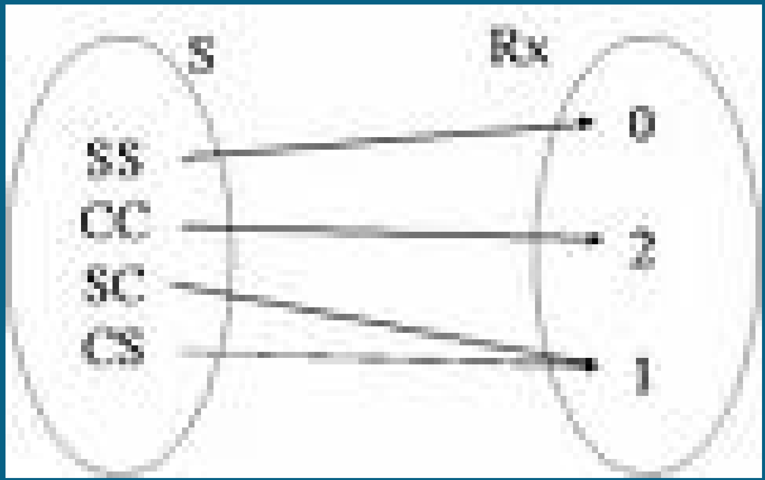
5. VARIABLES ALEATORIAS Y SUS MOMENTOS



Una variable aleatoria

Objetivos

Introducir la idea de una variable aleatoria y su distribución y sus características como la media, la varianza, los cuartíles etc.

Para leer

Podéis ver los *mini-videos* de Emilio Letón y su equipo sobre Variables Aleatorios.

Variables aleatorias

Hasta ahora, hemos tratado de sucesos, por ejemplo $A =$ "la suma de dos tiradas de un dado es 7". Ahora queremos generalizar y tratar de variables, por ejemplo "la suma de las dos tiradas" o "el número de llamadas telefónicas en una hora".

Formalmente, podemos pensar en una variable como una función que asocia un valor numérico a cada suceso elemental del espacio muestral, es decir que una *variable aleatoria* X es una función

$$X : \omega \in \Omega \to X(\omega) \in \mathbb{R}$$

Tipos de variables

Existen varios tipos de variables que necesitan tratamientos distintos.

Una variable es *discreta* si el conjunto de valores posibles un conjunto discreta.

Por ejemplo, $X =$ *el número de cruces en 10 tiradas de una moneda.*

Una variable es *continua* si el conjunto de valores posibles es un continuo o la unión de varios continuos.

El tiempo exacto hasta que reciba una llamada telefónica.

Una variable es *mixta* si puede tomar algunos valores de un conjunto discreta y otros valores de uno o más conjuntos continuos.

El tiempo que tengo que esperar en la cola antes de recibir servicio.

Variables discretas

Para una variable discreta, se puede definir directamente una función de probabilidad, $P(X = x) = \sum_{\omega \in \Omega, X(\omega)=x} P(\omega)$ para cada valor de x.

La función $P(X = x)$ se conoce como la *distribución de probabilidad* de X o la *función de masa* de X.

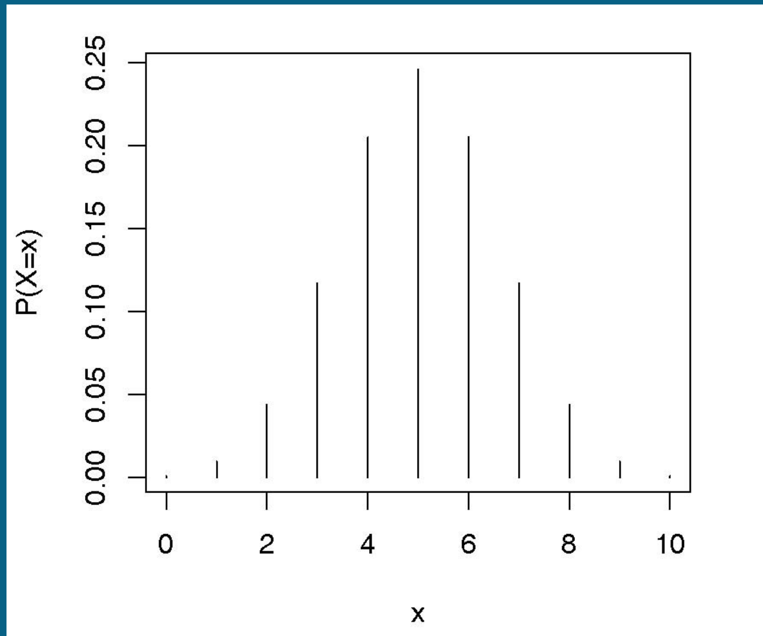
Supongamos que lanzamos una moneda con $P(cruz) = p$ un número n de veces y definimos $X =$ número de cruces.

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, 1, 2, \ldots, n \\ 0 & \text{para otros valores de } x \end{cases}$$

Esta función es un ejemplo de una *distribución binomial*.

