		0	1	2	3	4
y	0	$1{,}6 \times 10^{-7}$	$2{,}88 \times 10^{-5}$	$1{,}944 \times 10^{-3}$	0,05832	0,6561
	1	$2{,}66 \times 10^{-6}$	$3{,}456 \times 10^{-4}$	0,015552	0,23328	0
	2	$1{,}536 \times 10^{-5}$	$1{,}3824 \times 10^{-3}$	0,031104	0	0
	3	$4{,}096 \times 10^{-5}$	$1{,}8432 \times 10^{-3}$	0	0	0
	4	$4{,}096 \times 10^{-5}$	0	0	0	0

b) $P(X=2; Y=1) = 0{,}015552$; $P(X=2)\ 0{,}0486$; **c)** $P(Y=0\,/\,X=3) = 0{,}20$

d) $E(X) = 3{,}6$; $E(Y) = 0{,}32$; W: número de bits inaceptables $\Rightarrow E(W) = 0{,}08$

e) $Cov(X,Y) = -0{,}2903$; $Corr(X,Y) = -0{,}8917$; X y Y no son independientes.

Guía de Ejercitación 5. Parte II. Propiedad reproductiva de la distribución normal

Ejercicio 1. Para resolver las siguientes cuestiones, utilizar la **propiedad reproductiva de la distribución normal**. Si $X_1, X_2, \cdots, X_n$son variables aleatorias **normales independientes** con valores esperados $E(X_1) = \mu_1, E(X_2) = \mu_2, \cdots, E(X_n) = \mu_n$, y varianzas $V(X_1) = (\sigma_1)^2$, $V(X_2) = (\sigma_2)^2$, … , $V(X_n) = (\sigma_n)^2$, entonces

$$Y = c_1X_1 + c_2X_2 + \cdots + c_nX_n,$$

donde $c_1, c_2, \cdots, c_n$ son constantes no simultáneamente nulos, es una variable aleatoria normal con valor esperado y varianza

$$E(Y) = c_1E(X_{1)} + c_2E(X_2) + \cdots + c_nE(X_n), V(Y) = c_1^2V(X_{1)} + c_2^2V(X_2) + \cdots + c_n^2V(X_n)$$

<u>Como casos particulares</u>

Para $E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$ y varianzas $V(X_1) = V(X_2) = \cdots = V(X_n) = \sigma^2$,

- Si $c_1 = c_2 = \cdots = c_n = 1$, se tiene
$$T = X_1 + X_2 + \cdots + X_n \text{ con } T \sim N(E(T) = n\mu; V(T) = n\sigma^2);\ \sigma_T = \sqrt{n}\sigma$$
- Si $c_1 = c_2 = \cdots = c_n = 1/n$, se tiene
$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \text{ con } \bar{X} \sim N\left(E(\bar{X}) = \mu; V(\bar{X}) = \frac{\sigma^2}{n}\right);\ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

a) Las variables aleatorias X_1 y X_2 denotan la longitud y el ancho, respectivamente, de una pieza rectangular (suponer que los lados paralelos tienen igual longitud). Si X_1 tiene una distribución normal con media de 2,0 cm y desviación estándar de 0,1 cm, X_2 tiene también una distribución normal pero con media y desvío estándar de 5,0 cm y 0,2 cm, respectivamente, y se supone que ambas variables son independientes. Calcular la probabilidad de que el perímetro de la pieza resulte mayor que 14,5 cm.

b) Una máquina envasadora larga en cada impulsión de líquido una cantidad que es una variable aleatoria con distribución normal de valor esperado k cm^3 y desvío estándar $0{,}04k$ cm^3 con $k > 0$. Indicar si conviene más llenar un envase de un litro con cuatro impulsiones de valor medio de 250 cm^3 o con una sola de 1000 cm^3, si se quiere minimizar la cantidad de envases con menos de 990 cm^3.
Sin hacer nuevos cálculos, ¿se llegaría a la misma conclusión si se quiere minimizar la cantidad de envases con más de 1010 cm^3. Justificar.

c) Una máquina produce tornillos tales que el diámetro de su cuello se distribuye normalmente con media de 3,3mm y desvío estándar 0,05mm. Los tornillos se acoplan con arandelas cuyo diámetro interior se distribuye normalmente con media 3,5mm y desvío estándar 0,05mm. Si se toma un par tornillo-arandela al azar de la producción general, calcular la probabilidad de que el tornillo calce en la arandela.

d) Una máquina automática llena latas de una bebida. El volumen medio de llenado es de 350ml de líquido, y la desviación estándar es de 6ml. Si se supone que el volumen de llenado de las latas se corresponde con variables aleatorias normales independientes, ¿cuál es la probabilidad de que al seleccionar 9 latas, el volumen promedio de este proceso $(\bar{X})$ sea menor que 344ml de líquido? Comparar el resultado con el que el contenido de una lata sea menor a 344ml.

e) Hay unas zapatillas livianas que se expiden en cajas de cartón corrugado de 6 pares, cada par contenido en su caja individual. Frecuentemente los clientes reciben cajas con unidades faltantes, es decir que se encuentran 11 (o menos) zapatillas y reclaman furiosamente al vendedor. Para solucionar el problema, se ha decidido efectuar un control al final de la línea de empaques, pero como obviamente sería ilógico abrir cada caja para verificarla, se aplicará el siguiente procedimiento: se colocará una balanza al final de la línea y se pesarán todas las cajas, abriendo luego aquellas cuyo peso sea sospechoso. Ahora, para implementar este control debe fijarse un peso crítico C, tal que si una caja pesa menos, se la abrirá. A efectos de calcular el valor de C, se establece la condición de detectar al menos el 99% de las cajas con

11 zapatillas, y se sabe que los pesos de las zapatillas y las cajas son variables normales con los siguientes parámetros:

peso individual de las zapatillas en gramos, $X \approx N(\mu_X = 170; \sigma_X^2 = 7^2)$,

peso de las cajas individuales en gramos, $Y \approx N(\mu_Y = 50; \sigma_Y^2 = 5^2)$,

peso de las cajas de cartón corrugado en gramos, $W \approx N(\mu_W = 300; \sigma_W^2 = 40^2)$.

