CAPITULO III

MEDIDAS DE POSICIÓN, DISPERSIÓN Y DE ASIMETRÍA

3-1 CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA

Veremos algunas medidas descriptivas que pueden emplearse para caracterizar y describir la forma de la distribución de probabilidad de la variable aleatoria X y especificarla si todas son conocidas.

Es interesante utilizar unos pocos parámetros sencillos para localizar a la distribución en el plano euclídeo. Una vez caracterizada la distribución a partir de los parámetros de localización, interesa estudiar cómo se esparce la distribución en torno a los mismos. Para ello se definen los parámetros de dispersión o concentración. Sin olvidar la forma de la distribución, analizando simetría y kurtosis.

3-1-1 ESPERANZA MATEMÁTICA O VALOR ESPERADO

La esperanza de una variable aleatoria tiene sus orígenes en los juegos de azar, debido a que los apostadores deseaban saber cuál era su esperanza de ganar repetidamente un juego. En este sentido, el valor esperado representa la cantidad de dinero promedio que el jugador está dispuesto a ganar o perder después de un número muy grande de apuestas. Este significado también es válido para una variable aleatoria. Es decir, la esperanza representa el valor promedio de una variable aleatoria después de realizar repetidamente el experimento un gran número de veces. El valor esperado, da idea del centro de equilibrio o de tendencia central de la distribución de los valores de los valores de la variable aleatoria X.

3-1-1-1 DEFINICIÓN DE ESPERANZA MATEMÁTICA O VALOR ESPERADO

- ***Esperanza matemática para una variable aleatoria discreta:***

Sea X una variable aleatoria discreta definida en Ω, sea el espacio de probabilidad $(\Omega, \mathscr{P}(\Omega), P)$ y X(Ω) el conjunto de los valores de X; la esperanza de la variable aleatoria X es:

$$E(X) = \sum_{x_j \in X(\Omega)} x_j f_X(x_j) = \mu \qquad (3.1.1)$$

Donde f_X es la función de probabilidad de X.

- ***Esperanza matemática para una variable aleatoria continua:***

Sea X una variable aleatoria continua, con función de densidad f_X, la esperanza de la variable aleatoria X es:

$$E(X) = \int_{-\infty}^{\infty} x.f_X dx \qquad (3.1.2)$$

Ejemplo 3-1:
Considere el ejemplo 2-2, el experimento aleatorio que consiste en seleccionar dos componentes electrónicos de un paquete, sin reemplazo, probarlos y observar si cada uno de ellos es defectuoso o no. El paquete contiene 6 componentes, de los cuales hay 2 defectuosos. La variable aleatoria en estudio es: X: Número de componentes electrónicos defectuosos obtenidos en las dos extracciones.
En el ejemplo 2-5 se obtuvo su función de probabilidad.
Siendo su función de probabilidad:

X	0	1	2
f_X	$\frac{12}{30}$	$\frac{16}{30}$	$\frac{2}{30}$

Una situación que puede interesar es conocer ¿Cuántos componentes electrónicos defectuosos se espera en promedio encontrar en dos extracciones? Contestar a dicho interrogante es plantearnos sobre la esperanza matemática o valor esperado de la variable en estudio, en este caso X: “Cantidad de componentes defectuosos obtenidos en dos extracciones”

$$E(X) = \sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) = \mu$$

$$E(X) = 0.f(0) + 1.f(1) + 2.f(2)$$

$$E(X) = 0.\frac{12}{30} + 1.\frac{16}{30} + 2.\frac{2}{30} = \frac{16}{30} + \frac{4}{30} = \frac{20}{30} \cong 0.67$$

- Se puede observar que la esperanza E(X) no es necesariamente igual a uno de los valores posibles de X, de esta forma es completamente posible que una variable aleatoria nunca tome el valor de su esperanza. E(X) es un valor promedio y no necesariamente un posible resultado del experimento. Como nos muestra este ejemplo es imposible que obtengamos 0,67 componentes electrónicos defectuosos, o se obtiene uno o dos componente electrónico defectuoso, o ninguno.
- ¿Cómo lo interpretamos? Realizando el experimente varias veces, se espera encontrar en promedio de 0,67 componentes electrónicos defectuosos.
- Si X puede tomar únicamente un número finito de valores distintos, como en el ejemplo 2-8, entonces existe únicamente un número finito de términos en la suma de la ecuación (3.1.1). Sin embargo, si existe una sucesión infinita de valores posibles distintos de X, entonces la ecuación (3.1.1) es una serie infinita de términos. Tal serie puede no converger para una función de probabilidad concreta. Se dice que la esperanza E(X) existe si, y sólo si, la

suma de la ecuación (3.1.1) es absolutamente convergente, esto es, si y sólo si,

$$\sum_x |x|. f_X(x) < \infty \qquad (3.1.3)$$

En otras palabras, si se verifica la relación (3.1.3), entonces E(X) existe y su valor está dado por la ecuación (3.1.1). Si no se verifica la relación (3.1.3), entonces E(X) no existe.

Si X sea una variable aleatoria continua se dice que la E(X) existe si y solo si la integral de la ecuación (3.1.4) es absolutamente convergente.

$$\int_{-\infty}^{\infty} |x|. f_X(x) < \infty \qquad (3.1.4)$$

- La esperanza de una variable aleatoria se puede considerar como el centro de gravedad de los puntos de la variable con una masa asociada proporcional a la función de probabilidad de ellos. centro de masa de esa distribución.
- A partir de considerar a la esperanza como el centro de gravedad, se puede observar una desventaja en esta medida. La esperanza puede resultar enormemente afectada por la presencia de masas pequeñas en comparación con el resto de la distribución, pero que están situadas alejadas del grueso de las mismas.
- Otro significado proviene de la aplicación de la teoría de la probabilidad a los juegos de azar. Dado un juego y un espacio de probabilidad (Ω, $\mathscr{P}(\Omega)$, P) asociado, si se define la variable aleatoria X como la ganancia o pérdida que implica cada resultado del juego, la esperanza matemática de X, E(X), es la ganancia media por jugada. Un juego es equitativo si tiene esperanza nula.
- Supóngase que la función de probabilidad de una distribución es simétrica con respecto a un punto concreto x_0 en el eje x. Es decir que se cumple $f(x_0 + \delta) = f(x_0 - \delta)$, para todos los valores de δ. Supóngase que también existe la esperanza, E(X), de esta distribución. De acuerdo a la interpretación de que la esperanza es el centro de gravedad, resulta que E(X) debe ser igual a x_0, que es el centro de simetría. Es necesario asegurarse de la existencia de la esperanza antes de que se pueda concluir que $E(X) = x_0$.
- Se concluye que la esperanza matemática es una medida que caracteriza altamente a la distribución, sin embargo presenta algunas desventajas: a) no siempre está definida, b) puede resultar muy afectada por un pequeño cambio en la masa de probabilidad asignada a un valor extremo de la distribución. c) no es conveniente su utilización cuando la distribución presenta valores atípicos.

Ejemplo 3-2:

Analice si la siguiente variable aleatoria continua X, tiene esperanza. La variable X tiene una distribución de Cauchy, cuya función de densidad está dada por:

$$f_X(x) = \frac{1}{\pi(1+x^2)} \quad para - \infty < x < \infty \quad (3.1.5)$$

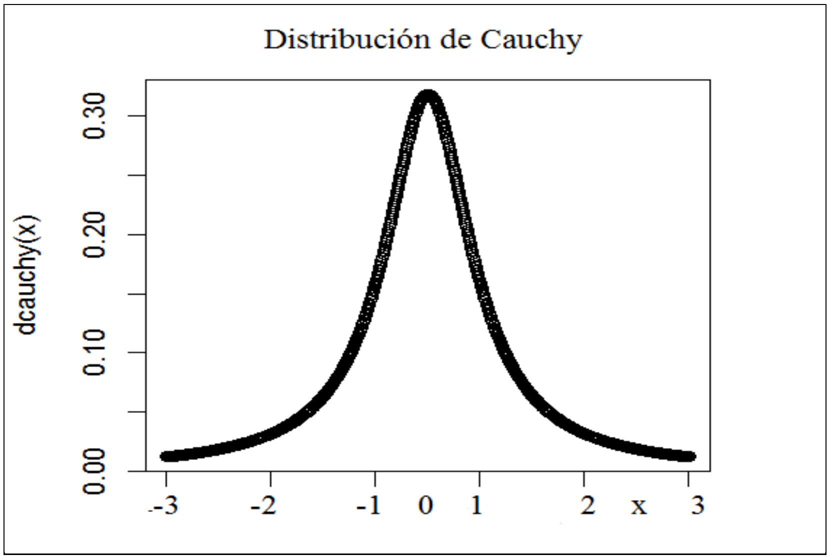


Figura 3.1 función de densidad

La función de densidad dada por la ecuación (3.1.5), se representa en la figura 3.1.Esta función es simétrica para $x = 0$.
Para analizar si esta distribución tiene esperanza se aplicará la ecuación (3.1.4), que debería darnos absolutamente convergente.
Siendo $f_X(x)$ la función densidad de la distribución de Cauchy, se reemplaza:

$$\int_{-\infty}^{\infty} |x|.\, f_X(x) = \frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2} = \infty$$

Dado que no es convergente, la distribución de Cauchy no tiene esperanza matemática.

3-1-2 *ESPERANZA DE UNA FUNCIÓN DE VARIABLE ALEATORIA*

En ocasiones se necesita calcular la esperanza matemática de una función que está relacionada con una variable aleatoria.

- ***Esperanza matemática de una función de variable aleatoria discreta:***

Sea una función real g definida: $g: R \to R \ y \ X: \Omega \to R$ una variable aleatoria discreta, entonces la función compuesta $g\ oX = g(X)$ es también una variable aleatoria, de forma que su esperanza se puede obtener aplicando la definición de esperanza de una variable aleatoria

$$E\big(g(X)\big) = \sum_{x_j \in X(\Omega)} g(x_j)\, f_X(x_j)$$

- ***Esperanza matemática de una función de variable aleatoria continua:***

Sea la función real g definida: $g: R \to R$ y X una variable aleatoria definida:

$X: \Omega \to R$, ambas funciones continuas, entonces la función compuesta: $g\, oX = g(X)$ es también una variable aleatoria, de forma que su esperanza es:

$$E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x).f(x)dx$$

La esperanza matemática existe siempre y cuando

$$\int_{-\infty}^{\infty} |g(x)|.f(x)dx < \infty$$

Ejemplo 3-3:
Si X es una variable aleatoria y sea la función compuesta g(X)= X^4; g(X) es también una variable aleatoria y su esperanza se calcula:

$$E(g(X)) = E(X^4) = \sum_{x_j \in X(\Omega)} x^4 f_X(x_j)$$

Ejemplo 3-4:
Un almacén vende diariamente 0, 1, 2, 3 ó 4 artículos con probabilidad 10%, 40%, 30, 15% y 5% respectivamente. Mantener el local le cuesta 40$ diariamente a la empresa. Por cada artículo que vende tiene una ganancia de 50$. Encuentre el valor esperado de la ganancia diaria.
Sea X: número de artículos que vende cada día.
La función de probabilidad de la variable aleatoria X es:

x	0	1	2	3	4
$f_X(x)$	0,1	0,4	0,3	0,15	0,05

Sea G: Ganancia diaria de la empresa
G es una función real, G(X) =50X-40
G(X) es una función de la variable aleatoria X, que representa la ganancia diaria de la empresa en función de los artículos que vende y el gasto del mantenimiento del local diario

$$E\big(g(X)\big) = \sum_{x_j \in X(\Omega)} g(x_j) f_X(x_j)$$

$$E\big(G(X)\big) = \sum_{x_j \epsilon X(\Omega)} (50.x_j - 40).f_X(x_j)$$

$$E\big(G(X)\big) = (50.0 - 40).0{,}1 + (50.1 - 40).0{,}4 + (50.2 - 40).0{,}3 + (50.3 - 40).0{,}15 + (50.4 - 40).0{,}05 = 42{,}5$$

Se espera tener una ganancia promedio diaria de 42,5$

Ejemplo 3-5:
El jefe de depósito en una fábrica ha desarrollado la siguiente distribución de probabilidad para la demanda diaria (o sea el número de veces que se utiliza) para una herramienta en particular. Se sabe que a la fábrica le cuesta $10 cada vez que se utiliza la herramienta.