Representación gráfica de la distribución

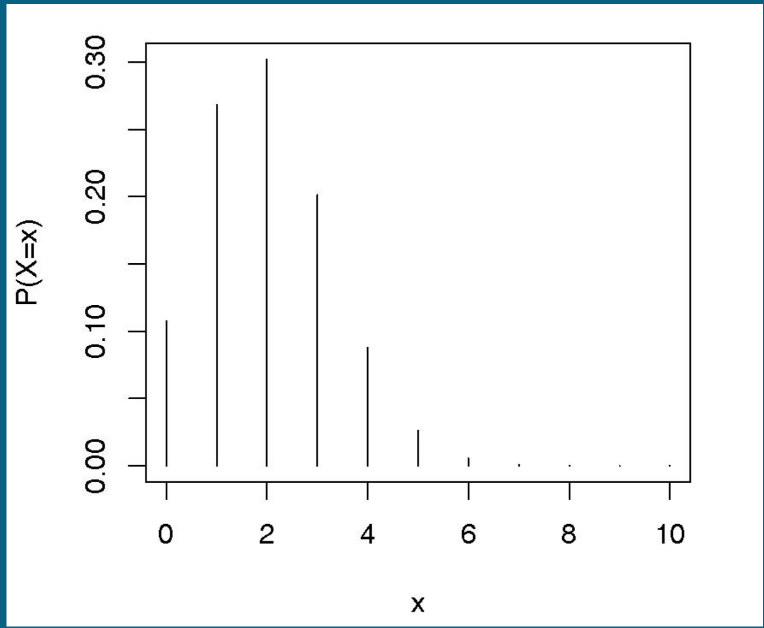
Supongamos que X *se distribuye como binomial con parámetros* $n = 10$ *y* $p = 0.5$.



Se ve que la distribución es *simétrica* y *unimodal* com *moda* 5.

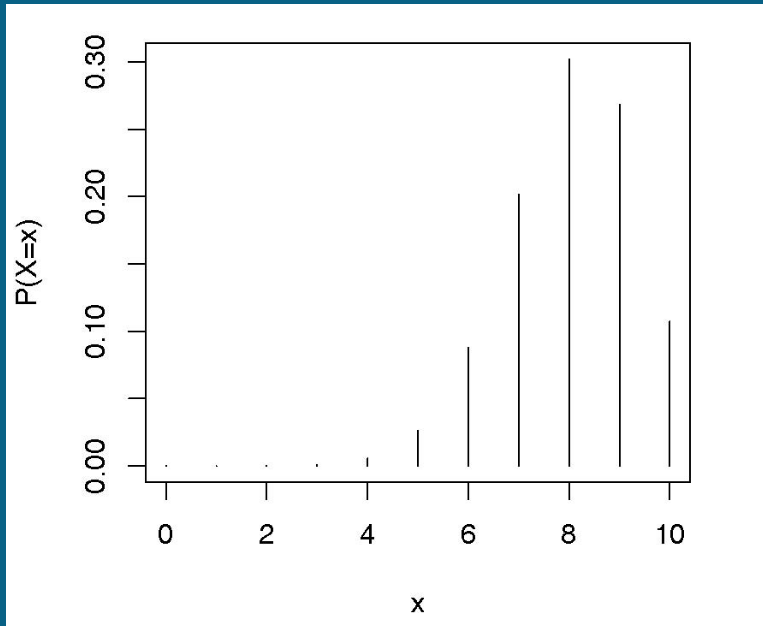
La moda de la distribución es el valor más probable.

Aquí vemos otro ejemplo con $n = 10$ *y* $p = 0.2$.



En este caso la distribución es *asimétrica a la derecha*.

Si $n = 10$ *y* $p = 0.8$,



tenemos una distribución *asimétrica a la izquierda*.

Propiedades de la función de probabilidad

- $0 \leq P(X = x) \leq 1$ para todo $x \in \mathbb{R}$.
- $\sum_i P(X = x_i) = 1$ donde se toma el sumatorio sobre todos los valores posibles de X, say $x_1, x_2, \ldots$.
- $P(X \leq x) = \sum_{i, x_i \leq x} P(X = x_i)$.
- $P(a \leq X \leq b) = \sum_{i, a \leq x_i \leq b} P(X = x_i)$.

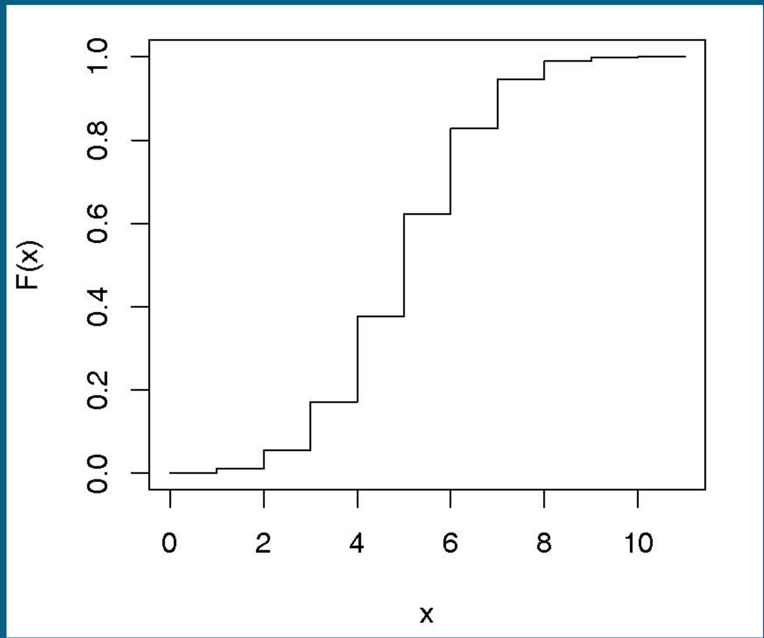
La tercera función se conoce como la función de distribución acumulativa de la variable.

La función de distribución acumulativa

Para una variable, X, se define la *función de distribución acumulativa* como

$$F(x) = P(X \leq x) \quad \text{para } x \in \mathbb{R}.$$

Ilustramos la función para la distribución binomial con $n = 10$ y $p = 0.5$.