Calcular: ***i***. el valor de C; ***ii.*** el porcentaje de las cajas completas que se revisa inútilmente.

Observaciones: Sean $X_1, X_2, \cdots, X_n$ variables aleatorias **independientes** con valores esperados $E(X_1) = \mu_1, E(X_2) = \mu_2, \cdots, E(X_n) = \mu_n$, y varianzas $V(X_1) = (\sigma_1)^2$, $V(X_2) = (\sigma_2)^2$, … , $V(X_n) = (\sigma_n)^2$. Sea Y la variable aleatoria correspondiente a la combinación lineal

$$Y = c_1X_1 + c_2X_2 + \cdots + c_nX_n,$$

donde $c_1, c_2, \cdots, c_n$ son constantes no simultáneamente nulos.

Siempre son válidas las siguientes expresiones para el valor esperado y la varianza de Y:

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_nE(X_n), V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_n^2V(X_n).$$

También vale para el caso particular que $\mu_1 = \mu_2 = \cdots = \mu_n = \mu$, y que $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$, que si

$T = X_1 + X_2 + \cdots + X_n$ al tomar $c_1 = c_2 = \cdots = c_n = 1 \Rightarrow E(T) = n\mu; V(T) = n\sigma^2$,

$\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ al tomar $c_1 = c_2 = \cdots = c_n = 1/n \Rightarrow E(\bar{X}) = \mu; V(\bar{X}) = \frac{\sigma^2}{n}$

pero, en general, **no se puede asegurar la forma de la distribución** de las variables Y, T y $\bar{X}$.

Ejercicio 2. Sea $\{X_1, X_2, \cdots, X_n\}$ un conjunto de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 comunes. Encontrar los valores de a y b que hace verdadera la siguiente igualdad.

$$E[(X_1 + X_2 + \cdots + X_n)^2] = aE(X_1^2) + b[E(X_1)]^2$$

Justificar todos los pasos de la demostración.

Respuestas Guía de Ejercitación 5. Parte II

Ej.1. a) 0,1314; **b)** convienen cuatro impulsiones; **c)** 0,9977; **d)** 0,00135;
e) ***i***. 2581,5 g; ***ii.*** 11,31%

Ej.2. $a = n; b = n^2 - n$

Guía de Ejercitación 5. Parte III.
Distribución del muestreo. Teorema Central del Límite

Cuando de una población se extrae una muestra aleatoria simple de valores numéricos, se puede considerar que cada unidad en la muestra constituye una variable aleatoria. A menos que la muestra represente una proporción grande (mayor que 5%) de la población, se puede tratar a las unidades en la muestra como independientes. A menos que se indique lo contrario, se supondrá que los valores de una muestra aleatoria simple son variables aleatorias independientes.

Muestra aleatoria $\{X_1, X_2, \cdots, X_n\}$
Conjunto de n variables aleatorias independientes con la misma distribución.

Ejercicio 1. Un vendedor ofrece de un mismo producto, tres calidades diferentes. De todos los clientes que compran un sólo artículo, el 20% compra el de más alta calidad cuyo costo es de \$250, el 30% compra de calidad media cuyo costo es de \$200, y el 50% restante compra el de más baja calidad cuyo costo es de \$160. Sea X el pago efectuado por un cliente seleccionado al azar que se llevan un sólo artículo.

a) Obtener la distribución de probabilidad puntual de X y hallar $E(X) = \mu$ y $V(X) = \sigma^2$.

b) Se considera ahora una muestra aleatoria de tamaño 2, esto es el conjunto de dos variables aleatorias independientes e idénticamente distribuidas, X_1 y X_2, correspondiente a los pagos efectuados por dos cliente. Determinar la distribución de probabilidad puntual del promedio muestral $\bar{X} = (X_1 + X_2)/\ 2$. Calcular $E(\bar{X})$ y comparar con μ.

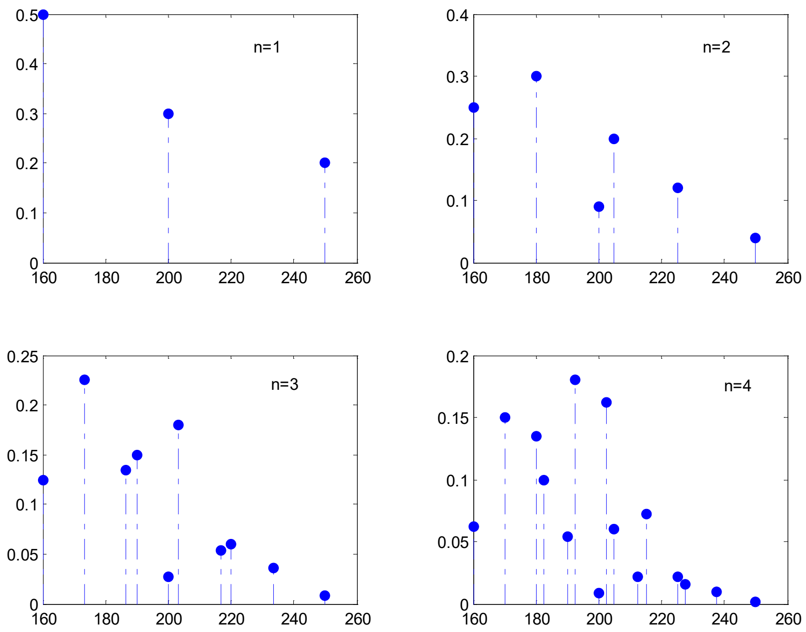
c) Calcular la probabilidad $P(\bar{X} > 210)$.