X	0	1	2
$f_X(x)$	0,1	0,5	0,4

a) Una situación que puede interesarnos, es conocer, ¿cuál es el valor esperado, el valor promedio, de uso diario de esta herramienta? Contestar este interrogante es plantearnos sobre la esperanza matemática o valor esperado de la variable en estudio, en este caso X: "Demanda diaria de la herramienta"

$$E(X) = \sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) = \mu$$

$$\boldsymbol{E(X) = 0.f_X(0) + 1.f_X(1) + 2.f_X(2)}$$
$$E(X) = 0.0{,}1 + 1.0{,}5 + 2.0{,}4 = 1{,}3$$

Se espera una frecuencia promedio de demanda de dicha herramienta de 1,3 veces por día.

b) El contador de la empresa puede preguntarse ¿cuál es el costo esperado por el uso diario de esa herramienta?
El costo C(X), es una función de la variable aleatoria X, que representa el costo diario de la empresa, en función de la demanda diaria de la herramienta

$$C(X) = 10.X$$

$$\boldsymbol{E(C(X)) = 10.\ 1{,}3 = 13\$}$$

Se espera un costo promedio por el uso de esa herramienta de 13 $ diarios.

3-1-4 PROPIEDADES DE LA ESPERANZA MATEMÁTICA

1. La esperanza de una constante es la misma constante.
   Si para todo $\omega \epsilon \Omega$, se cumple que $X(\omega) = c$, donde c es un número real
   Se cumple $P(X = c) = 1$, entonces:

$$E(c) = c$$

Demostración: Sea X una variable aleatoria discreta:

$$E(X) = \sum_{x_j \epsilon X(\Omega)} c.f_X(c) \quad (Definición\ de\ esperanza)$$
$$E(X) = c.\sum_{x_j \epsilon X(\Omega)} f_X(c) \quad (Extracción\ del\ factor\ común\ c)$$
$$E(X) = c.1 = c \ \ (Propiedad\ fundamental\ de\ función\ de\ probabilidad)$$

Demostración: Sea X una variable aleatoria continua:

$$E(X) = \int_{-\infty}^{\infty} c\,.\,f_X(c)dx \quad (Definición\ de\ esperanza)$$

$$E(X) = c.\int_{-\infty}^{\infty} f_X(c)dx \quad (Extracción\ del\ factor\ común\ c)$$

$$E(X) = c.1 = c \quad (Propiedad\ fundamental\ de\ función\ de\ densidad)$$

2. Sean g y h dos funciones cualquiera definidas en los reales, X una variable aleatoria, g(X) y h(X) funciones compuestas que también son variables aleatorias. La esperanza de la suma de dos funciones es igual a la suma de las esperanzas de dichas funciones:

$$E\big(g(X) + h(X)\big) = E\big(g(X)\big) + E\big(h(X)\big)$$

Demostración: Sea X una variable aleatoria discreta:

$$E\big(g(X) + h(X)\big) = \sum_{x_j \epsilon X(\Omega)} \Big(g\big(x_j\big) + h\big(x_j\big)\Big).f_X\big(x_j\big) \quad (esperanza\ de\ una\ fc\ de\ v.a.)$$

$$E\big(g(X) + h(X)\big) = \sum_{x_j \epsilon X(\Omega)} g(x_j)f_X(x_j) + h(x_j).f_X(x_j) \quad (distributiva)$$

$$E\big(g(X) + h(X)\big) = \sum_{x_j \epsilon X(\Omega)} g(x_j)f_X(x_j) + \sum_{x_j \epsilon X(\Omega)} h\big(x_j\big).f_X\big(x_j\big) \quad (asociativa)$$

$$E\big(g(X) + h(X)\big) = E\big(g(X)\big) + E\big(h(X)\big) \quad (esperanza\ de\ una\ fc\ de\ v.a.)$$

Demostración: Sea X una variable aleatoria continua:
Esta demostración queda para el lector.

3. La esperanza de una constante por una función es el producto de la constante por la esperanza de la función.
Siendo c un número real y g una función definida en los reales, g(X) también es una variable aleatoria entonces:

$$E\big(c.g(X)\big) = c.E\big(g(X)\big)$$

Demostración: Sea X una variable aleatoria discreta:

$$E\big(c.g(X)\big) = \sum_{x_j \epsilon X(\Omega)} c.\ g\big(x_j\big)f_X\big(x_j\big) \quad (esperanza\ de\ una\ función\ de\ v.a.)$$

$$E\big(c.g(X)\big) = c.\sum_{x_j \epsilon X(\Omega)} g\big(x_j\big)f_X\big(x_j\big), \quad (c\ puede\ ser\ extraído\ como\ factor\ común)$$

$$E\big(c.g(X)\big) = c.E\big(g(X)\big) \quad (esperanza\ de\ una\ función\ de\ v.a.)$$

Demostración: Sea X una variable aleatoria continua:
Esta demostración queda para el lector.

4. Sea a y b números reales y X es una variable aleatoria entonces:

$$E(aX+b) = aE(X) + b$$

Demostración: Sea X una variable aleatoria discreta:

$$E(aX+b) = \sum_{x_j \epsilon X(\Omega)} (a.x_j + b) f_X(x_j) \quad (definición\ de\ esperanza)$$

$$E(aX+b) = \sum_{x_j \epsilon X(\Omega)} a.x_j f_X(x_j) + \sum_{x_j \epsilon X(\Omega)} b.f_X(x_j) \quad (propiedad\ distributiva)$$

$$E(aX+b) = a.\sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) + b.\sum_{x_j \epsilon X(\Omega)} f_X(x_j) \quad (extracción\ de\ las\ constantes)$$

$$Reemplazando: \sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) = E(X) \ y \sum_{x_j \epsilon X(\Omega)} f_X(x_j) = 1$$

$$E(aX+b) = a.E(X) + b$$

Demostración: Sea X una variable aleatoria continua:
Esta demostración queda para el lector.

5. La esperanza de una variable centrada en su esperanza siempre es cero. Sea X una variable aleatoria entonces:

$$E(X - E(X)) = 0$$

- ***Demostración: Sea X una variable aleatoria discreta***:

$$E(X - E(X)) = \sum_{x_j \epsilon X(\Omega)} (x_j - \mu) f_X(x_j) \quad (definición\ de\ esperanza)$$

$$E\big(X - E(X)\big) = \sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) - \mu \sum_{x_j \epsilon X(\Omega)} f_X(x_j)$$

(Propiedad distributiva y 2da propiedad fundamental de la función de probabilidad)

$$E(X - E(X)) = E(X) - \mu = 0$$

- ***Demostración: Sea X una variable aleatoria continua***:

$$E(X - E(X)) = \int_{-\infty}^{\infty} (X - \mu) f_X(x)\, dx \qquad (definición\ de\ esperanza)$$

$$E(X-E(X)) = \int_{-\infty}^{\infty} x f_X(x)dx - \mu \int_{-\infty}^{\infty} f_X(x)\, dx$$

(Propiedad distributiva y 2da propiedad fundamental de la función de densidad.)

$$E(X-E(X)) = E(X) - \mu = 0$$

Observación:

Las propiedades de esperanza para el caso continuo se demuestran de la misma forma que para el caso discreto, difiere solamente en el hecho, que en vez de expresar sumatorias, en el caso continuo se utiliza integrales. Como puede verse en la demostración de la propiedad 1 y 5.

La esperanza de una variable aleatoria X es de especial importancia en estadística, pues describe el lugar donde se centra la distribución de los valores de la variable aleatoria. Sin embargo la esperanza no da una descripción adecuada de la forma de la distribución de dichos valores, no nos dice nada acerca de la dispersión de sus valores alrededor de la esperanza o valor medio.

Puede ocurrir que la distribución de dos variables aleatorias X_1 y X_2 tengan el mismo valor medio, es decir la misma esperanza y sin embargo sus funciones densidad sean diferentes.

Por ejemplo, si X representa el capital de una persona escogida al azar de un grupo de dos, tanto si la primera posee $99.000 y la segunda $1000, como si cada una tiene $50.000, el valor medio o esperanza matemática tiene el mismo valor de $50.000. Igualmente, el hecho de saber que la talla media de un conjunto de personas es 1,67 m. no se puede deducir si todas tienen una altura aproximada a este valor medio o si hay algunas muy altas y otras muy bajas. Otro ejemplo es el caudal de las aguas de un río durante las 52 semanas del año. A partir del conocimiento de su valor medio no se puede concluir si se trata de un río de caudal aproximadamente constante durante todo el año, o bien de un río muy caudaloso en verano y casi seco en invierno.

Por este motivo se debe analizar otras medidas que caracterizan a la variable X, estas medidas suelen ser la esperanza matemática de ciertas funciones g(X). Las más importantes son las potencias de X, que dan lugar a los llamados momentos.

3-1-3 MOMENTOS DE UNA VARIABLE ALEATORIA

Vamos a analizar la esperanza de ciertas funciones que nos permitirán medir la dispersión de los valores de la variable aleatoria X alrededor de su esperanza E(X).

Si tomamos la función $g(X) = |X - E(X)|$, valor absoluto de la diferencia $X - E(X)$, le calculamos su esperanza, evidentemente esta medida nos dará una idea de la mayor o menor concentración de los valores de X en torno de la E(X). Como trabajar con valores absolutos es muy incómodo matemáticamente, se prefiere otras funciones, las más importantes son las potencias de la variable aleatoria X, que dan lugar a los llamados momentos.

Los momentos son valores esperados de ciertas funciones de X, y pueden definirse alrededor del cero o de la $E(X)$. Si están definidos alrededor del cero se denominan momentos no centrados, si se definen alrededor de $E(X)$, se denominan momentos centrados. Por lo que se tienen dos tipos de momentos:

- Momentos no centrados de orden k
- Momentos centrados de orden k

3-1-3-1 DEFINICIÓN DE MOMENTOS NO CENTRADOS DE ORDEN K

Dada una variable aleatoria definida sobre un espacio de probabilidad (Ω, $\mathscr{P}(\Omega)$, P), se denomina momento no centrado de orden k, $k \epsilon N$, de la variable aleatoria X, a la esperanza matemática de la potencia X^k

$$\mu'_k = E\left(X^k\right)$$

- ***Si X es una variable aleatoria discreta su momento no centrado de orden k se calcula:*** Aplicando la esperanza de una función de variable aleatoria, de forma que:

$$\mu'_k = E\left(X^k\right) = \sum_{x_j \epsilon X(\Omega)} x_j^k \, f_X\left(x_j\right)$$

Siendo f_X la función de probabilidad de la variable X.

- ***Si X es una variable aleatoria continua su momento no centrado de orden k se calcula:*** Aplicando la esperanza de una función de variable aleatoria, de forma que:

$$\mu'_k = E\left(X^k\right) = \int_{-\infty}^{\infty} x^k \, f_X(x) dx$$

Siendo $f_X(x)$ la función de densidad de la variable X.

El momento existe cuando la serie o la integral correspondiente es absolutamente convergente.

Casos particulares:

1. Si k=1,

$$\mu'_1 = E(X^1)$$

Se obtiene lo que hemos definido como Esperanza de la variable aleatoria X. Por lo tanto la esperanza matemática es el momento no centrado de orden 1 de una variable aleatoria. Lo anotamos $\mu = E(X)$

- ***Si X es una variable aleatoria discreta*** su momento no centrado de orden 1 se calcula:

$$\mu_1' = E(X^1) = \sum_{x_j \epsilon X(\Omega)} x_j^1 \, f_X(x_j)$$

- ***Si X es una variable aleatoria continua*** su momento no centrado de orden 1 se calcula:

$$\mu_1' = E(X^1) = \int_{-\infty}^{\infty} x \, f_X(x) dx$$

2. Si k=2,

$$\mu_2' = E(X^2)$$

Se obtiene el momento no centrado de orden 2, que lo trabajaremos más adelante.

- ***Si X es una variable aleatoria discreta*** su momento no centrado de orden 2 se calcula:

$$\mu_2' = E(X^2) = \sum_{x_j \epsilon X(\Omega)} x_j^2 \, f_X(x_j)$$

- ***Si X es una variable aleatoria continua*** su momento no centrado de orden 2 se calcula:

$$\mu_2' = E(X^2) = \int_{-\infty}^{\infty} x^2 \, f_X(x) dx$$

3-1-3-2 DEFINICIÓN DE MOMENTOS CENTRADOS DE ORDEN K

Se definirá a continuación el momento centrado alrededor del valor medio o esperanza de la variable aleatoria X, de orden k, con $k \epsilon N$.