Propiedades de la función de distribución acumulativa

- $0 \leq F(x) \leq 1$ para todo $x \in \mathbb{R}$.
- $F(-\infty) = 0$ y $F(\infty) = 1$.
- Si $h > 0$, entonces $F(x+h) \geq F(x)$ para todo x.
- La función de distribución de una variable discreta es una *función escalera*.

Ejemplo

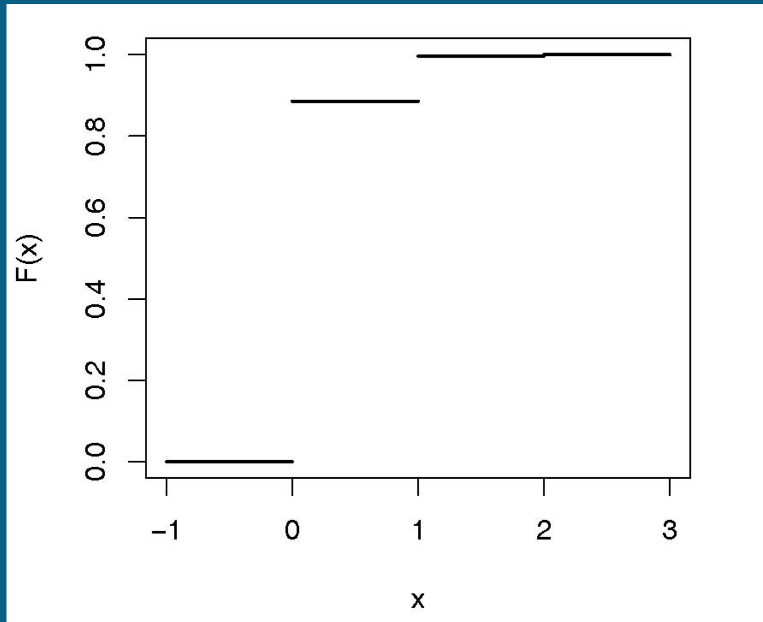
Supongamos que la producción de un día de 850 piezas manufacturadas contiene 50 piezas que no cumplen con los requerimientos del cliente. Se seleccionan del lote dos piezas al azar y sin reemplazo. Sea la variable aleatoria X igual al número de piezas de la muestra que no cumplen. ¿Cuál es la función de distribución acumulada de X?

Primero calculamos la función de probabilidad.

$$
\begin{aligned}
P(X=0) &= \frac{800}{850} \times \frac{799}{849} = 0.886 \\
P(X=1) &= \frac{800}{850} \times \frac{50}{849} + \frac{50}{850} \times \frac{800}{849} = 0.111 \\
P(X=2) &= \frac{50}{850} \times \frac{49}{849} = 0.003
\end{aligned}
$$

Ahora se tiene

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0.886 & \text{si } 0 \leq x < 1 \\ 0.997 & \text{si } 1 \leq x < 2 \\ 1 & \text{si } x \geq 2. \end{cases}$$



Ejemplo

En ocasiones algunas líneas aéreas venden más billetes de los disponibles en un vuelo. Una compañía ha vendido 205 billetes que corresponden a un avión con 200 plazas. Sea X la variable aleatoria correspondiente al número de pasajeros que se presentan en el aeropuerto para viajar en el avión.

x	198	199	200	201	202	203	204	205
$P(X = x)$	0.05	0.09	0.15	0.2	0.23	0.17	0.09	0.02

1. *Hallar la probabilidad de que todos los pasajeros que llegan a coger el avión tengan plaza.*

2. *Obtener la probabilidad de que alguno de los pasajeros que se presentan en el aeropuerto se quede sin plaza.*

Ejemplo

Se sacan tres cartas "al azar" de un mazo de 40 cartas. Denotamos por Y el número de ases que se obtienen. Hallar:

1. *La función de probabilidad de Y.*
2. *La función de distribución de Y*
3. *La probabilidad de ver por lo menos un As.*
4. *La probabilidad de observar 2 Ases o menos.*

Variables continuas

Para una variable continua, X, no tiene sentido definir una función de probabilidad, ya que $P(X = x) = 0$ para todo x. No obstante, es razonable definir la función de distribución, $F(x)$ como anteriormente, con la única diferencia que supongamos que está función es ya una función continua en lugar de una función escalera.

¿Cuáles de las siguientes funciones pueden ser funciones de distribución para una variable continua X?

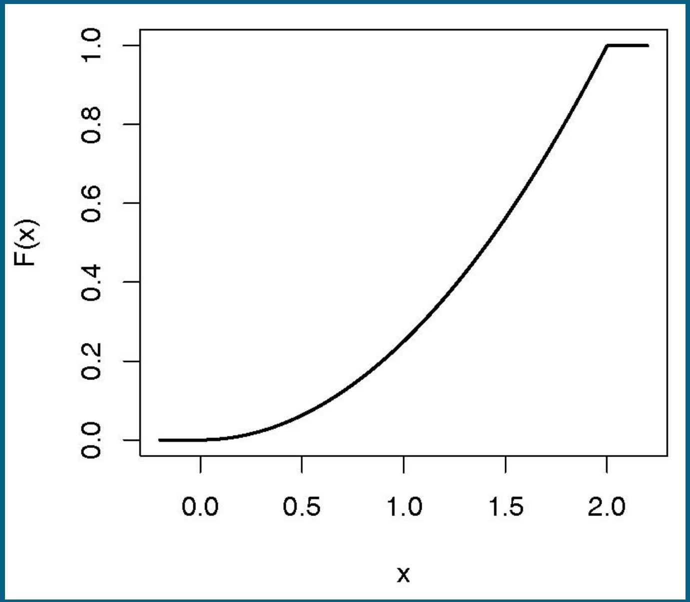
1. $F(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{x^2}{4} & para\ 0 \leq x \leq 2 \\ 1 & para\ x > 2 \end{cases}$