d) Para la misma muestra aleatoria de tamaño 2, demostrar que la varianza muestral

$$S^2 = \frac{1}{2-1}\sum_{i=1}^{2}(X_i - \bar{X})^2 \text{ es igual a } S^2 = \frac{1}{2}(X_1 - X_2)^2.$$

e) Hallar la distribución de probabilidad puntual de la varianza muestral S^2, calcular $E(S^2)$ y comparar con $V(X) = \sigma^2$.

f) Constatar que los gráficos correspondientes a la distribución de probabilidad del promedio muestral $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$, esto es el promedio de pagos efectuados por n clientes, para $n = 1$, 2, 3 y 4 son los dados a continuación.



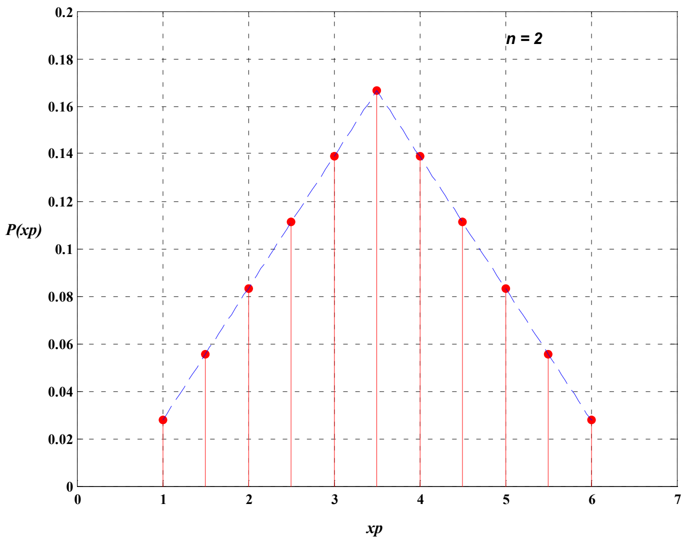
Ejercicio 2. Sea X la variable aleatoria discreta correspondiente al número que sale en la cara superior al arrojar un dado equilibrado.

a) Obtener la distribución de probabilidad puntual de X y hallar $E(X) = \mu$ y $V(X) = \sigma^2$. Calcular $P(3 \leq X \leq 4)$.

b) Considerar la enumeración de todas las muestras posibles de tamaño 2 correspondientes a la variable aleatoria X, esto es el conjunto de las dos variables aleatorias independientes e idénticamente distribuidas, X_1 y X_2, correspondiente a los resultados que se obtienen al tirar dos veces el dado. Determinar la distribución de probabilidad puntual del promedio muestral $\bar{X} = (X_1 + X_2)/2$ y corroborar que la misma está definida por

$$P(\bar{X} = \bar{x}) = \frac{1 - \left|\bar{x} - \frac{7}{2}\right|}{36} \qquad \bar{x} = \frac{k}{2} \wedge k = 2,3,4,\cdots,12$$

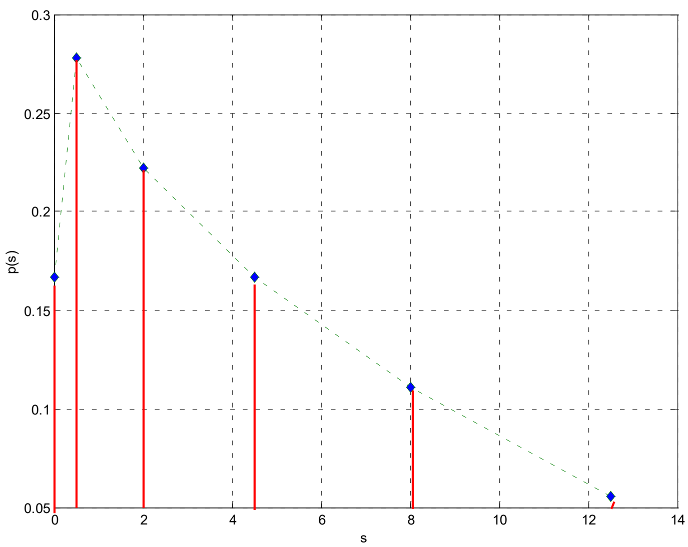
cuya representación gráfica es la dada en la figura a.



Ej.2. Fig a

Hallar $E(\bar{X})$ y comparar con μ. Calcular $P(3 \leq X \leq 4)$.

c) Hallar la distribución de probabilidad puntual de la varianza muestral S^2, calcular $E(S^2)$ y comparar con $V(X) = \sigma^2$. Constatar que la representación gráfica es la dada en la figura b.