Los momentos centrados de orden k, se definen por la ecuación:

$$\mu_k = E\left((X - E(X))^k\right) \quad (3.1.3)$$

Como $\mu = E(X)$ se puede reemplazar en la ecuación (3.1.3) obteniendo:

$$\mu_k = E\left((X - \mu)^k\right)$$

- ***Momento centrado de orden k de una variable aleatoria discreta:***

Por las propiedades de la esperanza matemática el momento central de orden k se obtiene:

$$\mu_k = E\left((X - \mu)^k\right) = \sum_{x_j \epsilon X(\Omega)} \left(x_j - \mu\right)^k f_X(x_j)$$

- ***Momento centrado de orden k de una variable aleatoria continua:***

Por las propiedades de la esperanza matemática el momento central de orden k se obtiene:

$$\mu_k = E\left((X-\mu)^k\right) = \int_{-\infty}^{\infty} \left(x_j-\mu\right)^k f_X(x)dx$$

Casos particulares:

1. Si k=1, obtendremos:

$$\mu_1 = E((X-\mu)^1) = E(X) - E(\mu) = E(X) - \mu = \mu - \mu = 0$$

   Para cualquier variable aleatoria el momento centrado de orden 1 es igual a cero.

2. Si k=2, obtendremos:

$$\mu_2 = E[(X-\mu)^2] = Var\ (X)$$

   El momento centrado de segundo orden se denomina varianza de la variable aleatoria X

$$\mu_2 = Var\ (X) = \sigma_X^2$$

- ***Si X es una variable aleatoria discreta el momento centrado de orden k s***e calcula aplicando la definición de esperanza de una función de variable aleatoria:

$$\mu_2 = E(X-\mu)^k = \sum_{x_j \epsilon X(\Omega)} \left(x_j-\mu\right)^k f_X\left(x_j\right) = Var\ (X)$$

- ***Si X es una variable aleatoria continua el momento centrado de orden k s***e calcula aplicando la definición de esperanza de una función de variable aleatoria:

$$\mu_2 = E(X-\mu)^k = \int_{-\infty}^{\infty} \left(x_j-\mu\right)^k f_X(x)dx = Var\ (X)$$

El momento de segundo orden se denomina varianza, pero que significa varianza? Qué nos indica la varianza.

Para estudiar cómo se distribuyen los valores de una variable alrededor de la media hay que analizar las medidas de variabilidad. Una medida de variabilidad es la varianza, esta medida cuantifica el nivel de dispersión de los valores alrededor de la media.

3-1-4 VARIANZA DE UNA VARIABLE ALEATORIA

La varianza de una variable aleatoria da una idea de la "forma de la distribución", es decir en cuanto se concentran o se dispersan los valores de la variable respecto de algún valor central, como la esperanza. Se denota por Var(X) o σ_X^2.

La varianza es una medida de dispersión.

3-1-4-1 DEFINICIÓN DE VARIANZA

Sea X una variable aleatoria discreta con función de probabilidad f_X y esperanza $E(X) = \mu$. Se llama varianza de X al momento centrado de segundo orden.

$$\sigma_X^2 = Var(X) = E(X - E(X))^2$$

- ***Si X es una variable aleatoria discreta su varianza se calcula:*** Aplicando la definición de esperanza de una función de variable aleatoria tenemos:

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_{x_j \in X(\Omega)} \left(x_j - E(X)\right)^2 \ f_X(x_j)$$

La varianza de una variable aleatoria discreta se calcula como la suma de los productos entre el cuadrado de los desvíos y su probabilidad. Es una medida no negativa.

- ***Si X es una variable aleatoria continua su varianza se calcula:*** Aplicando la definición de esperanza de una función de variable aleatoria tenemos:

$$\sigma_X^2 = \mathrm{Var}(X) = \int_{-\infty}^{\infty} \left(x_j - E(X)\right)^2 \ f_X(x)dx$$

La varianza mide en promedio cuánto se alejan los valores de la variable aleatoria X respecto de su valor medio. Es decir que da una idea de la mayor o menor concentración de los valores de X entorna de la E(X). En este sentido la varianza junto con la media caracteriza a la distribución.

Observaciones:

- Si consideramos la función de la variable aleatoria X, $g_1(X) = \left(X - E(X)\right)^2$, calcular la var(X) es calcular la esperanza de esta función, es decir $E\left(g_1(X)\right)$.
- Esta función $g_1(X) = \left(X - E(X)\right)^2$ nos indica cuánto se alejan los valores de la variable aleatoria respecto de su “media” y la varianza considera el “promedio” de todas esas distancias.
- La cantidad $(x_i - \mu)$ se llama desvío de una observación respecto de su media o esperanza.
- La varianza por su definición es una medida no negativa.

- Como la varianza σ^2es una suma de términos positivos, si la varianza es pequeña, significa que todos ellos deben ser pequeños y por lo tanto, o bien $(x_i - \mu)$ es pequeño para todo x_i o bien es pequeña la probabilidad $f(x_i)$. Al contrario si la varianza σ^2 es grande, ello significa que hay valores de X muy separados de su valor medio. En síntesis, la idea que se debe tener de la varianza σ^2, es que si tiene un valor pequeño se trata de una variable aleatoria X cuyos valores difieren poco de su valor medio y, si tiene un valor grande, significa que hay valores de X alejados de su valor medio.

- Debido a que la unidad de medida de la varianza de una variable es el cuadrado de la unidad con que trabaja la variable aleatoria X, para poder interpretar ese dato, comparándolo con la esperanza, E(X), se utiliza la raíz cuadrada positiva de la varianza, esta medida se denomina desviación estándar de la variable X.

3-1-4-2 PROPIEDADES DE LA VARIANZA

1. La varianza de una constante es igual a cero

$$Var(c) = 0, c\epsilon R$$

Demostración:
$Sea\ X: \Omega \rightarrow R, \quad X(\omega) = c,\ \forall \omega \epsilon \Omega$
$E(c) = c,\ c\epsilon R$
$Var(c) = E\big(c - E(c)\big)^2,\ \ (por\ definición\ de\ varianza)$
$Var(c) = E(c - c) = E(0) = 0,\ \ (\,propiedad\ 1\ de\ esperanza)$

2. La varianza de una constante por una variable, es igual al producto de la constante elevada al cuadrado por la varianza de la variable

$$Var(c.X) = c^2.var(X)$$

Demostración:
$Var(c.X) = E(c.X - E(c.X))^2$ (Definición de varianza)

$Var(c.X) = E[(c.X)^2 - 2.c.X.c.E(X) + (c.E(X))^2]$ (Cuadrado de un binomio y esperanza de una constante por una variable aleatoria)

$Var(c.X) = E(c^2.X^2) - 2.c^2.[E(X)]^2 + c^2.[E(X)]^2$ (Esperanza de la suma de funciones)

$Var(c.X) = c^2.E(X^2) - c^2.[E(X)]^2$ (Esperanza de una constante por una variable aleatoria)

$Var(c.X) = c^2(E(X^2) - [E(X)]^2)$ (Factor común)

$Var(c.X) = c^2 var(X)$ (Definición de varianza)

3. $Var(X) = E(X^2) - [E(X)]^2$

Demostración:

$Var(X) = E\left(X - E(X)\right)^2$ (Definición de varianza)

$Var(X) = E(X^2 - 2.X.E(X) + [E(X)]^2)$ (Binomio al cuadrado)

$Var(X) = E(X^2) - 2.E(X).E(X) + [E(X)]^2)$ (Esperanza de la suma de funciones y esperanza de una constante)

$Var(X) = E(X^2) - [E(X)]^2$

4. La varianza de una constante más una variable aleatoria, es la varianza de la variable aleatoria, sea c una constante:

$$Var(c + X) = var(X)$$

Demostración:

$Var\ (c + X) = E(c + X)^2 - [E(c + X)]^2$ *(Se aplica propiedad anterior)*

$Var\ (c + X) = E(c^2 + 2.c.X + X^2) - (c + E(X))^2$ *(binomio al cuadrado y esperanza de una constante)*

$Var\ (c + X) = E(c^2) + 2.c.E(X) + E(X^2) - (c^2 + 2c.E(X) + [E(X)]^2)$

(Propiedad de esperanza de una suma de funciones y binomio al cuadrado)

$Var\ (c + X) = c^2 + 2.c.E(X) + E(X^2) - c^2 - 2c.E(X) - [E(X)]^2$ *(Esperanza de una constante)*

$Var\ (c + X) = E(X^2) - [E(X)]^2$ *(Propiedad 3 de varianza)*

$Var\ (c + X) = Var(X)$

5. Si X es una variable aleatoria y a, b, son constantes entonces:

$$Var\ (aX + b) = a^2 Var(X)$$

La demostración queda como ejercitación para el lector, dado que es inmediata aplicando las propiedades anteriores.

3-1-5 DESVIACIÓN ESTÁNDAR DE UNA VARIABLE ALEATORIA

3-1-5-1 DEFINICIÓN DE DESVIACIÓN ESTÁNDAR DE UNA VARIABLE ALEATORIA

El número no negativo σ, raíz cuadrada de la varianza, se llama desviación típica o desviación estándar de la variable aleatoria X.

$$\sigma_X = +\sqrt{Var(X)}$$

- Cómo se dijo no se puede comparar la varianza de la variable aleatoria con la esperanza de ésta dado que ambas no tienen las mismas dimensiones físicas. Dado que la desviación estándar y la esperanza de una variable aleatoria tienen las mismas dimensiones físicas que la variable aleatoria, en consecuencia entre ellas si se pueden comparar y a partir de estos valores se puede saber cómo están los datos distribuidos alrededor de la media.
- La desviación estándar nos indica cuanto se espera que en promedio los valores de la variable se desvíen respecto de su media.

3-2 RECORRIDO O RANGO DE UNA DISTRIBUCIÓN

Es una medida que caracteriza a la distribución. Es una medida de dispersión. Si toda la masa de la distribución está situada en un intervalo finito, existe una cota g de todos los puntos x tales que $F_X(g) = 0$ y otra cota G de todos los puntos x tales que $F_X(G) = 1$. La longitud G-g del intervalo finito (g, G) se conoce con el nombre de Recorrido o Rango de la distribución.

Análogamente se puede definir un recorrido para distribuciones no contenidas en un intervalo finito. Se puede utilizar como recorridos, intervalos tales como $(\mu - \sigma;\ \mu + \sigma)$, o $(x_{0,25};\ x_{0,75})$, entre otros, aunque estos no contengan el total de la masa de la distribución.

3-3 COEFICIENTE DE VARIACIÓN

A menudo nos interesa comparar variabilidad entre dos o más poblaciones. Puede hacerse esto con sus respectivas varianzas o desviaciones estándar, cuando las variables se dan en las mismas unidades de medida, y estas son aproximadamente iguales. Cuando no sucede esto se utiliza una medida relativa de variabilidad llamada coeficiente de variación.

3-2-1 DEFINICIÓN DE COEFICIENTE DE VARIACIÓN

El coeficiente de variación es una medida de dispersión relativa de un conjunto de datos, se obtiene a partir del cociente de la desviación estándar y su esperanza.

$$C.V. = \frac{\sigma}{\mu}$$

El coeficiente de variación expresa la magnitud de la dispersión de una variable aleatoria con respecto a su valor esperado. Es una medida estandarizada de la variación con respecto a la media.

Debido a que el coeficiente de variación es adimensional es especialmente utilizado para comparar la dispersión relativa de dos distribuciones de probabilidad, principalmente cuando la escala de medición difiere de manera apreciable entre éstas. Se acostumbra a expresar el coeficiente de variación en porcentaje.

$$C.V. = \frac{\sigma}{\mu}.100$$

Un inconveniente del coeficiente de variación es que deja de ser útil cuando el valor de la esperanza de la variable es próximo a cero. Es necesario para el cálculo de este coeficiente que: $\mu \neq 0$

Ejemplo 3-6:
Considere el ejercicio 2-11. El gerente de un almacén en una fábrica ha construido la siguiente distribución de probabilidad para la demanda diaria (número de veces utilizada) de una herramienta en particular. Se sabe que la esperanza de frecuencia promedio de dicha herramienta es de 1,3. Ahora el gerente quiere conocer más sobre la distribución, por lo que quiere calcular su varianza, desviación estándar y coeficiente de variación.

X: "Número de veces que es utilizada la herramienta por día"

$$E(X) = 1{,}3 \; veces \; por \; día$$

Función de probabilidad de X

X	**0**	**1**	**2**
f_X	**0,1**	**0,5**	**0,4**

Se calculará la varianza de la variable:

$$\sigma_X^2 = Var(X) = E(X - E(X))^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_{x_j \in X(\Omega)} \left(x_j - E(X)\right)^2 \; f_X(x_j)$$

$$\sigma_X^2 = Var(X) = (0 - 1{,}3)^2 . 0{,}1 + (1 - 1{,}3)^2 . 0{,}5 + (2 - 1{,}3)^2 . 0{,}4$$

$$\sigma_X^2 = Var(X) = 0{,}41 \; (veces/día)^2$$

Este resultado no dice mucho, se debe calcular su desviación estándar para poder interpretar.

$$\sigma_X = \sqrt{Var(X)} = \sqrt{0{,}41 \; (veces/día)^2} = 0{,}64 \; veces \; por \; día$$

Interpretación: Se espera que en promedio el número veces que se utilice la herramienta por día se desvíe respecto de su media en 0,64.