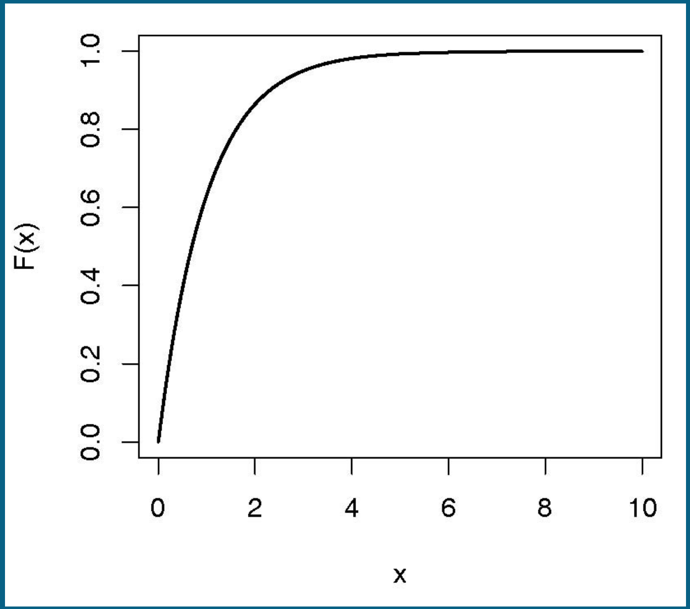
2. $F(x) = \begin{cases} 0 & para\ x < -1 \\ \frac{x}{2} & para\ -1 \leq x \leq 2 \\ 1 & para\ x > 2 \end{cases}$

3. $F(x) = \begin{cases} 0 & si\ x \leq 0 \\ 1 - e^{-x} & para\ 0 < x < \infty \end{cases}$

4. $F(x) = \begin{cases} 0 & si\ x < 0 \\ 4(x - 0.5)^2 & si\ 0 \leq x < 1 \\ 1 & si\ x \geq 1 \end{cases}$

Funciones 1 y 3 cumplen los requisitos para ser funciones de distribución. La función 2 es negativa en parte de su rango y no es una función de distribución. La función 4 da un salto en $x = 0$ y no es estrictamente continua y además, es decreciente para $0 < x < 0.5$ y entonces no puede ser una función de distribución. Los gráficos muestran las funciones 1 y 3.





La mediana y los cuantiles

Se puede utilizar la función de distribución para calcular algunas medidas de localización y de dispersión de una variable.

En primer lugar, definimos la *mediana* de una variable continua, X, como el punto $\tilde{x}$ tal que $F(\tilde{x}) = 0.5$.

Con más generalidad, definimos el $p \times 100\%$ *cuantil* como el punto x^p para que $F(x^p) = p$.

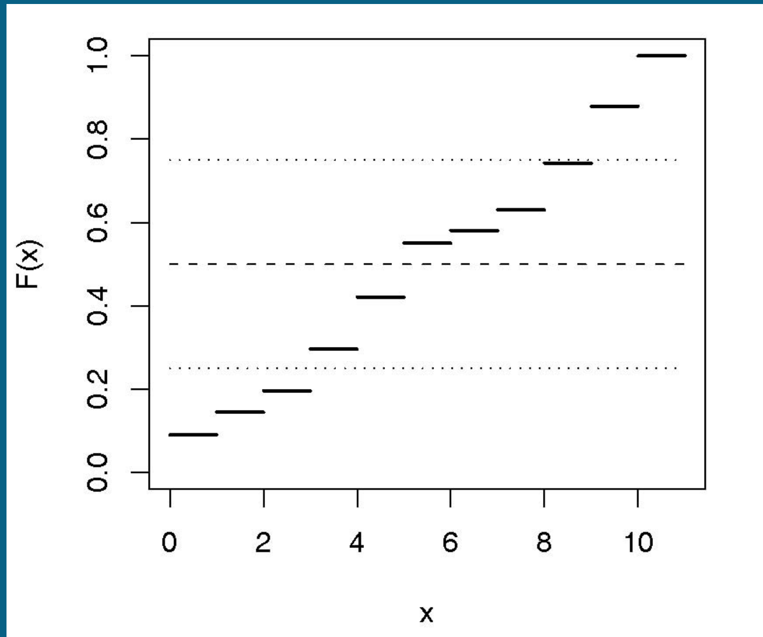
En particular, el *primer cuartil* es $Q_1 = x^{0.25}$ y el *tercer cuartil* es el punto $Q_3 = x^{0.75}$. El rango intercuartilico es $Q_3 - Q_1$.

Calculamos la mediana y los cuartiles para el ejemplo con $F(x) = x^2/4$ para $0 \leq x \leq 2$.

$$
\begin{aligned}
F\left(x^p\right) &= p \Rightarrow \\
\frac{x_p^2}{4} &= p \\
x_p &= 2\sqrt{p} \\
\tilde{x} &= 2\sqrt{0.5} \approx 1.414 \\
Q_1 &= 2\sqrt{0.25} = 1 \\
Q_3 &= 2\sqrt{0.75} \approx 1.732
\end{aligned}
$$

Mediana y cuartiles de variables discretas

En el caso de variables discretas, esta definición de la mediana no es adecuada, ya que puede ser posible de que no exista ningún valor $\tilde{x}$ tal que $F(\tilde{x}) = \frac{1}{2}$.