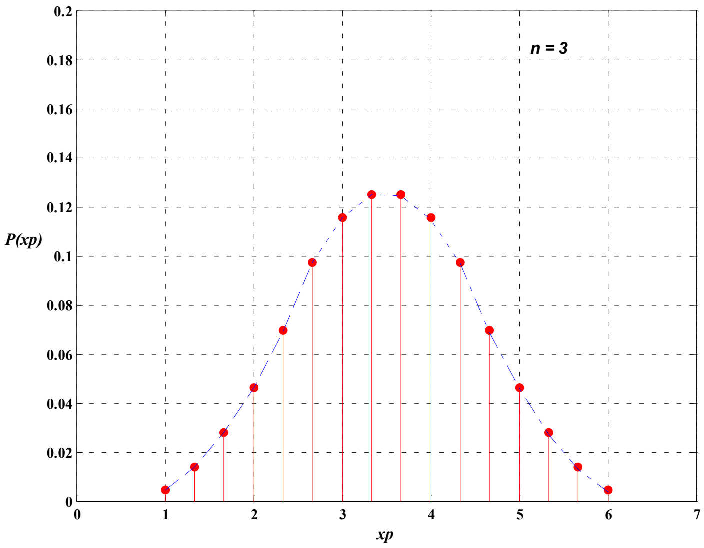


Ej.2. Fig b

d) Considerar la enumeración de todas las muestras posibles de tamaño 3 correspondientes a la variable aleatoria X, esto es el conjunto de las dos variables aleatorias independientes e idénticamente distribuidas, X_1,X_2 y X_3 correspondiente a los resultados que se obtienen al tirar tres veces el dado. Determinar la distribución de probabilidad puntual del promedio muestral $\bar{X} = (X_1 + X_2 + X_3)/3$ y corroborar que la misma está definida por

$$\begin{cases} \dfrac{9}{432}\bar{x}^2 - \dfrac{9}{432}\bar{x} + \dfrac{1}{216} & 1 \le \bar{x} < 3 \\ -\dfrac{9}{216}\bar{x}^2 + \dfrac{63}{216}\bar{x} - \dfrac{83}{216} & 3 \le \bar{x} < 4 \text{ para } \bar{x} \in R_{\bar{X}}, \\ \dfrac{9}{432}\bar{x}^2 + \dfrac{111}{432}\bar{x} - \dfrac{171}{216} & 4 \le \bar{x} \le 6 \end{cases}$$

cuya representación gráfica es la dada en la figura c.



Ej.2. Fig c

Hallar $E(\bar{X})$ y comparar con μ. Calcular $P(3 \le X \le 4)$.

Ejercicio 3. Sea $X_1, X_2, \cdots, X_n$ una muestra aleatoria (**conjunto de variables aleatorias idénticamente distribuidas e independientes**) de una variable aleatoria X con valor esperado μ y varianza σ^2. Demostrar que la variable aleatoria varianza muestral definida por

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

donde $\bar{X}$ es el promedio muestral, satisface que $E(S^2) = \sigma^2$.

Ejercicio 4. **Transporte en camiones.** El peso X, en kg, de las cajas transportadas por la Camionera Fierrorrap se distribuye de manera **aproximadamente normal** con $E(X) = \mu = 20$ kg y $V(X) = \sigma^2 = 9$kg^2.

Calcular la probabilidad de que:

- **a)** el peso de una caja tomada al azar esté comprendido entre 19,7kg y 20,6kg.
- **b)** El peso medio de una muestra de 100 cajas, $\bar{X} = (X_1 + X_2 + \cdots + X_{100})/100$, esté comprendido entre 19,7 kg y 20,6 kg.
- **c)** Un camión que transporta 100 cajas sea multado si se conoce que dicha sanción se aplica si el peso neto de su carga, $T = X_1 + X_2 + \cdots + X_{100}$, excede los 2060 kg.
- **d)** Determinar un intervalo simétrico con respecto a µ que abarque el peso medio de 100 de estas cajas con una probabilidad de 0,95. Ayuda: determinar el número real positivo a de forma que $P(\mu - a < \bar{X} < \mu + a) = 0{,}95$.

Ejercicio 5. Suma y Teorema Central del Límite.

Aplicar el Teorema Central del Límite a las siguientes situaciones correspondiente a la suma de variables aleatorias independientes con igual distribución.

a) El depósito de un avión se carga con 100 paquetes cuyos pesos son variables aleatorias independientes que están uniformemente distribuidas entre 5 y 50 kilos. ¿Cuál es la probabilidad de que el peso total de la carga exceda los 3000 kilos?

b) Treinta y seis instrumentos electrónicos se usan de la siguiente manera: tan pronto falla el primero empieza a actuar el segundo, cuando éste falla empieza a actuar el tercero y así sucesivamente. Si la variable aleatoria asociada a la duración de cada instrumento, X_i, tiene una distribución exponencial de parámetro 0,1 h^{-1}, cada instrumento tiene un tiempo de operación independiente de los demás, y T es el tiempo total de operación de los 36 instrumentos, ¿cuál es la probabilidad de que T exceda las 370 hs? Justificar el planteo del cálculo.

c) Sea $\{X_1, X_2, \cdots, X_{50}\}$ un conjunto de variables aleatorias independientes que tienen cada una distribución Poisson con parámetro $\lambda = 0{,}3$, y sea $T_{50} = X_1 + X_2 + \cdots + X_{50}$ su suma. Evaluar usando el Teorema Central del Límite $P(T_{50} \geq 18)$.

d) **Tiempo de procesado de partes y cantidad de unidades.** Una máquina procesa partes, una a la vez. El tiempo de procesado de las diferentes partes son variables aleatorias independientes uniformemente distribuidas en el intervalo entre 1 y 5 minutos. Calcular la probabilidad de que el número de partes procesadas en un intervalo de 320 minutos sea al menos 100.

e) **Número de clientes atendidos en una caja.** Los tiempos de servicio para los clientes que llegan a una caja de supermercado son variables aleatorias independientes con un promedio de 1,5 minutos y una dispersión de 1 minuto.

  i. Calcular la probabilidad aproximada de que se pueda atender a 100 clientes en menos de 2 horas de tiempo total de servicio de esta caja.

  ii. Calcular el número de clientes n tal que la probabilidad de dar servicio a todos en menos de 2 horas sea aproximadamente 0,1.