Se calculará a continuación el coeficiente de variación:

$$C.V. = \frac{\sigma}{\mu} . 100$$

$$C.V. = \frac{0{,}64}{1{,}3} . 100 = 49{,}23 \; \%$$

Interpretación: El número promedio de veces que se utiliza la herramienta por día se dispersa en un 49,23% de la media.

3-3 ASIMETRÍA - COEFICIENTE DE ASIMETRÍA

Asimetría o simetría es una característica de la variable aleatoria. Es una medida de forma de la distribución que permite identificar y describir la manera como los datos tiende a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución.

Se dice que una distribución es simétrica cuando sus datos se distribuyen aproximadamente de igual forma a ambos lados de la esperanza matemática. En caso contrario la distribución se denomina asimétrica. Presentará un alargamiento hacia uno de ambos lados o sesgo.

Los momentos centrales de tercer y cuarto orden de una variable aleatoria X, proporcionan información respecto a la forma de la distribución de probabilidad de la variable aleatoria X. El tercer momento central se define:

$$\mu_3 = E(X-\mu)^3$$

Este momento está relacionado con la asimetría de la distribución de probabilidad de la variable X. Para las distribuciones de probabilidad que presentan una sola moda, si $\mu_3 < 0$, se dice que la distribución es asimétrica negativamente, si $\mu_3 > 0$, la distribución es asimétrica positivamente, y si $\mu_3 = 0$ la distribución es simétrica.

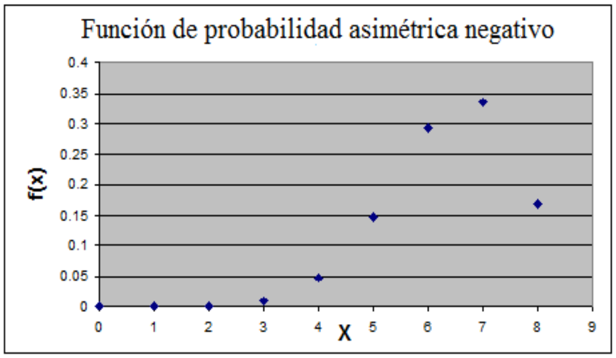
Si la curva de la distribución de una variable aleatoria es simétrica, todos los momentos centrales de orden impar de la distribución son nulos. Dado que los valores $\left(x - E(X)\right)^{2n+1} f(x)$ para $x \epsilon (-\infty, E(X))$ coinciden con signo contrario con los obtenidos para $x \epsilon (E(X), +\infty)$ por lo que se cancelan dando cero. Por ello algunos de estos momentos pueden dar idea del grado de asimetría o de deformación de la distribución. En general se utiliza el momento central de orden tres.

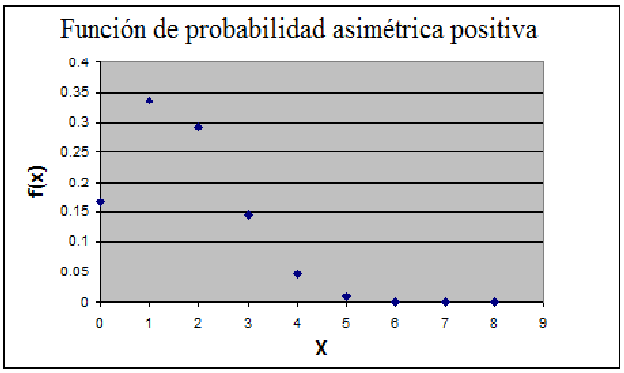
3-3-1 DEFINICIÓN DE COEFICIENTE DE ASIMETRÍA

Una medida de asimetría adimensional será el tercer momento centrado estandarizado:

$$\gamma_3 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} = \frac{\mu_3}{{\sigma_X}^3}$$

Que recibe el nombre de coeficiente de asimetría. El coeficiente γ_3 es una medida de asimetría de una distribución de probabilidad con respecto a su dispersión. Una distribución de probabilidad puede ser simétrica o asimétrica positiva o negativa. Si $\gamma_3 = 0$, entonces $\mu_3 = 0$ y la curva es simétrica. En el caso que $\gamma_3 > 0$ se cumple que $\mu_3 > 0,$ la curva es asimétrica por la derecha o asimétrica positiva, (es mayor la rama derecha), si $\gamma_3 < 0$ se cumple que $\mu_3 < 0$, la curva es asimétrica por la izquierda o asimétrica negativa en este caso es mayor la rama izquierda.





3-4 COEFICIENTE DE APUNTAMIENTO O CURTOSIS

La curtosis o medida de apuntamiento de una variable aleatoria, es una característica de forma de la distribución. Es una medida estadística que determina el grado de concentración que presenta los valores de una variable aleatoria alrededor de una medida central.

La curtosis mide la concentración de datos alrededor de su esperanza, dando una idea de su mayor o menor concentración de datos alrededor de la misma. De forma que su distribución sea más escarpada o más achatada respectivamente.

El cuarto momento central es una medida de qué tan puntiaguda es una distribución de probabilidad. Recibe el nombre de curtosis o kurtosis.

$$\mu_4 = E(X-\mu)^4$$

3-4-1 DEFINICIÓN DE COEFICIENTE DE APUNTAMIENTO

Es preferible utilizar el cuarto momento estandarizado como una medida de curtosis, y recibe el nombre de coeficiente de apuntamiento o curtosis.

$$\gamma_4 = \frac{\mu_4}{(\mu_2)^2} = \frac{\mu_4}{{\sigma_X}^4}$$

Si $\gamma_4 > 3$ la distribución de probabilidad presenta un pico relativamente alto, es decir exhibe un elevado grado de concentración alrededor de la media, recibe el nombre de leptocúrtica; si $\gamma_4 < 3$ la distribución es relativamente plana, es decir presenta un reducido grado de concentración alrededor de la media, recibe el nombre de platicúrtica; y si $\gamma_4 = 3$, la distribución de probabilidad no presenta un pico ni muy alto, ni muy bajo, es decir exhibe un grado medio de concentración de los valores alrededor de la media, recibe el nombre de mesocúrtica. Los tres modelos de distribuciones se encuentran graficadas en la fig.

El coeficiente de curtosis de una variable con distribución normal es $\gamma_4 = 3$, por lo que ésta es un ejemplo de distribución mesocúrtica. Si la curva tiene un pico mayor a la normal, entonces estaríamos en un caso denominada leptocúrtica. Si la curva de distribución tiene un pico menor que la normal, entonces estamos en presencia de una distribución platicúrtica.

Como se dijo el coeficiente de una distribución normal es $\gamma_4 = 3$, por ello en ocasiones se utiliza el coeficiente llamado de exceso:

$$\gamma_4' = \frac{\mu_4}{(\mu_2)^2} - 3$$

Proporciona el exceso o defecto de apuntamiento de una distribución respecto a una variable aleatoria con distribución normal.

Si $\gamma_4' > 0$ se dice que el exceso es positivo y la curva de la distribución es más alta y esbelta que la normal. A este tipo de distribuciones se las denomina leptocúrticas. Si $\gamma_4' < 0$ se dice que el exceso es negativo y la curva de la distribución es más aplanada que la de la normal. Se dice entonces que la distribución es platicúrtica.

Ejemplo 3-7:

Para ensamblar una máquina se utilizan dos componentes mecánicos. De experiencias anteriores se conoce que la probabilidad que el primer componente cumpla con las especificaciones de calidad es de 0,92 y que el segundo componente cumpla las especificaciones es de 0,85. Se sabe también que ambas componentes trabajan independientemente.

a) Encuentre la función de distribución del número de componentes que cumplen las especificaciones.
b) Represente la función de probabilidad.
c) Encuentre la función de distribución acumulada del número de componentes que cumplen las especificaciones.
d) Encuentre la esperanza de la variable discreta, X: Número de componentes que cumplen con las especificaciones.
e) Calcule el momento centrado de tercer orden.
f) Calcule el momento centrado de cuarto orden.
g) Encuentre su varianza y desviación estándar.
h) Calcule el coeficiente de variación, el coeficiente de simetría y el coeficiente de curtosis.

Solución

Sea X: “El número de componentes que cumplen con las especificaciones de calidad”

a) Sea E_i: “El componente i cumple con las especificaciones de calidad” con $i = 1, 2$
$\bar{E}_i$: “El componente i no cumple con las especificaciones de calidad” con $i = 1, 2$

$P(E_1) = 0{,}92 \quad P(\bar{E}_1) = 0{,}08$
$P(E_2) = 0{,}85 \quad P(\bar{E}_2) = 0{,}15$

Siendo $X(\Omega) = \{0, 1{,}2\}$

Función de probabilidad:

$$f_X(x) = P(\{\omega/\omega \,\epsilon\, \Omega \wedge \mathrm{X}(\omega) = \mathrm{x}\}) = P(X = x)$$

$$f_X(0) = P(X = 0) = P(\{\omega/\omega \,\epsilon\, \Omega \wedge \mathrm{X}(\omega) = 0\}) = P(\{(\bar{e}_1, \bar{e}_2)\}) = 0{,}08.\,0{,}15 = 0{,}012$$
$$f_X(1) = P(X = 1) = P(\{\omega/\omega \,\epsilon\, \Omega \wedge \mathrm{X}(\omega) = 1\}) = P(\{(e_1, \bar{e}_2), (\bar{e}_1, e_2)\}) = 0{,}206$$
$$f_X(2) = P(X = 2) = P(\{\omega/\omega \,\epsilon\, \Omega \wedge \mathrm{X}(\omega) = 2\}) = P(\{(e_1, e_2)\}) = 0{,}92.0{,}85 = 0{,}782$$

Esta función puede escribirse de la siguiente forma:

$$f_X(x) = \begin{cases} 0{,}012 & si\ x = 0 \\ 0{,}206 & si\ x = 1 \\ 0{,}782 & si\ x = 2 \\ 0 & cualquier\ otro\ caso \end{cases}$$

b) Representación gráfica de la función de probabilidad de la variable aleatoria discreta:
X: “Número de componentes que cumplen con las especificaciones de calidad”

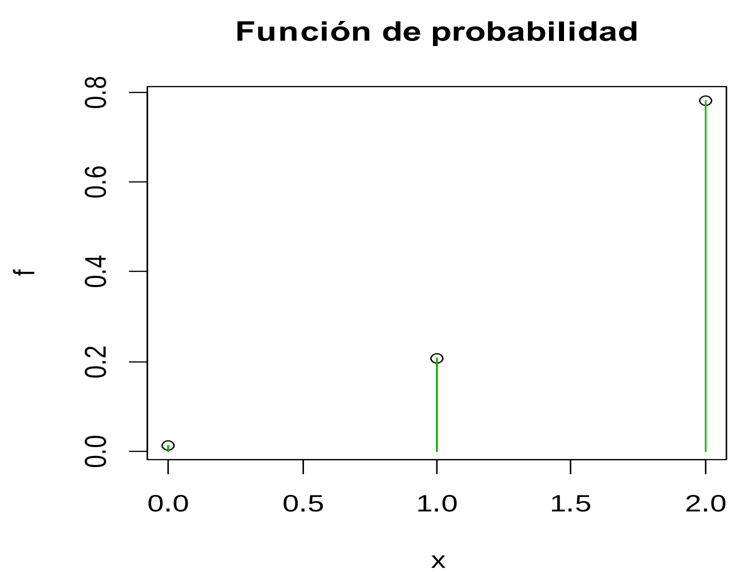


Gráfico 3-7-1. Ejercicio 3-7 item b

c) Función de distribución acumulada de X:

$$F_X(0) = P(X \leq 0) = P(\{\omega/\omega \epsilon \Omega \wedge \mathrm{X}(\omega) \leq 0\}) = P(\{(\bar{e}_1, \bar{e}_2)\}) = 0{,}012$$
$$F_X(1) = P(X \leq 1) = P(\{\omega/\omega \epsilon \Omega \wedge \mathrm{X}(\omega) \leq 1\}) = P(\{(\bar{e}_1, \bar{e}_2), (e_1, \bar{e}_2)(\bar{e}_1, e_2)\}) = 0{,}218$$
$$F_X(2) = P(X \leq 2) = P(\{\omega/\omega \epsilon \Omega \wedge \mathrm{X}(\omega) \leq 2\}) = P(\{(\bar{e}_1, \bar{e}_2), (e_1, \bar{e}_2)(\bar{e}_1, e_2), (e_1, e_2)\}) = 1$$