En este caso, se dice que la mediana es *cualquier* valor, $\tilde{x}$, tal que

$$P(X \leq x) \geq 0.5 \quad \text{y} \quad P(X \geq x) \geq 0.5.$$

En la ilustración, la mediana es cualquier valor $\tilde{x} \in [5, 6)$.

Igualmente, podemos definir los cuartiles Q_1, y Q_3 como valores que cumplen

$$P(X \leq Q_1) \geq 0.25 \quad \text{y} \quad P(X \geq Q_1) \geq 0.75$$
$$P(X \leq Q_3) \geq 0.75 \quad \text{y} \quad P(X \geq Q_3) \geq 0.25$$

Otra vez, esta definición no proporciona un valor único.

La función de densidad

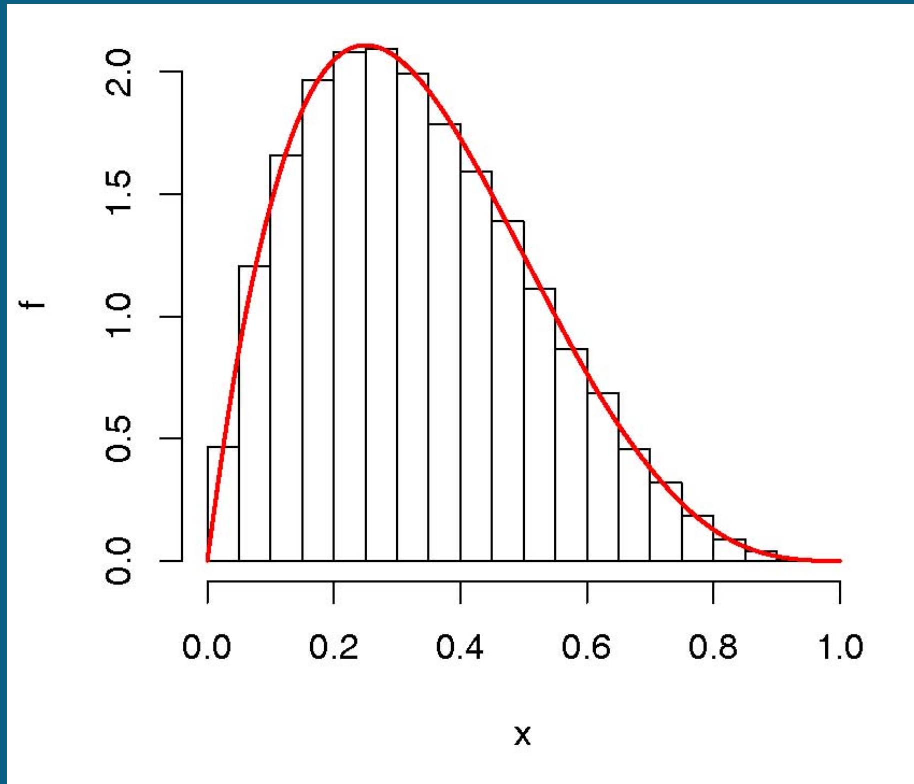
Como se observó anteriormente, no se puede definir una función de probabilidad para variables continuas. No obstante, existe una función con características parecidas.

Para una variable continua, X, con función de distribución $F(\cdot)$, se define la *función de densidad* de X como

$$f(x) = \frac{dF}{dx}.$$

Interpretación de la función de densidad

Supongamos que generamos una muestra grande de datos $x_1, x_2, \ldots, x_n$ de una variable con función de distribución F y que construimos un histograma con la área normalizada a 1.



La densidad es como el límite del polígono de frecuencias cuando la muestra es muy grande y las barras son muy finas.

Propiedades de la función de densidad

Obviamente, la función de densidad tiene las siguientes propiedades:

1. $0 \leq f(x) \; \forall \; x$.

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

3. $\int_{-\infty}^{x} f(u)\, du = F(x)$.

4. $\int_{a}^{b} f(x)\, dx = P(a < X < b)$.

Estas propiedades son semejantes a las propiedades de la función de probabilidad para una variable discreta pero observamos que la densidad no está restringida a ser menor de 1.

Ejemplos

Volvemos a las dos funciones de distribución que hemos visto antes.

$$F(x) = \begin{cases} 0 & si\ x < 0 \\ \frac{x^2}{4} & para\ 0 \leq x \leq 2 \\ 1 & para\ x > 2 \end{cases}$$

En este caso, $f(x) = 0$ si $x < 0$ o si $x > 2$, porque $\frac{d}{dx}0 = \frac{d}{dx}1 = 0$. Si $0 \leq x \leq 2$, tenemos

$$f(x) = \frac{d}{dx}\frac{x^2}{4} = \frac{2x}{4} = \frac{x}{2}.$$

$$f(x) = \begin{cases} \frac{x}{2} & \text{si } 0 \leq x \leq 2 \\ 0 & \text{en caso contrario} \end{cases}$$

$$F(x) = \begin{cases} 0 & si \ x \leq 0 \\ 1 - e^{-x} & para \ 0 < x < \infty \end{cases}$$