Ejercicio 6. Aproximación normal de una variable aleatoria binomial.

Una variable aleatoria binomial S_n con parámetros n y p se puede ver como la suma de n variables aleatorias independientes con distribución Bernoulli, $\{X_1, X_2, \cdots, X_n\}$, con parámetro común p:

$$S_n = X_1 + X_2 + \cdots + X_n; E(X_i) = p; V(X_i = p(1-p) \Rightarrow E(S_n)np; V(S_n) = np(1-p).$$

Si n es lo suficientemente grande y el histograma de probabilidad binomial no está demasiado sesgado ($np \geq 10$ y $n(1–p) \geq 10$), entoncesS_n tiene distribución aproximadamente normal con valor medio np y varianza $np(1–p)$.

A los efectos prácticos, se considera la conocida corrección de medio punto de modo que

$$P(S_n \leq a) = B(a; n, p) \cong P(S_n|_{NOR} \leq a + 0{,}5) = P\left(Z \leq \frac{a + 0{,}5 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{a + 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(a \leq S_n \leq b) \cong P(a - 0{,}5 \leq S_n|_{NOR} \leq b + 0{,}5) = \Phi\left(\frac{b + 0{,}5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - 0{,}5 - np}{\sqrt{np(1-p)}}\right)$$

a) Sea X una variable aleatoria binomial con parámetros $n = 36$ y $p = 0{,}5$. Calcular el valor de $P(X \leq 21)$ y comparar con él, el valor aproximado utilizando el teorema central del límite con y sin la corrección de medio punto.

b) Para la misma variable del ítem anterior, comparar $P(X = 19)$ con el valor aproximado utilizando el teorema central del límite dado por

$$P(X=19) \cong P(19-0{,}5 \leq X|_{NOR} \leq 19+0{,}5) = \Phi\left(\frac{19{,}5-18}{3}\right) - \Phi\left(\frac{18{,}5-18}{3}\right).$$

Ejercicio 7 Promedio muestral y Teorema Central del Límite.

a) Sea $\{X_1, X_2, \cdots, X_{40}\}$ una muestra aleatoria de mediciones de la proporción de impurezas en muestras de una aleación. Cada una de las variables aleatorias X_k tiene una densidad de probabilidad definida por

$$f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{en otro caso} \end{cases}.$$

Un comprador potencial rechaza una partida de mineral si la media de la proporción de impurezas de una muestra de tamaño 40 es superior a 0,8. Calcular la probabilidad de que la partida sea rechazada. Comentar claramente el planteo del problema.

b) Una máquina automática de llenado de botellas tiene una varianza en las cantidades de llenado de $\sigma^2 = 1$ onza^2. Sin embargo, el promedio de las onzas de llenado μ depende de un ajuste diario y de operador a operador. Si en un día se llevan a cabo 36 observaciones de la cantidad (en onzas) del líquido de llenado y con el mismo ajuste de máquina, calcular:

i. la probabilidad de que el promedio muestral quede dentro de 0,3 onzas de diferencia con respecto al promedio verdadero de la población, para ese ajuste.

ii. ¿Cuál debe ser el tamaño de la muestra n, para que $\bar{X}$ quede a menos de 0,3 onzas de μ con una probabilidad de 0,95?

c) En la fabricación de cierto tipo de cojinetes para motor se sabe que el diámetro promedio es de 5 cm con un desvío estándar de 0,005 cm. El proceso es controlado en forma periódica mediante la selección aleatoria de 64 cojinetes, midiendo sus correspondientes diámetros. El proceso se supone bajo control si la media muestral se encuentra entre dos límites especificados en el 95% de las veces que se extraen las muestras para realizar el control del proceso. Determinar los valores de esos límites si son simétricos respecto de la media poblacional.

Ejercicio 8. Se sabe que el monto de las ventas realizadas por una empresa tiene distribución normal con media \$343200 y desvío típico \$48152. De las 1815 ventas realizadas en el mes se saca una muestra de 16 facturas.

a) ¿Cuál es la probabilidad de que la media muestral difiera de la media poblacional en más de \$20000?

b) ¿Cuál es el valor de la media muestral que es superado con probabilidad 0,05?

Ejercicio 9. Se considera un evento A definido en el contexto de algún experimento aleatorio. Sea $p = P(A)$ la probabilidad de que el evento ocurra. Si se consideran n repeticiones independientes, se puede definir la **frecuencia empírica de ocurrencia de *A*** como

$$f_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

donde X_i es 1 si ocurre A en la repetición i-ésima, y es 0 en caso contrario, con $E(X_i) = p$. Así, $E(f_n) = p$ y $V(f_n) = p(1-p)/n$.

i. Una fábrica produce determinados artículos de tal manera que el 5% resulta defectuoso. Se inspeccionan n artículos y se determina la frecuencia empírica f_n de defectuosos entre los seleccionados. ¿Cuál debe ser el tamaño n si se desea que f_n difiera de la verdadera proporción $p = 0{,}05$ de defectuosos en menos de 0,01 (1 punto porcentual) con una probabilidad no menor a 0,95, esto es $P(|f_n - 0{,}05| < 0{,}01) \geq 0{,}95$?

ii. Si en el ítem anterior no se conociera el porcentaje p de defectuosos en lo producido, para evaluar el tamaño n de la muestra se utiliza que la dispersión será máxima cuando la función $p(1–p)$ tenga un máximo, esto es para $p = 1/2$. A partir de esto, evaluar n para que $P(|f_n - p| < 0{,}01) \geq 0{,}95$.