Esta función puede escribirse de la siguiente forma:

$$F_X(x) = \begin{cases} 0 & si \quad x < 0 \\ 0{,}012 & si\ 0 \leq x < 1 \\ 0{,}218 & si\ 1 \leq x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

d) Esperanza de X:

$$E(X) = \sum_{x_j \epsilon X(\Omega)} x_j f_X(x_j) = \mu$$

$$E(X) = 0.f(0) + 1.f(1) + 2.f(2)$$
$$E(X) = 0.0{,}012 + 1.0{,}206 + 2.0{,}782$$
$$E(X) = 1{,}77$$

Se espera que el número promedio de componentes que cumplan con las especificaciones de calidad sea 1,77.

e) Momento de tercer orden:

$$\mu_3 = E(X - \mu)^3 = \sum_{i=0}^{2} (x_i - 1{,}77)^3 f(x_i)$$

$$\mu_3 = (0 - 1{,}77)^3 f(0) + (1 - 1{,}77)^3 f(1) + (2 - 1{,}77)^3 f(2)$$

$$\mu_3 = (0 - 1{,}77)^3.\ 0{,}012 + (1 - 1{,}77)^3. 0{,}206 + (2 - 1{,}77)^3. 0{,}782 = -0{,}15$$

Como $\mu_3 < 0$, se dice que la distribución es asimétrica negativamente.

f) Momento de cuarto orden

$$\mu_4 = E(X - \mu)^4 = \sum_{i=0}^{2} (x_i - 1{,}77)^4 f(x_i)$$

$$\mu_4 = (0 - 1{,}77)^4 f(0) + (1 - 1{,}77)^4 f(1) + (2 - 1{,}77)^4 f(2)$$

$$\mu_4 = (0 - 1{,}77)^4. 0{,}012 + (1 - 1{,}77)^4. 0{,}206 + (2 - 1{,}77)^4 .0{,}782 = 0{,}1924$$

g) Varianza(X)

$$\mu_2 = \sigma_X^2 = Var(X) = E(X - E(X))^2$$

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_{x_j \in X(\Omega)} \left(x_j - E(X)\right)^2 f_X(x_j)$$

$$\sigma_X^2 = Var(X) = (0 - 1{,}77)^2 . 0{,}012 + (1 - 1{,}77)^2 . 0{,}206 + (2 - 1{,}77)^2 . 0{,}782$$

$$\sigma_X^2 = Var(X) = 0{,}2011\ (componentes)^2$$

Desviación estándar (X)

$$\sigma_X = \sqrt{Var(X)} = \sqrt{0{,}2011\ (componentes)^2}$$

$$\sigma_X = 0{,}448\ componentes$$

Interpretación: Se espera que en promedio la cantidad de componentes que cumplen con las especificaciones se desvíen respecto de su media en 0,448 unidades.

h) Coeficiente de variación

$$C.V. = \frac{\sigma}{\mu}$$

$$C.V. = \frac{0{,}448}{1{,}77} = 0{,}2531$$

$$C.V. = 25{,}31\%$$

Interpretación: El número de componentes que cumplen con las especificaciones en promedio se dispersa en 25,31% con respecto a la media.

i) Coeficiente de simetría

$$\gamma_3 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} = \frac{\mu_3}{{\sigma_X}^3}$$

$$\gamma_3 = \frac{E(X-\mu)^3}{(E(X-\mu)^2)^{\frac{3}{2}}} = \frac{\mu_3}{{\sigma_X}^3}$$

Como $\mu_2 = \sigma_X^2 = Var(X) = 0{,}2011$ se reemplaza

$$\gamma_3 = \frac{-0{,}15}{(0{,}2011)^{\frac{3}{2}}} = \frac{-0{,}15}{0{,}0902} = -1{,}663$$

Como $\gamma_3 < 0$, la curva es asimétrica negativa, como se observa en la representación gráfica de la función de probabilidad, gráfico 3-7-1

j) Coeficiente de curtosis

$$\gamma_4 = \frac{\mu_4}{(\mu_2)^2} = \frac{\mu_4}{{\sigma_X}^4}$$

Como $\mu_2 = \sigma_X^2 = Var(X) = 0{,}2011$ se reemplaza

$$\gamma_4 = \frac{0{,}1924}{(0{,}2011)^2} = \frac{0{,}1924}{0{,}04044} = 4{,}7577$$

Como $\gamma_4 > 3$ la distribución de probabilidad presenta un pico relativamente alto y recibe el nombre de leptocúrtica.

3-5 FUNCIÓN GENERADORA DE MOMENTOS DE UNA VARIABLE ALEATORIA

A veces es muy complicado encontrar los momentos de una variable aleatoria, la función generadora de momentos es una forma alternativa para encontrar los momentos en especial para calcular la esperanza y la varianza de una variable aleatoria.

Se define la función generadora de momentos de una variable aleatoria X, definida sobre un espacio de probabilidad $(\Omega, \mathscr{P}(\Omega)$, P), como la esperanza matemática de la variable aleatoria e^{tX}. La función generadora de momentos tiene como dominio un entorno de los reales incluido el cero y como codominio los reales:

$$m_X: \varepsilon_0 \to R$$
$$m_X(t) = E(e^{tX})$$

Donde t es una variable real y X es una variable aleatoria.

- ***Si X es una variable aleatoria discreta su función generadora de momentos se calcula:***

$$m_X(t) = \sum_{x_j \epsilon X(\Omega)} e^{tx_j} f_X(x_j)$$

- ***Si X es una variable aleatoria continua su función generadora de momentos se calcula:***

$$m_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

3-5-1 OBTENCIÓN DE LOS MOMENTOS NO CENTRADOS A PARTIR DE LA FUNCIÓN GENERADORA DE MOMENTOS

Si la función generadora de momentos existe para $-c < t < c$, entonces existen las derivadas de ésta de todos los órdenes para $t = 0$. Esto asegura que $m_X(t)$ generará todos los momentos de X alrededor del cero.

Los distintos momentos pueden ser obtenidos mediante la diferenciación sucesiva de la función generadora de momentos $m_X(t)$, con respecto a t, y luego se evalúa la derivada en $t = 0$

Ejemplo:

$$m'_X(t) = \frac{d}{dt}E(e^{tX})$$

$$m'_X(t) = E\left[\frac{d}{dt}(e^{tX})\right]$$

$$m'_X(t) = E[Xe^{tX}] \qquad (2.6.8.1)$$

Donde hemos asumido que el intercambio de los operadores de diferenciación y de expectativas es legítimo. Es decir, se ha supuesto que

$$\frac{d}{dt}\left[\sum_x e^{tX}f(x)\right] = \sum_x \frac{d}{dt}[e^{tX}f(x)]$$

Este supuesto casi siempre se puede justificar y, de hecho, es válido para todas las distribuciones consideradas en este texto. Por lo tanto la ecuación obtenida en (2.6.8.1) evaluada en t = 0:

$$m'_X(t = 0) = E[X]$$

Da el momento no centrado de 1er orden de la variable X

Similarmente

$$m''_X(t) = \frac{d}{dt}m'_X(t)$$

$$m''_X(t) = \frac{d}{dt}E[Xe^{tX}]$$

$$m''_X(t) = E\left[\frac{d}{dt}(Xe^{tX})\right]$$

$$m''_X(t) = E[X^2.e^{tX}] \quad (2.6.8.2)$$

Por lo tanto la ecuación obtenida en (2.6.8.2) evaluada en t=0:

$$m''_X(0) = E[X^2]$$

Da el momento no centrado de orden 2 de la variable aleatoria X.

En general la derivada enésima de $m_X(t)$ esta dada por

$$m^n{}_X(t) = E[X^n . e^{tX}] \ para\, n \geq 1$$

Evaluada luego en t=0, nos queda:

$$m^n{}_X(0) = E[X^n] \ para\, n \geq 1$$

De esta forma obtenemos los distintos momentos no centrados de orden n.

Propiedad de la unicidad para funciones generadoras de momentos: Si la función generadora de momentos existe, puede demostrarse que es única y que determina por completo la distribución de probabilidad de la variable X. Es decir, si dos variables aleatorias tienen la misma función generadora de momentos, entonces tienen la misma distribución de probabilidad.

3-6 CUANTILES

Los cuantiles son medidas de posición no centradas que permiten conocer otros puntos característicos de la distribución. Es una medida que describe la posición de un valor específico de la variable en relación al resto de los datos.

En numerosas ocasiones es preciso establecer un valor de la variable aleatoria que tiene una posición fija dentro de la distribución de valores que tiene la variable. Por ejemplo si se estudia el tiempo de vida útil de una pieza fabricada por una empresa puede importar conocer el tiempo de vida útil máximo que tiene por lo menos el 80% de las piezas fabricadas por la empresa, o nos puede importar conocer el tiempo de vida útil mínimo que tiene el 25% de las piezas que tienen mayor vida útil. Estos valores de la variable aleatoria pueden encontrarse a partir de las medidas denominadas cuantiles.

Se puede definir unos valores "x_q" de una variable aleatoria X, que delimiten unas ciertas proporciones de la distribución. A estos valores "x_q" se denominan cuantiles de orden q de la distribución, donde $0 < \mathrm{q} < 1$

3-6-1 DEFINICIÓN DE CUANTILES

- ***Si X es una variable aleatoria discreta,*** se denomina cuantil "$\mathrm{x_q}$", de orden "q", de una variable aleatoria X, al menor valor de la variable para el cual
  $F_X = \mathrm{P(X \leq x_q) \geq q}$, donde $0 < \mathrm{q} < 1$.

- ***Si X es una variable aleatoria continua***, se denomina cuantil de orden "q", al valor "$\mathrm{x_q}$", de la variable aleatoria X, para el cual $F_X(\mathrm{x_q}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x_q}) = \mathrm{q}$.
  Siendo F_X la función de distribución acumulada de X y $0 < \mathrm{q} < 1$.

Los cuantiles más utilizados son los cuartiles, deciles y percentiles.

Aquellos cuantiles que dividen a la distribución en cuatro intervalos cada uno con probabilidad de 0,25, se denominan *"cuartiles de la distribución"*, se representan $x_{0,25}$; $x_{0,50}$; $x_{0,75}$. Se observa que existen tres cuartiles. En particular el segundo cuartil recibe el nombre de mediana.

$x_{0,25}$: es el primer cuartil de orden q=0,25

$x_{0,50}$: es el segundo cuartil de orden q=0,50.

$x_{0,75}$: es el tercer cuartil de orden q=0,75

Aquellos cuantiles que dividen a la distribución en diez intervalos cada uno con probabilidad de 0,10; se denominan "*deciles de la distribución",* se representan $x_{0,10}$; $x_{0,20}$; $x_{0,30}$; ... ; $x_{0,90}$. Se observa que existen nueve deciles. El decil 5 se denomina mediana.

Los percentiles de la distribución son los puntos que dividen a la distribución de probabilidad en 100 intervalos, cada uno con probabilidad 0,01. *Se o*bserva que existen 99 percentiles. Los percentiles resultan muy útiles en la interpretación porcentual de la información. En particular el percentil 50 es la mediana.

La longitud del intervalo entre el primer cuartil y el tercer cuartil, se denomina recorrido intercuartil. Este intervalo toma el 50% central de la distribución. Es muy utilizado cuando existen valores extremos en la distribución.

Este recorrido intercuartil $(x_{0,25},\ x_{0,75})$ fue mencionado cuando se definió recorrido.

3-7 MEDIANA

La mediana es una medida que caracteriza a la distribución. Es una medida de posición centrada. La mediana junto a la esperanza matemática y a la moda son las medidas de tendencia central o localización más importantes que tenemos al analizar una distribución. Se denominan medidas de tendencia central o de localización centradas.

Como se vio la mediana es un cuantil es el cuantil "$x_{0,50}$", de orden $q=0{,}50$.

La mediana es el valor de la variable aleatoria que ocupa la posición central, cuando los datos se disponen en orden de su magnitud. Es decir están ordenados de menor a mayor. El 50% de las observaciones tienen valores menores o iguales a la mediana y el otro 50% de las observaciones tienen valores mayores o iguales a la mediana.

- ***Si X es una variable aleatoria discreta,*** se debe resolver la siguiente ecuación para encontrar el valor de la mediana.
  $F_X(\mathrm{x}_{0,50}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x}_{0,50}) \geq 0{,}50$

- ***Si X es una variable aleatoria continua***, la ecuación a resolver es
  $F_X(\mathrm{x}_{0,50}) = \mathrm{P}(\mathrm{X} \leq \mathrm{x}_{0,50}) = 0{,}50.$
  Siendo F_X la función de distribución acumulada de X

Ejemplo 3-8:

Una empresa ha calculado que el tiempo en horas por obrero, que le demanda para ensamblar un equipo electrónico, es una variable aleatoria cuya función de probabilidad está dada por:

$$f_X(x) = \begin{cases} 3x^{-4} & si\ x > 1 \\ 0 & para\ cualquier\ otro\ caso \end{cases}$$

Sea X: Tiempo demandado para ensamblar un equipo electrónico.

a) Encuentre el valor del cuantil mediana.