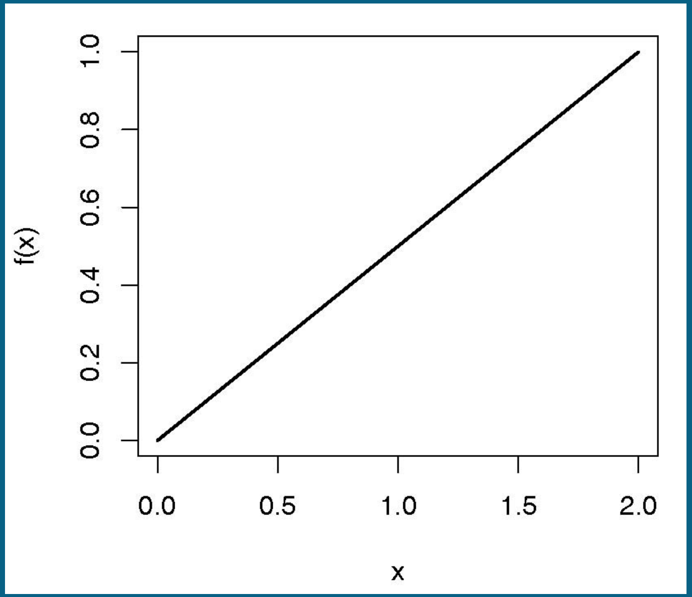
En este caso, para $x > 0$,

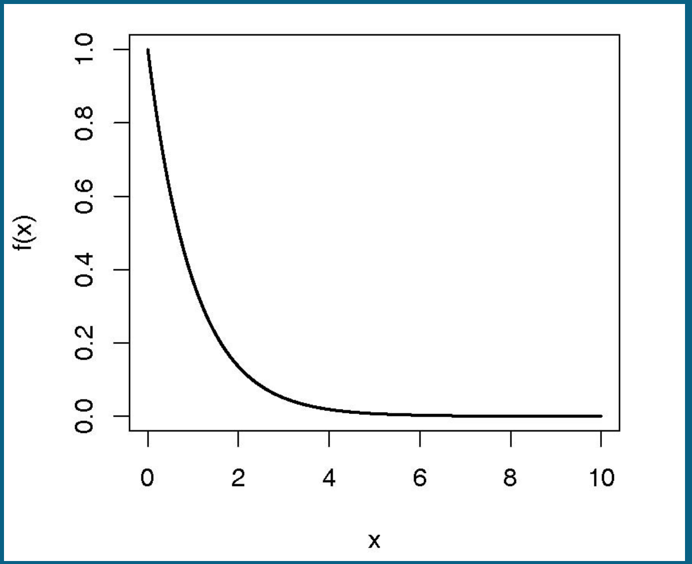
$$f(x) = \frac{d}{dx}\left(1 - e^{-x}\right) = 0 - \left(-e^{-x}\right) = e^{-x}$$

y luego

$$f(x) = \begin{cases} e^{-x} & \text{si } x > 0 \\ 0 & \text{en caso contrario} \end{cases}$$

Gráficos de la función de densidad





Ambas distribuciones son muy asimétricas.

La moda de una distribución continua

La moda de la distribución es el punto de máxima densidad. En los dos ejemplos anteriores, cuando $x \longrightarrow 2$ en el primer caso y cuando $x \longrightarrow 0$ en el segundo caso.

No obstante en la mayoría de los casos, no se encuentra la moda en un extremo de la distribución. Entonces, la moda será un punto, $\hat{x}$, donde $f'(\hat{x}) = 0$.

Ejemplo

Una variable aleatoria Y *tiene la función de densidad*

$$f(y) = \begin{cases} cy^2(1-y) & si\ 0 < y < 1 \\ 0 & si\ no \end{cases}$$

1. *¿Cuál es el valor de c?*

2. *Hallar la función de distribución de* Y*.*

3. *Calcular la moda de* Y

4. *¿Cuál es la mediana?*

1. Observamos que $\int_{-\infty}^{\infty} f(y)\,dy = 1$ y luego

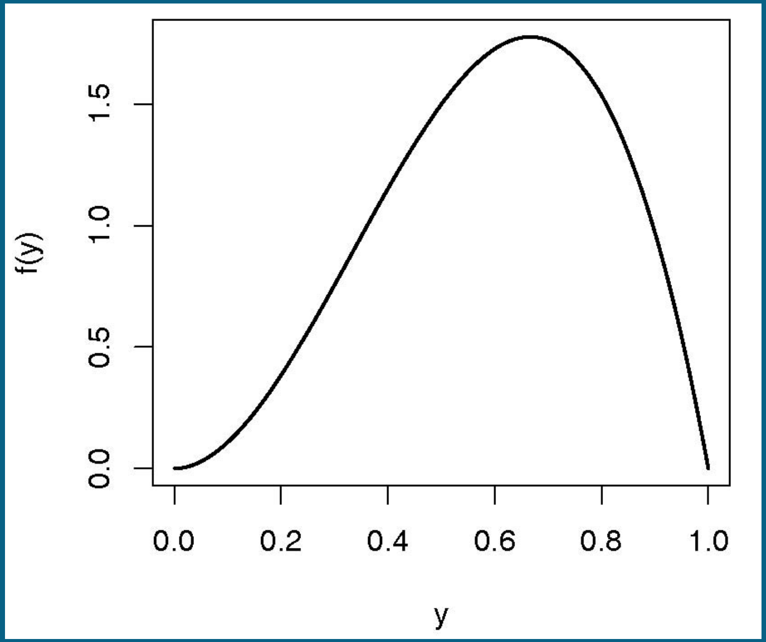
$$
\begin{aligned}
1 &= \int_{-\infty}^{0} 0\,dy + \int_{0}^{1} cy^2(1-y)\,dy + \int_{1}^{\infty} 0\,dy \\
&= c\int_{0}^{1} y^2 - y^3\,dy \\
&= c\left[\frac{y^3}{3} - \frac{y^4}{4}\right]_0^1 \\
&= c\left(\frac{1}{3} - \frac{1}{4}\right) \Rightarrow \\
&= \frac{c}{12} \\
\Rightarrow c &= 12
\end{aligned}
$$