Ejercicio 10. Diferencias de medias muestrales

a) Se toman muestras independientes de tamaño 10 y 15 de una variable aleatoria distribuida normalmente con valor esperado 10 y varianza 3. ¿Cuál es la probabilidad de que el promedio de las dos muestras se diferencie (en valor absoluto) en más de 0,3?

b) La vida útil en horas de un componente utilizado en la turbina de una aeronave es una variable aleatoria X con media $\mu_X = 5000$ y desviación estándar $\sigma_X = 40$, siguiendo una distribución normal. Se produce una mejora en el diseño del componente que modifica el tiempo de vida útil, variable aleatoria Y con distribución normal, pero de mayor valor medio, siendo ahora $\mu_Y = 50500$ y disminuyen la desviación estándar a $\sigma_Y = 30$. Suponiendo que el primer proceso utiliza una muestra aleatoria de $n_X = 16$ componentes y el mejorado una muestra aleatoria de $n_Y = 25$ elementos, ¿cuál es la probabilidad de que la diferencia de las dos medias muestrales $(\bar{X} - \bar{Y})$ sea al menos de 25 horas (para ser significativa)?

Respuestas Guía de Ejercitación 5. Parte III

Ej. 1. **a)** $E(X) = 190;\ V(X) = 1200.$

x	160	200	250
$p(x)$	0,50	0,30	0,20

b) $n = 2: E(\bar{X}) = 190;\ V(\bar{X}) = 600$

$\bar{x}$	160	180	200	205	225	250
$p(\bar{x})$	0,25	0,30	0,09	0,20	0,12	0,04

c) 0,16

e) $E(S^2) = 1200 = V(X)$

s^2	0		800	1250	4050
$p(s^2)$	0,38		0,30	0,12	0,20

f) $n = 3: E(\bar{X}) = E(X) = 190;\ V(\bar{X}) = 400 = \frac{V(X)}{3} = \frac{1}{3}\sigma^2; \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{3}}$

$\bar{x}$	160	173,33	186,67	190	200	203,33	216,67	220	233,33	250
$p(\bar{x})$	0,125	0,225	0,135	0,150	0,027	0,180	0,054	0,060	0,036	0,008

$n = 4: E(\bar{X}) = E(X) = 190;\ V(\bar{X}) = 300 = \frac{V(X)}{4} = \frac{1}{4}\sigma^2; \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{4}}$

$\bar{x}$	160	170	180	182,5	190	192,5	200	202,5
$p(\bar{x})$	0,0625	0,1500	0,1350	0,1000	0,0540	0,1800	0,0081	0,1080

$\bar{x}$	205	212,5	215	225	227,5	237,5	250
$p(\bar{x})$	0,0600	0,0216	0,0720	0,0216	0,0160	0,0096	0,0016

Ej. 2. **a)** $E(X) = 7/2;\ V(X) = 35/12;\ P(3 \leq X \leq 4) = 2/6 \cong 0{,}3333$

x	1	2	3	4	5	6
$p(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

b) $n = 2 \colon E(\bar{X}) = 7/2;\ V(\bar{X}) = 35/24;\ P(3 \leq \bar{X} \leq 4) = 12/36 \cong 0{,}4444$

$\bar{x}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	4	$\frac{9}{2}$	5	$\frac{11}{2}$	6
$p(\bar{x})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

c) $E(S^2) = 35/12 = V(X) = \sigma^2$

s^2	0	$\frac{1}{2}$	$\frac{4}{2}$	$\frac{9}{2}$	$\frac{16}{2}$	$\frac{25}{2}$
$p(s^2)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

d) $n = 3 \colon E(\bar{X}) = 7/2;\ V(\bar{X}) = 35/36;\ P(3 \leq \bar{X} \leq 4) = 104/216 \cong 0{,}4815$

$\bar{x}$	1	$\frac{4}{3}$	$\frac{5}{3}$	2	$\frac{7}{3}$	$\frac{8}{3}$	3	$\frac{10}{3}$	$\frac{11}{3}$	4	$\frac{13}{3}$	$\frac{14}{3}$	5	$\frac{16}{3}$	$\frac{17}{3}$	6
$p(\bar{x})$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$

Ej. 3. Se puede descargar la demostración del material en la página PyE del CVG

Ej. 4. **a)** $0{,}1191$; **b)** $0{,}8185$; **c)** $0{,}0228$; **d)** $(19{,}412;\ 20{,}588)$ en kg.

Ej. 5. **a)** $\cong 0{,}0274$; **b)** $\cong 0{,}4325$; **c)** $\cong 0{,}219$ (T.C.L.); **d)** $\cong 0{,}9582$; **e) i.** $\cong 0{,}0013$; **ii.** $n = 88$

Ej. 6 **a)** En Excel. DISTR.BINOM(21,36,0.5,VERDADERO)=0,87851;
con aprox. normal: 0,84134 sin corrección de medio punto; 0,87833 con corrección de medio punto. Igual respuesta trabajando con la App Probability Distribution
b) En Excel. DISTR.BINOM(19,36,0.5,FALSO)=0,12511; aprox. normal 0,12528

Ej. 7. **a)** $0{,}0516$; **b) i.** $0{,}9282$; **ii.** 43; **c)** bajo control si $\bar{X} \in (4{,}998775; 5{,}001225)$

Ej. 8. **a)** $0{,}0970$; **b)** $\$363002{,}51$.

Ej. 9. **a)** 1825; **b)** 9604.