Para calcular la mediana obtendremos primeramente la función de distribución acumulada.

$$F_X(x) = P(X \leq x) = \int_1^x 3x^{-4}dx = \int_1^x 3u^{-4}du = -\,u^{-3}|_1^x = 1 - \frac{1}{x^3}; \qquad x > 1$$

$$F_X(x) = 1 - \frac{1}{x^3}; \qquad si\ \ x > 1$$

Entonces la mediana será el valor de la variable aleatoria en el cual la función de distribución acumulada toma el valor 0,5. Entonces:

$$F_X(x_{0,5}) = P(X \leq x_{0,50}) = 0{,}5$$

$$F_X(x_{0,5}) = 1 - \frac{1}{x_{0,5}^3} = 0{,}5$$

$$\frac{1}{x_{0,5}^3} = 0{,}5 \ \Rightarrow x_{0,5}^3 = 2$$

$$x_{0,5} = \sqrt[3]{2} = 1{,}259921 \cong 1{,}26$$

Interpretación: Significa que el tiempo demandado por el 50% de los equipos ensamblados fue aproximadamente 1,26 hs./ obrero o menos, y el 50% demandó 1,26 hs./obrero o más.

b) Encuentre el primer cuartil

El primer cuartil será el valor de la variable aleatoria en el cual la función de distribución acumulada toma el valor 0,25.

$$F_X(x_{0,25}) = P(X \leq x_{0,25}) = 0{,}25$$

$$F_X(x_{0,25}) = 1 - \frac{1}{x_{0,25}^3} = 0{,}25$$

$$\frac{1}{x_{0,25}^3} = 0{,}75 \ \Rightarrow x_{0,25}^3 \cong 1.33$$

$$x_{0,25} \cong \sqrt[3]{1{,}33} \cong 1{,}1$$

Interpretación: Significa que el tiempo demandado por el 25% de los equipos ensamblados fue aproximadamente 1,1 hs./ obrero o menos, y el 75% demandó 1,1 hs./obrero o más.

c) Encuentre el primer decil

$$F_X(x_{0,10}) = P(X \leq x_{0,10}) = 0{,}10$$

$$F_X(x_{0,10}) = 1 - \frac{1}{x_{0,10}^3} = 0{,}10$$

$$\frac{1}{x_{0,10}^3} = 0{,}90$$

$$x_{0,10}^3 = \frac{1}{0{,}90} \cong 1{,}11$$

$$x_{0,10} \cong \sqrt[3]{1{,}11} \cong 1{,}0354$$

Interpretación: Significa que el tiempo demandado por el 10% de los equipos ensamblados fue aproximadamente 1,03 hs./obrero o menos, y el 90% demandó 1,03 hs./obrero o más.

d) Encuentre el cuantil 0,99 o percentil 99

$$F_X(x_{0,99}) = P(X \leq x_{0,99}) = 0{,}99$$

$$F_X(x_{0,99}) = 1 - \frac{1}{x_{0,99}^3} = 0{,}99$$

$$\frac{1}{x_{0,99}^3} = 0{,}01$$

$$x_{0,99}^3 = \frac{1}{0{,}01} = 100$$

$$x_{0,99} \cong \sqrt[3]{100} \cong 4{,}6416 \cong 4{,}64$$

Interpretación: Significa que el tiempo demandado por el 99% de los equipos ensamblados fue aproximadamente 4,64 hs./obrero o menos, y el 1% demandó 4,64 hs./obrero o más.

3-8 MODA

Como se ha dicho la moda es una medida que caracteriza a la distribución. Es una medida de tendencia central o de localización centrada. Podemos decir que moda es un valor de la variable aleatoria donde hay un máximo relativo.

- ***Si X es una variable aleatoria discreta,*** una moda es un valor x_i, tal que

$$f_X(\mathrm{x_i}) \geq f_X(\mathrm{x_{i-1}}) \;\; y \;\; f_X(\mathrm{x_i}) \geq f_X(\mathrm{x_{i+1}})$$

- ***Si X es una variable aleatoria continua***, se define moda como cualquier punto $\mathrm{x_i}$, al que le corresponda un máximo de la función densidad f_X

Una distribución puede tener más de una moda. Si la distribución tiene una moda se denomina unimodal, en el caso de tener dos modas se denomina bimodal, en el caso de tener tres modas se denomina trimodal, etc, según el número de modas o máximos que presente.

Si la distribución es simétrica y unimodal la moda coincide con el valor medio o esperanza y la mediana.

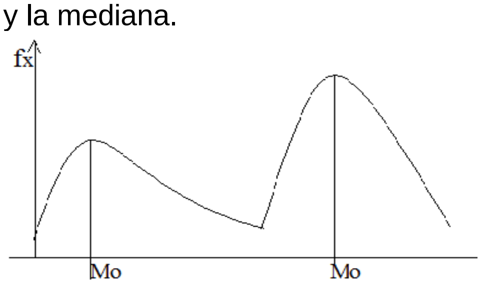


Fig: 3-7-1 Distribución bimodal

Ejemplo 3-8:
Para ensamblar una máquina se utilizan dos componentes mecánicos. De experiencias anteriores se conoce que la probabilidad que el primer componente cumpla con las especificaciones de calidad es de 0,92 y que el segundo componente cumpla las especificaciones es de 0,85. Se sabe también que ambas componentes trabajan independientemente. Calcular la mediana y el segundo decil.

Sea X: "El número de componentes que cumplen con las especificaciones de calidad"

Para el cálculo de los cuantiles debemos recordar la función de distribución acumulada calculada en el ejemplo 2-13:

$$F_X = \begin{cases} 0 & si \quad x < 0 \\ 0{,}012 & si\ 0 \leq x < 1 \\ 0{,}218 & si\ 1 \leq x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

Cálculo de la mediana:

$$F\left(x_q\right) = (P(X \leq x_q) \geq q$$
$$F\left(x_{0,50}\right) = (P(X \leq x_{0,50}) \geq 0{,}50$$
$$F\left(x_{0,50}\right) = P\left(X \leq x_{0,5}\right) = P(X \leq 2) = 1 > 0{,}5 \quad (3.7.1)$$
$$x_{0,5} = 2$$

El valor del segundo cuartil o mediana es $x_{0,5} = 2$, ya que es el menor valor de la variable aleatoria que cumple con (3.7.1)

Interpretación: Por lo menos en el 50% de los casos se espera encontrar a lo sumo dos componentes que cumplan con las especificaciones de calidad.

Cálculo del segundo decil:

$$P\left(X \leq x_q\right) \geq q$$
$$P(X \leq x_{0,20}) \geq 0{,}20$$
$$P\left(X \leq x_{0,20}\right) = P(X \leq 1) = 0.218 > 0{,}20$$

Interpretación: Por lo menos en el 20% de los casos se espera encontrar a lo sumo un componente que cumpla con las especificaciones de calidad.

3-9 TEOREMA DE TCHEBYCHEV

Permite acotar la probabilidad de dispersión de los valores de una variable aleatoria respecto a la media. Su importancia reside que es aplicable para cualquier distribución. Aun cuando la función de distribución de la variable aleatoria es desconocida.

Sea X una variable aleatoria definida sobre un espacio de probabilidad (Ω, $\mathscr{P}(\Omega)$, P), con esperanza y varianza finitas, luego para algún valor $k \in R\ y\ k > 0$, se verifica:

$$P[|X - E(X)| \geq k] \leq \frac{var(X)}{k^2}$$

Demostración para el caso que X es una variable aleatoria absolutamente continua:

$$var(X) = \int_{-\infty}^{\infty} \left(x - E(X)\right)^2 . f_X(x)dx =$$

$$= \int_{-\infty}^{E(X)-k} \left(x - E(X)\right)^2 f_X(x)dx + \int_{E(X)-k}^{E(X)+k} \left(x - E(X)\right)^2 f_X(x)dx \ + \int_{E(X)+k}^{\infty} \left(x - E(X)\right)^2 f_X(x)dx) \geq \text{(3.9.1)}$$

Se integra solo los x_i para los cuales $|X - E(X)| \geq k$ y como

$$|X - E(X)| \geq k \Rightarrow \left(X - E(X)\right)^2 \ \geq k^2$$

Se reemplaza en (3.9.1)

$$var(X) \geq k^2 \int_{-\infty}^{E(X)-k} f_X(x)dx + k^2 \int_{E(X)+k}^{\infty} f_X(x)dx) =$$

$$= k^2 \left[P[X \leq E(X) - k] + P[X \geq (X) + k]\right] = k^2 . P[|X - E(X)| \geq k]$$

$$var(X) \geq k^2 \ P[|X - E(X)| \geq k]$$

Por lo tanto

$$P[|X - E(X)| \geq k] \leq \frac{var(X)}{k^2}$$

Que es lo que queríamos demostrar. La desigualdad de Tchebychev es muy importante y aunque lo hemos demostrado para variables continuas, es válida con análoga demostración para las variables aleatorias discretas.

Si reemplazamos en la ecuación $k = h\sigma$

$$P[|X - E(X)| \geq h\sigma] \leq \frac{1}{h^2}$$

De esta forma queda acotada la probabilidad encerrada en las colas de la distribución a partir del valor de la varianza.

Se observa que cuanto menor sea la varianza menor es la cota de esta probabilidad. La idea física es que un sistema de masas que tenga el momento de inercia pequeño no puede tener mucha masa que esté lejos del centro de gravedad.

La importancia de la desigualdad de Tchebyshev es que nos permite encontrar cotas en probabilidades, cuando sólo la media y la varianza de la distribución de probabilidad son conocidas. Por supuesto si se conoce la distribución real, entonces las probabilidades deseadas podrían ser calculadas exactamente.

A veces nos interesa la probabilidad entre $E(X) - h \ y \ E(X) + h$, es decir en un entorno de la $E(X)$.

$$P[|X - E(X)| \leq h\sigma] \geq 1 - \frac{1}{h^2}$$

Por lo tanto la probabilidad de que una variable tome un valor dentro de h desviaciones estándares respecto a su media, es por lo menos $1 - \frac{1}{h^2}$. Es decir:

Si $h = 2,$ tenemos:

$$P[|X - E(X)| \leq 2\sigma] \geq 1 - \frac{1}{2^2} = \frac{3}{4}$$

El teorema plantea que la variable aleatoria X tiene una probabilidad de que por lo menos $\frac{3}{4}$ de las observaciones caen dentro de las dos desviaciones estándares de la media. Es decir que tres cuartas partes o más de las observaciones de cualquier distribución están en el intervalo $E(X) \pm 2\sigma$.

Si $h = 3$, de igual forma nos queda:

$$P[|X - E(X)| \leq 3\sigma] \geq 1 - \frac{1}{3^2} = \frac{8}{9}$$

El teorema expresa que por lo menos ocho novenas partes de las observaciones de cualquier distribución caen en el intervalo $E(X) \pm 3\sigma$.

Conocer la esperanza y la desviación estándar de una variable aleatoria discreta o continua, nos permite calcular la proporción de la distribución que está situada entre $E(X) \pm h\sigma$.

Ejemplo 3-9:
Supongamos que se sabe que el número de artículos producidos en una fábrica durante una semana es una variable aleatoria con una media de 50 y la varianza de la producción de la semana es 25, entonces, ¿qué se puede decir sobre la probabilidad de que la producción de esta semana este entre 40 y 60 artículos?
Datos:

$$E(X) = 50$$
$$\sigma^2 = 25 \Rightarrow \sigma = 5$$
$$k = h\sigma = h.5$$
$$E(X) + h\sigma = 60 \Rightarrow 50 + h.5 = 60$$
$$E(X) - h\sigma = 40 \Rightarrow 50 - h.5 = 40$$

Por lo tanto tenemos que:

$$h = 2$$
$$k = h\sigma = h.5 = 10$$

Consideremos la relación:

$$P[|X - E(X)| \geq k] \leq \frac{var(X)}{k^2}$$

Se reemplaza por los datos dados en el problema:

$$P[|X - 50| \geq 10] \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Por lo tanto:

$$P[|X - 50| < 10] \geq 1 - \frac{1}{4}$$
$$P[|X - 50| < 10] \geq \frac{3}{4}$$
$$P[-10 < X - 50 < 10] \geq \frac{3}{4}$$
$$P[-10 + 50 < X < 10 + 50] \geq \frac{3}{4}$$

$$P[40 < X < 60] \geq \frac{3}{4}$$

Así la probabilidad de que la producción de esta semana este entre 40 y 60 artículos es de al menos 0,75.