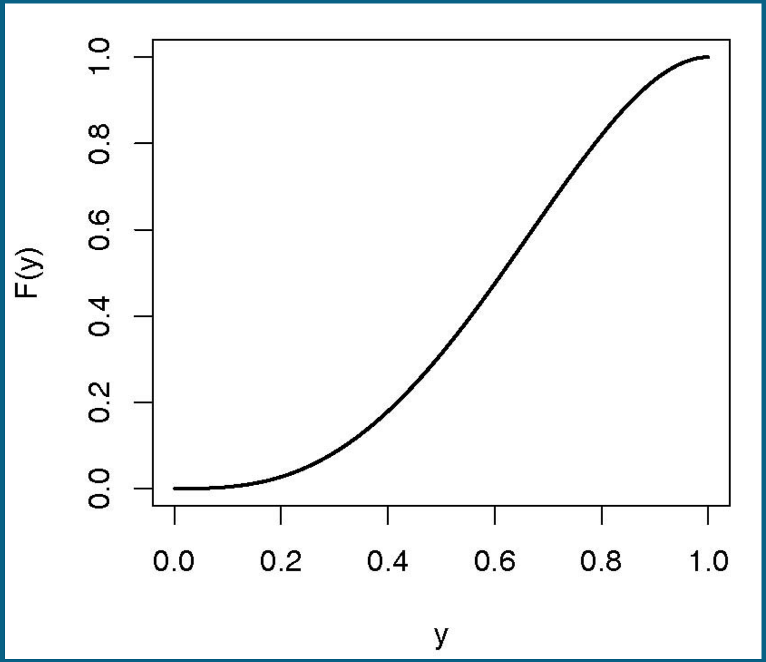
2. Tenemos $F(y) = \int_{-\infty}^{y} f(y)\,dy$. Luego, si $y \leq 0$, tenemos $F(y) = 0$ y si $y \geq 1$, tenemos $F(y) = 1$. Para $0 < y < 1$ tenemos

$$\begin{aligned} F(y) &= \int_0^y 12u^2(1-u)\,du \\ &= 12\left[\frac{u^3}{3} - \frac{u^4}{4}\right]_0^y \\ &= 4y^3 - 3y^4 \end{aligned}$$

$$F(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ 4y^3 - 3y^4 & \text{si } 0 < y < 1 \\ 1 & \text{si } y \geq 1 \end{cases}$$

Los diagramas ilustran la función de densidad y la función de distribución. Se ve que la densidad es asimétrica a la izquierda.





3.

$$
\begin{aligned}
\frac{df}{dy} &= 12(2y - 3y^2) \\
\frac{df}{dy} &= 0 \\
\Rightarrow 0 &= 2y - 3y^2 \\
\Rightarrow y &= 0 \quad \text{o} \quad \frac{2}{3}
\end{aligned}
$$

Claramente el valor 0 es un mínimo y entonces, la moda es $\frac{2}{3}$.

4. Para calcular la mediana, necesitamos resolver la ecuación $F(y) = 0.5$, es decir que

$$4y^3 - 3y^4 = 0.5.$$

Resolviéndola numéricamente en R, estimamos la moda en $\hat{y} \approx 0.615$.

Ejemplo

Sea X una variable aleatoria continua de función de densidad:

$$f(x) = \begin{cases} C(1+x^2) & \textit{si } x \in (0,3) \\ 0 & \textit{en otro caso} \end{cases}$$

a. *Hallar el valor de la constante C y la función de distribución acumulativa de probabilidad. Dibujar ambas funciones*

b. *Calcular la probabilidad de que X esté comprendido entre 0 y 1*

c. *Hallar la probabilidad de que X sea menor que 1.*

d. *¿Cuál es la moda de X?*

Ejemplo

La función de densidad de una variable aleatoria continua es:

$$f(x) = \begin{cases} ax^2 + b & si\ x \in (0,2) \\ 0 & en\ caso\ contrario \end{cases}$$

Sabiendo que $P(1/2 < x < 1) = 1/8$, *calcular:*

a. a *y* b

b. *La función de distribución.*

c. $P(1/4 < x < 3/4)$.

Momentos de una variable

Hasta ahora, se han visto dos medidas de localización de una variable;

- la moda, o el valor más probable

- la mediana, definida tal que hay una probabilidad de $(\geq)$ 50% de caer a cada lado de la mediana.

Otra medida de localización es la media.

La media

Supongamos que generamos una muestra, de una variable discreta con función de probabilidad $P(\cdot)$. En este caso, la media muestral es

$$\bar{x} = \sum_i x_i f_i$$

donde f_i es la frecuencia relativa de x_i. Si dejamos que el tamaño de la muestra crezca, entonces, recordando la definición frecuentista de la probabilidad, tenemos que $f_i \to P(X = x_i)$ para $i = 1, 2, \ldots$ y luego,

$$\bar{x} \to \sum_i x_i P(X = x_i) = E[X].$$

De manera semejante, podemos pensar en estimar la media de una muestra de datos continuos a través de un histograma cuando

$$\bar{x} \approx \sum_i x_i f_i$$

y x_i y f_i representan el centro y la proporción de datos en la barra i'ésima. Dejando el tamaño de la muestra acercarse al infinito y suponiendo que la anchura de las barras se acerca a 0, se tiene

$$\bar{x} \to \int x f(x)\, dx.$$

Para una variable discreta, X, con posibles valores $x_1, x_2, \ldots$ y función de probabilidad $P(\cdot)$, la *media* o *esperanza* de X es

$$E[X] = \sum_{i=1} x_i P(X = x_i).$$

Para una variable continua, X, con función de densidad $f(\cdot)$, la *media* es

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Es común utilizar el símbolo μ para representar la media.