Ej. 10 **a)** $0{,}6708$;

$$\mathbf{b)}\ P(\bar{Y} - \bar{X} > 25) 1 - P\left(\frac{(\bar{Y} - \bar{X}) - (5050 - 5000)}{\sqrt{\frac{40^2}{16} + \frac{30^2}{25}}} \leq \frac{25 - 50}{\sqrt{\frac{40^2}{16} + \frac{30^2}{25}}} \right) \cong 0{,}9840$$

Guía Ejercitación 5. Material Adicional

Lectura y comprensión

Desigualdad de Tchebychev. Pafnuti Livovic Tchebychev, 1821-1894, fue uno de los primeros matemáticos del siglo XIX que intentó dar rigor al estudio de la Probabilidad. Dentro de los trabajos que realizó, uno de los más conocidos dentro del área es la desigualdad que lleva su nombre. Con ella se puede estimar cotas para probabilidades de eventos asociados con variables aleatorias a partir del conocimiento de su valor esperado y su varianza, sin conocer las características de la distribución de dicha variable incluyendo el no tener que tener en cuenta si es discreta o continua.

Esta desigualdad establece que si X es una variable aleatoria con valor esperado μ y varianza σ^2, cualquiera sea la distribución, se verifica que

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}, \quad \forall c > 0.$$

Si $c = k\sigma$, la expresión es

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad \forall k > 0 \text{ ó } P(|X-\mu| < k\sigma) \geq 1-\frac{1}{k^2}, \quad \forall k > 0.$$

Si se toma una muestra aleatoria de X, $\{X_1, X_2, \cdots, X_n\}$, se puede expresar la desigualdad para las variables aleatorias suma $T = X_1 + X_2 + \cdots + X_n$ y promedio muestral $\bar{X} = \frac{X_1+X_2+\cdots+X_n}{n}$, obteniéndose, respectivamente,

$$P\left(|T-n\mu| \geq k\sqrt{n}\sigma\right) \leq \frac{1}{k^2}, \quad \forall k > 0 \text{ ó } P\left(|T-n\mu| < k\sqrt{n}\sigma\right) \geq 1-\frac{1}{k^2}, \quad \forall k > 0;$$

$$P\left(|\bar{X}-\mu| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}, \quad \forall k > 0 \text{ ó } P\left(|\bar{X}-\mu| < k\frac{\sigma}{\sqrt{n}}\right) \geq 1-\frac{1}{k^2}, \quad \forall k > 0.$$

Si se considera un evento A definido en el contexto de algún experimento aleatorio, con la probabilidad de que el evento ocurra indicada con $p = P(A)$, y se realizan n repeticiones independientes, se puede definir la frecuencia empírica de ocurrencia de A como

$$f_n = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

donde X_i es 1 si ocurre A en la repetición i-ésima, y es 0 en caso contrario con $E(X_i) = p$. La desigualdad de Tchebyshev se puede expresar

$$P(|f_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}, \quad \epsilon > 0.$$

Esta última expresión tiene una cota superior dada por

$$P(|f_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}, \quad \epsilon > 0.$$

Ejemplos de aplicación.

Ejemplo 1. Un cañón efectúa disparos, siendo la probabilidad de dar en el blanco igual a 0,8 (se supone constante y los disparos independientes). Se efectúan 100 disparos y se desea estimar con la desigualdad de Tchebychev la probabilidad de que el número de blancos X satisfaga $70 \leq X \leq 90$.

$$X \sim Bin\,(n = 100;\ p = 0{,}8);\ E(X) = 80;\ V(X) = 16 \Rightarrow \sigma = 4$$

Usando la desigualdad de Tchebychev, en la forma $P(|X-\mu| < k\sigma) \geq 1-\frac{1}{k^2}, k > 0$:

$$P(70 \leq X \leq 90) = P(|X-80| \leq 2{,}5 \times 4) \geq 1-\frac{1}{2{,}5^2} = 0{,}84.$$

Se puede comparar este resultado con el cálculo usando la probabilidad acumulada de la distribución binomial usando las funciones del Excel:

=DISTR.BINOM(90,100,0.8,VERDADERO)–DISTR.BINOM(69,100,0.8,VERDADERO)

$P(70 \leq X \leq 90) = 0{,}99766644 - 0{,}00605934 = 0{,}9916071.$

O bien, utilizar la aproximación de la distribución a normal, con el siguiente resultado:

$P(70 \leq X \leq 90) \cong 0{,}99565.$

Ejemplo 2. Si se comparan las $P(|X-\mu| < \sigma)$, $P(|X-\mu| < 2\sigma)$ y $P(|X-\mu| < 3\sigma)$ calculadas a partir la desigualdad de Tchebychev en forma general con las correspondientes a considerar una variable aleatoria X con distribución normal, se obtiene el siguiente cuadro:

	$P(\|X-\mu\| < \sigma)$	$P(\|X-\mu\| < 2\sigma)$	$P(\|X-\mu\| < 3\sigma)$
Des. Tchebychev	≥ 0	$\geq 0{,}75$	$\geq 0{,}8889$
Normal	0,6826	0,9544	0,9974

Ejemplo 3. Al sumar números, una computadora aproxima cada número al entero más próximo. Se supone que todos los errores de aproximación son variables aleatorias X_i independientes con distribución uniforme en el intervalo $(-0{,}5, 0{,}5)$. Si se estima con la expresión

$$P(|T - n\mu| \geq c) \leq \frac{n\sigma^2}{c^2}, \forall c > 0,$$

la probabilidad de que cuando se sumen 1500 números, el valor absoluto del error total T sea superior a 15, se obtiene que

$$P(|T| \geq 15) \leq 0{,}5555.$$

Ejemplo 4. Una fábrica produce determinados artículos de tal manera que el 5% resulta defectuoso. Se inspeccionan n artículos y se determina la frecuencia empírica f_n de defectuosos entre los seleccionados. Si se evalúa a partir de la desigualdad de Tchebychev cuál debe ser el tamaño n, para que f_n difiera de la verdadera proporción $p = 0{,}05$ de defectuosos en menos de 0,01 (1 punto porcentual) con una probabilidad no menor a 0,95, esto es

$$P(|f_n - 0{,}05| < 0{,}01) \geq 0{,}95,$$

el resultado que se obtiene es $n \geq 9500$. En condiciones de total incertidumbre sobre el valor de p, $P(|f_n - p| < 0{,}01) \geq 0{,}95$, conduce a un tamaño muestral $n \geq 50000$. Ambos tamaños muestrales corresponden a sobreestimaciones del valor buscado.