Se puede decir que se sabe que la probabilidad de que el número de artículos producidos en una fábrica, por semana, se encuentra entre 40 y 60 artículos es mayor a $\frac{3}{4}$; pero no se conoce cuánto mayor puede ser en realidad. Sólo cuando se conoce la distribución de probabilidad de la variable aleatoria X es posible determinar en forma exacta.

Como la desigualdad de Tchebychev es válida para todas las distribuciones de la variable aleatoria X, no se puede esperar que la probabilidad calculada este muy cerca de la probabilidad real en la mayoría de los casos.

Ejemplo 3-10:

Si X se distribuye uniformemente en el intervalo (0; 10) con E(X)=5, Var(X)=$\frac{25}{3}$.

Se quiere conocer la $P[|X-5| \geq 4]$.

Se calculará aplicando la desigualdad de Tchebychev y luego aplicando la distribución verdadera que en este caso la conocemos, para poder comparar los resultados.

- Aplicando desigualdad de Tchebychev tenemos:

$$P[|X-E(X)| \geq k] \leq \frac{var(X)}{k^2}$$

$$P[|X-5| \geq 4] \leq \frac{25}{3(16)} \cong 0{,}52$$

- Aplicando distribución uniforme tenemos:

$$P[|X-5| \geq 4] = 0{,}20$$

Así, aunque la desigualdad de Tchebychev es correcta, el límite superior que proporciona no es particularmente cercano de la probabilidad real.

Ejemplo 3-11:

Si X es variable aleatoria normal con media μ y desviación estándar σ, se quiere conocer cuál es la probabilidad de que una variable aleatoria se encuentre a más de dos desviaciones estándares de su media.

Se resolverá aplicando desigualdad de Tchebychev, en la cual no es necesaria conocer la distribución de la variable aleatoria, y luego utilizando su distribución de probabilidad normal.

- La desigualdad de Tchebychev establece que:

$$P[|X-E(X)| \geq h\sigma] \leq \frac{1}{h^2}$$

$$P[|X-\mu| \geq 2\sigma] \leq \frac{1}{4}$$

La desigualdad de Tchebychev dice que la probabilidad de que una variable aleatoria se encuentre a más de dos desviaciones estándares de su media es menor o igual a 0,25

- Aplicando la distribución normal, para obtener la probabilidad exacta:

$$P[|X-\mu| \geq 2\sigma] = P\left\{\left|\frac{X-\mu}{\sigma}\right| \geq 2\right\} = 2[P(Z \geq 2)] = 2\big(1 - P(Z < 2)\big) = 0{,}0455 \quad (3.9.1)$$

La probabilidad de que una variable aleatoria normal se encuentre a más de dos desviaciones estándares de su media es igual a 0,0455

Nota:

En (3.9.1), se reemplazó la variable X estandarizada, $\frac{X-\mu}{\sigma}$, por la variable aleatoria Z.

Se observa en este ejemplo igual que en el 3-10, que la desigualdad de Tchebychev y la probabilidad real son diferentes.

3-10 PROBLEMAS PROPUESTOS SOBRE MEDIDAS QUE CARACTERIZAN A UNA DISTRIBUCIÓN

3.1 En una fábrica de productos químicos, la probabilidad de que se produzca un incendio es de 0,43. La intensidad del incendio puede ocasionar deterioros de 1, 2, 3 ó 4 millones de pesos con probabilidades de 0,375; 0,125; 0,125; 0,025 respectivamente. Un sistema eficiente de prevención de incendios reduce las pérdidas anteriores a 0; 0,5; 0,75 y 1 millón de pesos, con probabilidades de 0,725; 0,125; 0,075 y 0,075 respectivamente, su implementación tiene un costo de $ 300.000. También existe la posibilidad de contratar un seguro contra incendios que cuesta $ 250.000 y cubre riesgos hasta $ 2.500.000. ¿Cuál de las opciones es la más conveniente?

3.2 Considere el ejercicio 2-10 del capítulo anterior. En una empresa de hormigón elaborado se ha pedido al laboratorio de nuevos productos que elabore un hormigón más resistente a grietas por sismos. El director del laboratorio estima la siguiente función de probabilidades para el tiempo que requiere la investigación en meses.

x	2	3	4	5	6	7	8	9	10	11	12
f_X(x)	0,20	0,30	0,15	0,10	0,08	0,06	0,04	0,03	0,02	0,01	0,01

a) Calcule el tiempo esperado, en meses, que requiere la investigación.
b) Calcule la desviación estándar esperada, de la variable aleatoria X.
c) Calcule la probabilidad de que la variable aleatoria tome un valor en el intervalo $(\mu - 2\sigma;\ \mu + 2\sigma)$.
d) El costo de la investigación por mes es de $ 10.000. Calcule el costo esperado de la investigación y la desviación estándar de la misma.
e) Cuál es la probabilidad de que el costo de la investigación sea mayor que $10.000?
f) Observando la gráfica de la función densidad, sin realizar cálculos, determine qué tipo de asimetría identifica en la distribución.
g) Encuentre el coeficiente de apuntamiento y el coeficiente de asimetría. Interprete los resultados.
h) ¿Cuál es el valor modal?, interprete en el contexto del problema.
i) Calcule el cuantil de orden 0,25, primer cuartil y la mediana. Interprete ambos resultados.
j) Determine $E(X^2 - 1)\ y\ var(2.X + 1)$

3.3 Considere el ejercicio 2-13, del capítulo anterior. En un lote de 12 piezas hay 3 no conforme a especificaciones. Se toma al azar tres de ellas. Considere la variable aleatoria X, "Número de piezas no conformes extraídas"

a) Calcule el número de piezas no conformes que se esperada obtener al extraer dos de ellas.
b) Encontrar varianza y desviación estándar de la variable aleatoria X, e interprete los valores obtenidos.

c) A partir de la función de distribución acumulada encuentre e interprete el valor de la mediana y el valor del percentil veintiocho.

3.4 Considere el ejercicio 2-16, del capítulo anterior. Sea la variable aleatoria continua: X: "Tiempo de vida útil de una plaqueta de video, en cientos de horas", cuya función de densidad está dada por:

$$f_X\,(x) = \begin{cases} \dfrac{100}{x^2} & para\ x \geq 100 \\ 0 & para\ x < 100 \end{cases}$$

a) Analice si la variable aleatoria tiene esperanza. En caso de que la variable tenga esperanza, calcule el tiempo esperado de vida útil de una plaqueta de video.
b) Determine el valor de q tal que $P(X \leq q) = 0{,}25$. ¿Qué nombre recibe q? Interprete.
c) Determine el valor de q tal que $P(X \leq q) = 0{,}50$. ¿Qué nombre recibe q? Interprete en términos del problema.
d) ¿Cuál es el tiempo mínimo de vida útil que deberá tener una plaqueta de video para pertenecer al 5% superior de la distribución?
e) ¿Cuál es el tiempo máximo de vida útil que deberá tener una plaqueta de video para pertenecer al 5% inferior de la distribución?
f) ¿Cuál es el tiempo máximo de vida útil correspondiente al 50% central de la distribución?

3.4 Considere el ejercicio 2-18, del capítulo anterior. En una embotelladora la adecuación del proceso por cambio de producto requiere 1, 2, 3 ó 4 hs. dependiendo de las características del producto específico que será producido. Sea la variable aleatoria discreta X: "El número de horas que se requiere para adecuar el proceso".
La función de probabilidad puede ser utilizada para determinar la probabilidad de la cantidad de horas que se requiere para la adecuación del proceso al próximo producto. Sea la función de distribución acumulativa:

$$F_X(x) = \begin{cases} 0 & si\ x < 1 \\ \dfrac{1}{10} & si\ 1 \leq\ x < 2 \\ \dfrac{3}{10} & s\ 2 \leq\ x < 3 \\ \dfrac{6}{10} & si\ 3 \leq\ x < 4 \\ 1 & si\ x \geq 4 \end{cases}$$

a. Calcule el número esperado de horas que se requiere para adecuar el proceso.
b. Calcule la desviación estándar de la variable aleatoria X, interprete.
c. Observando la gráfica de la función de probabilidad realizado en el punto 2-18-b, determine qué tipo de asimetría identifica a la distribución de estos datos. Interprete el resultado.
d. Halle el coeficiente de asimetría y el coeficiente de apuntamiento e interprete los resultados.
e. Encuentre el valor modal e interprete en el contexto del problema.
f. Compruebe los resultados del inciso a) y b) utilizando la función generadora de momentos.

g. Determine los cuantiles de orden 0,25 y 0,50 e interprete en términos del problema.
h. ¿Entre que valores se encuentra el 50% central de la distribución?
i. Halle la probabilidad de que la variable difiera de la media en menos de una desviación estándar.

3.5 Sea X la variable aleatoria que describe el lanzamiento de un dado no legal (cargado). Su función de probabilidad esta dada por la siguiente tabla:

x_i	1	2	3	4	5	6
$f_X(x_i)$	$1/6$	$1/6$	$1/6$	$1/12$	$1/4$	$1/a$

a) Determinar el valor de a.
b) Determinar la esperanza y la varianza de X
c) Obtenga su función de distribución acumulada y represéntela gráficamente.
d) A partir de la función de distribución acumulada obtenida escriba la expresión que le permite calcular la $P(X=2)\ y\ P(1{,}5 < X \leq 5)$. Represente gráficamente estas probabilidades en el gráfico obtenido en el ítem c.

3.6 Considere el ejercicio 2-19 del capítulo anterior. El diámetro de un conductor eléctrico medido en cm, es una variable aleatoria continua X, se puede considerar que su función de densidad está dada por:

$$f_X = \begin{cases} 6.x.(1-x) & si\ 0 \leq x \leq 1 \\ 0 & otro\ caso \end{cases}$$

a) Calcule el diámetro esperado de un conductor eléctrico medido en cm.
b) Calcule la desviación estándar de la variable aleatoria X, interprete.
c) Compruebe los resultados de los incisos a) y b) utilizando la función generadora de momentos.
d) Calcule la probabilidad de que la variable tome un valor en el intervalo $(\mu - 2\sigma;\ \mu + 2\sigma)$
e) Determine el cuantil mediana y los cuartiles de orden 0,25 y 0,75 e interprete en términos del problema.
f) ¿A partir de lo calculado en el ítem anterior, diga entre que valores se encuentra el 50% central de la distribución?
g) Calcule el rango de la distribución y el rango intercuartil. Interprete.

3.7 Considere el ejercicio 2-20 del capítulo anterior. El tiempo que una aeronave tiene que esperar para poder aterrizar en cierto aeropuerto internacional, en minutos, es una variable aleatoria X, cuya función densidad está dada por:

$$f_X(x) = \begin{cases} \frac{1}{3}e^{-\frac{x}{3}} & x > 0 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

a) En promedio, cuánto debe esperar una aeronave para poder aterrizar?
b) Calcule e interprete el valor de la mediana y el valor del 3er cuartil.

c) ¿Cuál es el tiempo mínimo que deberá esperar una aeronave para poder aterrizar para pertenecer al 25% superior de la distribución?
d) ¿Cuál es el tiempo máximo que deberá esperar una aeronave para poder aterrizar para pertenecer al 10% inferir de la distribución?
e) ¿Entre que valores se encuentra el 50% central de la distribución?
f) ¿Qué puede decir de la simetría o asimetría de la distribución de la variable X?
g) Calcule el coeficiente de asimetría y el coeficiente de apuntamiento. Interprete.

3.8 Considere el ejercicio 2-22 del capítulo anterior. La función de distribución de probabilidad de la variable aleatoria X, definida como el número de defectos cada 10 m^2 de una membrana impermeabilizante para techos en rollos está dada por:

X	0	1	2	3	4
$f_X(x)$	0,25	0,49	0,16	0,09	0,01

a) Represente gráficamente la función de probabilidad de la variable X, f_X.
b) Calcule el número esperado de defectos de una membrana impermeabilizante para techos cada 10 m^2.
c) Calcule la varianza y desviación estándar de la variable aleatoria X.
d) Encuentre la moda de la distribución de X e interprete en términos del problema.
e) Determine la mediana de esta distribución e interprete en términos del problema.
f) Marque en el gráfico de la función de probabilidad la Esperanza de la variable X, la mediana y su moda.
g) Encuentre la probabilidad de que la variable difiera de la media en más de un desvío.
h) Determine $E(3.X^4 + 2)$ y $var(X + 5)$
i) Se selecciona un rollo de membrana impermeabilizante y se sabe que tiene menos de dos defectos cada 10 m^2, ¿cuál es la probabilidad de que tenga un defecto cada 10 m^2?