======================

Simulaciones. Generación de números aleatorios y su implementación computacional.

El desarrollo de los sistemas de cómputos ha convertido a los experimentos de simulación en técnicas útiles para el análisis de sistemas complejos. En la simulación de estos sistemas surge la necesidad de simular fenómenos aleatorios que le son característicos. Por ejemplo, si un negocio desea estudiar su política de servicios al cliente, puede simular el flujo de clientes como así también el tiempo necesario de atención a cada uno de ellos. Si una empresa desea examinar su política de mantenimiento de planta, puede simular el tiempo entre fallas de las componentes como así también el tiempo necesario para la reparación de cada una de ellas. Estos sucesos constituyen eventos aleatorios.

Para encarar este tipo de problemas se supone, en general, una distribución de probabilidad adecuada para cada fenómeno y se genera una secuencia de valores para la correspondiente variable aleatoria por computadora. La función de distribución uniforme en el intervalo $(0; 1)$ juega un papel importante en la generación de números aleatorios por computadora. Los sistemas de cómputos tienen en su estructura la capacidad de generar valores aleatorios que responden a este tipo de distribución. A partir de ellos se puede generar números aleatorios con otras distribuciones de probabilidad específica, como se muestra en los siguientes ejemplos. Algunos *software* tiene incorporada esta posibilidad automáticamente, sin necesidad de conversiones a una función de distribución uniforme.

Ejemplo 1. Sea T una variable aleatoria continua que represente el tiempo entre dos fallas sucesivas de un dispositivo que sigue un proceso Poisson y se desea simular. La función de densidad de probabilidad de T es exponencial con parámetro b, y su correspondiente función acumulada de probabilidad son

$$f(t) = \begin{cases} be^{-bt} & t \geq 0 \\ 0 & t < 0 \end{cases} \text{con } b > 0 \qquad \Bigg| \qquad F(t) = \begin{cases} 1 - e^{-bt} & t \geq 0 \\ 0 & t < 0 \end{cases} \text{con } b > 0$$

A partir de la función *RAND* de calculadoras, o algún paquete estadístico para computadora, se puede obtener un número aleatorio entre 0 y 1 (que responde a una distribución uniforme entre dichos valores). Sea a el número sorteado. Este número se puede hacer corresponder con un valor de t, asociándolo con un valor de la función acumulada de distribución $F(t)$ y despejando en la forma

$$a = 1 - e^{-bt} \Rightarrow t = \frac{-\ln(1-a)}{b}.$$

Así, por ejemplo, si $a = 0{,}5965, t = 90{,}7579$.

Ejemplo 2. Sea X una variable aleatoria discreta que represente las ventas diarias de un negocio. De los datos históricos se obtiene la siguiente distribución de probabilidad.

x	1	2	3	4
$p(x)$	0,20	0,40	0,30	0,10

Queda definida la función de distribución acumulada de la forma,

$$F(x) = \begin{cases} 0 & x < 1 \\ 0{,}20 & 1 \leq x < 2 \\ 0{,}60 & 2 \leq x < 3. \\ 0{,}90 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

Con un número aleatorio a obtenido con la función *RAND* de la calculadora, o de algún paquete estadístico, se puede simular el número de ventas de un día en el negocio. Este número sorteado se puede hacer corresponder con un valor de la función de distribución acumulada de la variable ventas de un día X. Como puede observarse, quedan definidos los siguientes valores críticos $\{0; 0{,}20;\ 0{,}60;\ 0{,}90; 1\}$. Así, si $a \in [0;\ 0{,}20)$, entonces no se produce ninguna venta en el día, si $a \in [0{,}20;\ 0{,}60)$ entonces se produce 1 venta en el día, y así sucesivamente.

Ejemplo 3. Sea Z una variable aleatoria con distribución normal estandarizada, media 0 y desvío estándar 1. La función de distribución acumulada que le corresponde es

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt.$$

Se podría asociar esta expresión con un número aleatorio a obtenido con la función *RAND* de la calculadora, como en los ejemplos anteriores. Sin embargo, la resolución de $F(x) = a$ no puede resolverse en forma cerrada para x (como en el ejemplo 1), esto es no existe forma analítica de despejar x en función de a. Una alternativa es recurrir a las tablas numéricas de $F(x)$. Otra alternativa es utilizar el hecho demostrable que si U_1 y U_2 son dos variables aleatorias independientes con distribución uniforme sobre el intervalo unitario, entonces

$$Z_1 = (-2\ \ln U_1)^{1/2}\ \mathrm{sen}(2\pi\ U_2) \text{ y } Z_2 = (-2\ \ln U_1)^{1/2}\ \cos(2\pi\ U_2)$$

son dos variables aleatorias normales estandarizadas e independientes. Cabe señalar que, con respecto a la distribución normal, muchos son los software que tienen incorporada la posibilidad de su generación en forma automatizada.

El equivalente a *RAND* en las calculadoras en el programa Excel es *ALEATORIO*(). El valor de la celda de esta función se refresca con la tecla F9. Para generar números aleatorios con otra distribución se puede explorar siguiendo la ruta Análisis de datos, Herramientas, Generación de números aleatorios.