3.8 Una empresa industrial fabrica un producto cuya venta le reporta un beneficio de $1000, si el producto pasa las inspecciones de calidad y $750 si el producto presenta algún defecto. La empresa está pensando en implementar un sistema de gestión de la calidad ISO 9001 y quiere saber si esto es conveniente. Si implementan el sistema de gestión de la calidad tendrán un gasto adicional de $20 por producto. Si resulta algún producto con defecto puede ser subsanado antes de ser entregado con un costo de reparación de $40, evitando de esta forma los costos de la no calidad que esto conlleva. Si el porcentaje de defectuosos puede ser estimada en el 5%, que le conviene a la empresa implementar o no el sistema de gestión de la calidad.

3.9 Considere el ejercicio 2-2 del capítulo anterior. Una caja contiene 3 tornillos defectuosos y 6 no defectuosos. Se extraen dos tornillos aleatoriamente, con reemplazo. A partir de la función de probabilidad de la variable aleatoria X: "Número de tornillos no defectuosos que se obtiene", encuentre:
a) Encuentre la esperanza de la variable aleatoria X.
b) Encuentre la varianza y desviación estándar de la variable X.
c) Determine la mediana de esta distribución e interprete en términos del problema.
d) Calcule la moda de la distribución e interprete en términos del problema.
e) ¿Qué puede decir con respecto a la simetría de la distribución?

f) Calcule el coeficiente de asimetría e interprete.

3.10 De experiencias pasadas un profesor sabe que el puntaje de la prueba de un estudiante en su examen final es una variable aleatoria con media 75, y varianza 25. ¿Qué puede decirse acerca de la probabilidad de que un estudiante obtenga entre 65 y 85 puntos?

Ayuda: En este problema no se conoce la distribución de la variable aleatoria.

3.11 Considere el ejercicio 2-24 del capítulo anterior. Considere la siguiente función de distribución acumulativa de la variable aleatoria X:

$$F_X(x) = \begin{cases} 0 & si \quad x < 0 \\ \dfrac{x^3}{8} & si \quad 0 \leq x \leq 2 \\ 1 & si \quad x > 2 \end{cases}$$

a) Encuentre el valor de la mediana y el valor de los cuartiles.
b) Encuentre el valor de la Esperanza de la variable aleatoria X.
c) Encuentre el recorrido intercuartil. Interprete.
d) Encuentre el cuantil de orden 0,80. Interprete.
e) Encuentre el cuantil de orden 0,80 gráficamente a partir de la función de distribución acumulativa.
f) Encuentre el cuantil de orden 0,5 gráficamente a partir de la función de distribución acumulativa.
g) Represente los cuantiles en la gráfica de la función densidad.

3.12 Considere el ejercicio 2-25 del capítulo anterior. La función de densidad del diámetro de contacto de roscas de un conducto de acero inoxidable está dada por:

$$f_X(x) = \begin{cases} 0 & si \quad x < 0 \\ \dfrac{1}{(1+x)^2} & si \quad x \geq 0 \end{cases}$$

a) A partir de la función de distribución acumulada encuentre el valor de la mediana e interprete.
b) Encuentre los cuartiles y el cuantil 0,99 y represéntelos en la función de distribución acumulada.
c) Encuentre el recorrido intercuartil.
d) Calcule la probabilidad de que X se encuentre a menos de una desviación estándar de su media.

3.13 Considere el ejercicio 2-26. Suponga que un investigador requiere realizar un sondeo de opinión sobre un acontecimiento coyuntural y contando con los recursos suficientes, decide utilizar en su trabajo de indagación, información primaria. La técnica que decide utilizar es la entrevista telefónica. La principal atracción de esta técnica es que permite recolectar información de lugares alejados en forma más económica y rápida que otras técnicas. La proporción de personas que responden a cierta encuesta telefónica es una variable aleatoria continua, cuya función de densidad es:

$$f_X(x) = \begin{cases} 2(5.x-2) & si \quad 0 < x < 1 \\ 0 & cualquier\ otro\ caso \end{cases}$$

a) ¿Qué proporción de personas se espera que respondan la encuesta telefónica?
b) Determine e interprete la mediana de la distribución.
c) Encuentre e interprete el valor del percentil veinticinco.
d) Encuentre la varianza y la desviación estándar de la variable aleatoria e interprete.
e) ¿Qué puede decir de la simetría o asimetría de la distribución?
f) Calcule el coeficiente de asimetría y el coeficiente de apuntamiento. Interprete.

3.14 Considere el ejercicio 2-27 de la unidad anterior. El contenido de ceniza en el carbón (en porcentaje) se puede considerar como una variable aleatoria continua, cuya función de densidad es:

$$f_X(x) = \begin{cases} \dfrac{1}{4875}x^2 & si \quad 10 < x < 25 \\ 0 & para\ cualquier\ otro\ caso \end{cases}$$

a) Hallar el promedio de ceniza que contiene el carbón.
b) Hallar la varianza y la desviación estándar. Interprete.
c) Encuentre entre que valores se encuentra el 50% central de la distribución.
d) Calcule el rango intercuartil.
e) Calcule la probabilidad de que la variable aleatoria X se encuentre a menos de una desviación estándar de su media.

3.15 Considere el ejercicio 2-28 de la unidad anterior. La duración, en cientos de horas, de la batería de un celular recargable es una variable aleatoria continua con la siguiente función de densidad:

$$f_X(x) = \begin{cases} 125x^{-2} & si \quad 100 < x < 500 \\ 0 & para\ cualquier\ otro\ caso \end{cases}$$

a) Encuentre la media de la variable aleatoria en estudio. Represéntela en la gráfica de la función de densidad.
b) Encuentre la mediana de la variable aleatoria en estudio y represéntela en la gráfica de la función de densidad.
c) Encuentre los cuartiles de la distribución de la variable aleatoria X. Interprételos.
d) Encuentre la varianza y la desviación estándar de la variable aleatoria en estudio.
e) Calcule la probabilidad de que la variable aleatoria X se encuentre a menos de dos desviaciones estándar de su media.

3.16 Una fábrica de carrocerías conoce que el número de piezas defectuosas es una variable aleatoria con media 12 por semana. Si la desviación estándar es una medida de variabilidad conocida y es igual a un tercio de la media. ¿Qué probabilidad hay que en una semana determinada la producción de defectuosos este entre 4 y 20 piezas?

3.17 Considere el ejercicio 2-30 del capítulo anterior. Una vinería ofrece distintas clases de vinos, los cuales tienen distintos tiempos de añejamiento, sus clientes pueden elegir entre, 1; 2; 3 o 4 años. En la siguiente tabla se muestra la distribución acumulada del número de años de añejamiento de los vinos seleccionados aleatoriamente.

x_i	$(-\infty, 1)$	[1,2)	[2,3)	[3,4)	[4, ∞)
$F_X(x_i)$	0	0,15	0,39	0,74	1

a) ¿Cuál es el tiempo esperado de añejamiento de los vinos de esta vinería?
b) Encuentre la varianza y desviación estándar de la variable aleatoria X. Interprete en términos del problema.
c) Halle la probabilidad de que la variable difiera de la media en un desvío estándar o más.
d) Halle la probabilidad de que la variable aleatoria difiera de la media en dos o más veces su desvío estándar.
e) Determine la mediana de la distribución e interprete en términos del problema.

3.18 Un corredor de bolsa estudia intensamente el mercado de valores para estar atento a cualquier inversión potencial. En estos momentos examina la posibilidad de invertir en acciones de empresas mineras. Mediante el análisis del rendimiento en el pasado, y la capacidad actual de estas empresas mineras, ha desglosado los resultados potenciales en 5 clases, con sus posibles tasas de rendimiento anual asociadas, sobre una acción que en la actualidad cuesta 10 U$D.

a) Encuentre el valor esperado del rendimiento sobre la inversión de una acción en este grupo de empresas.
b) Recomendará a sus clientes que compren este grupo de acciones siempre que la tasa de rendimiento esperada exceda al 10%, ¿asesorará a sus clientes que compren estas acciones?
c) ¿Cuál es el riesgo de esta inversión?
Nota: En el mundo de la inversión, existe otra variable importante que es el riesgo de una inversión. Es la posibilidad de que el rendimiento real de una inversión se desvíe del rendimiento esperado. Se dice que el riesgo es mayor cuando mayor es la desviación y mayor probabilidad de que ocurra.

Rendimiento	-10	0	10	20	30
Tasa	0,20	0,20	0,30	0,20	0,10

3.19 El número de clientes que concurren al banco a realizar algún trámite financiero un día a la semana, es una variable aleatoria con media 20 clientes y desviación estándar de 3 clientes. ¿Cuál es la probabilidad de que un día de la semana concurran al banco entre 10 y 30 clientes?

3.20 Un vendedor de paraguas puede ganar $6000 por día cuando llueve y pierde $1200 por día cuando no llueve. La probabilidad de que llueva en un día es 0,3. Encuentre la ganancia esperada del vendedor.

3.21 Sea (Ω, P) un espacio de probabilidad y X una variable aleatoria discreta definida en Ω. $Siendo\ X(\Omega) = \{1{,}5, x_3, 15{,}20\}$ y $f_X(x)$ su función de probabilidad, cuyos valores están dados por la siguiente tabla:

x_i	1	5	x_3	15	20

$f_X(x_i)$	0,5	0,1	0,25	0,1	p_5

a) Calcule $x_3\ y\ p_5\ si\ se\ sabe\ que\ E(X^2) = 37\ y\ x_1 < x_2 < x_3 < x_4 < x_5$
b) Calcule la esperanza de X
c) Calcule la mediana y el cuantil 0,25 de X.
d) Encuentre la función de distribución acumulada de la variable aleatoria X. Represéntela gráficamente y marque en ella $f_X(x_2)\ y\ f_X(x_3)$

3.22 Se venden 4000 cupones para un sorteo aún valor de $100 cada uno. El primer y único premio es de $40000. Una persona compra 3 cupones, cuánto debe esperar ganar?. Considere la variable X: "Cantidad de dinero obtenido en el juego"

3.23 Dos personas juegan a cara o cruz y han convenido en continuar la partida hasta tanto la cara o la cruz se haya presentado 3 veces.
a) Dada la variable aleatoria X: "Número de tiradas hasta terminar el juego de un jugador". Encuentre el espacio muestral. Utilice un diagrama de árbol para encontrar todos los sucesos.
b) Encuentre la función de probabilidad de X. Represente gráficamente, f_X.
c) Encuentre la probabilidad de que el juego no se acabe cuando un jugador realizó 4 tiradas o más.
d) ¿Cuántas tiradas se espera debe realizar un jugador para terminar su juego?

3.24 En una fábrica el departamento de control de calidad someten al control diario a sus productos y se estima que la probabilidad de que en un día sean producidos x productos defectuosos está dado por la función $f_X(x)$

$$f_X(x) = 0{,}85.\,0{,}15^x \qquad si\ x = 0{,}1{,}2\ \ldots..$$

Encuentre la probabilidad de que en un día de los productos producidos:
a) Cinco sean defectuosos
b) Dos o más sean defectuosos
c) Tres o menos sean defectuosos
d) Más de 3 sean defectuosos, si se sabe que se han producido más de 2.

3.25 En las siguientes proposiciones completa con V ó F, según corresponda

1		La esperanza de una variable aleatoria continua X, siempre existe.
2		Si X es una variable aleatoria se cumple que $E(2.X) = 2.X$
3		Cuando una variable aleatoria continua X toma el particular valor denominado mediana, x=mediana, la función de distribución acumulada toma el valor 0,5

4		Sea (Ω, P) un espacio de probabilidad y X es una variable aleatoria continua definida en él, con $\mu\ y\ \sigma$, finitas. La función densidad se indica con f_X. Entonces $var(X) = \int_{-\infty}^{\infty}(x-\mu)f_X(x)dx$
5		Sea (Ω, P) un espacio de probabilidad y X es una variable aleatoria continua definida en él. La función densidad se indica con f_X. Entonces siempre se cumple: $$E(5) = 5.\int_{-\infty}^{\infty} f_X.\,dx = 5$$
6		Sea (Ω, P) un espacio de probabilidad y X e Y dos variables aleatorias definida en él. Entonces se cumple: $E(X+Y) = E(X) + E(Y)$
7		Sea (Ω, P) un espacio de probabilidad y X es una variable aleatoria discreta definida en él, cuyos valores con probabilidad positiva son los números naturales. $var(c.X+a) = c.var(X)$
8		Para aplicar el teorema de Tchebychev es condición necesaria que se conozca la distribución de la variable en estudio.
9		Sea (Ω, P) un espacio de probabilidad y X es una variable aleatoria discreta definida sobre él. Entonces se cumple que si la distribución no es simétrica entonces la variable aleatoria X, no tiene moda.
10		El cuantil de orden q de una variable aleatoria X discreta, se define como el menor valor de X para el cual $P(X \le x) \ge q, donde\ 0 < q < 1$. Se lo indica x_q, en particular q=0,5 , se denomina mediana.