Ejemplo

Volvemos al ejemplo sobre los pasajeros del avión.

x	198	199	200	201	202	203	204	205
$P(X = x)$	0.05	0.09	0.15	0.2	0.23	0.17	0.09	0.02

¿Cuál es el número medio de pasajeros que llegan al aeropuerto?

$$\begin{aligned} E[X] &= 198 \times 0.05 + 199 \times 0.09 + \cdots + 205 \times 0.23 \\ &= 201.44 \end{aligned}$$

Observamos que la media no tiene que ser uno de los valores posibles de X.

Ejemplo

Volvemos a la variable aleatoria Y con función de densidad

$$f(y) = \begin{cases} 12y^2(1-y) & si\ 0 < y < 1 \\ 0 & si\ no \end{cases}$$

En este caso,

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y f(y)\, dy \\ &= \int_0^1 y \times 12y^2(1-y)\, dy \\ &= 12 \int_0^1 y^3 - y^4\, dy \\ &= 12 \left[\frac{y^4}{4} - \frac{y^5}{5} \right]_0^1 = \frac{3}{5} \end{aligned}$$

Ejemplo

Calcular la media y la varianza para una variable continua con densidad

$$f(x) = \begin{cases} \frac{1+x^2}{12} & si\ x \in (0,3) \\ 0 & en\ otro\ caso \end{cases}$$

Cuando la media no existe

Si una variable discreta está definida sobre un conjunto finito de valores, por ejemplo $x_1 \leq x_2 \leq \cdots \leq x_n$, entonces su media siempre existe, ya que

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i) \leq \sum_{i=1}^{n} x_n P(X = x_i) = x_n.$$

Igualmente, si una variable continua tiene densidad positiva en un intervalo finito $[a, b]$, se tiene

$$E[X] = \int_a^b x f(x)\, dx \leq \int_a^b b f(x)\, dx = b.$$

En caso contrario, es posible que la media no exista.

Ejemplos

Sea X una variable discreta con función de probabilidad

$$P(X = x) = \frac{e^{-1}}{x!} \quad \textit{para } x = 0, 1, 2, \ldots, \infty.$$

¿Cuál es la media de X?

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} x\frac{e^{-1}}{x!} \\ &= \sum_{x=1}^{\infty} \frac{e^{-1}}{(x-1)!} \\ &= \sum_{y=0}^{\infty} \frac{e^{-1}}{y!} = 1 \end{aligned}$$

Supongamos que X es una variable discreta con función de probabilidad

$$P(X = x) = \frac{6}{\pi^2 x^2} \quad \textit{para } x = 1, 2, \ldots, \infty.$$

Calculamos la media de X.

$$E[X] = \sum_{x=1}^{\infty} x \times \frac{6}{\pi^2 x^2} = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{1}{x} = \infty$$

Sea X una variable continua con

$$f(x) = \frac{1}{\pi(1 + x^2)} \quad \textit{para } -\infty < x < \infty.$$

$$E[X] = \int_{-\infty}^{\infty} \frac{x}{\pi(1 + x^2)} = \infty$$

Esperanza de una función

Supongamos que X es discreta y sea $g(X)$ una función de X. Luego la esperanza de $g(X)$ es

$$E[g(X)] = \sum_i P(X = x_i) \times g(x_i).$$

Igualmente, si X es continua, definimos

$$E[g(X)] = \int g(x) f(x) \, dx$$

Ejemplo

En el ejemplo sobre los pasajeros supongamos que la compañía aérea recibe 250 euros por cada billete que vende pero que tiene que devolver el precio del ticket y además pagar una multa de 1000 euros a cada pasajero que no puede montar en el avión. Calcular la cantidad de dinero que espera cobrar la compañía en este vuelo.

Sean $g(X)$ las ganancias de la compañía. Las ventas totales de tickets son $250 \times 205 = 51250$ euros. Si llegan $x \leq 200$ personas entonces $g(x) = 51250$. Si llegan $x > 200$ personas, $g(x) = 51250 - (x - 200) \times (1250)$. Entonces

$$
\begin{aligned}
E[g(X)] &= 51250 \times .05 + 51250 \times .09 + 51250 \times .15 + \\
&\quad (51250 - (201 - 200) * 1250) \times .20 + \\
&\quad (51250 - (202 - 200) * 1250) \times .23 + \\
&\quad \ldots \\
&\quad +(51250 - (205 - 200) * 1250) \times .02 \\
&= 49212.5 \text{ euros}
\end{aligned}
$$

Propiedades de la esperanza

Teorema 8

Para una variable, X, constantes b y c y funciones g y h, se tiene

$$\begin{aligned} E[c] &= c \\ E[bg(X)] &= bE[g(X)] \\ E[g(X)+h(X)] &= E[g(X)]+E[h(X)] \end{aligned}$$

Demostración Supongamos que X es continua.

$$\begin{aligned}
E[c] &= \int_{-\infty}^{\infty} cf(x)\,dx = c\int_{-\infty}^{\infty} f(x)\,dx \\
&= c \times 1 = c \\
E[bg(X)] &= \int_{-\infty}^{\infty} bg(x)f(x)\,dx = b\int_{-\infty}^{\infty} g(x)f(x)\,dx \\
&= bE[g(X)] \\
E[g(X)+h(X)] &= \int_{-\infty}^{\infty} (g(x)+h(x))f(x)\,dx \\
&= \int_{-\infty}^{\infty} g(x)f(x)+h(x)f(x)\,dx \\
&= \int_{-\infty}^{\infty} g(x)f(x)\,dx + \int_{-\infty}^{\infty} h(x)f(x)\,dx \\
&= E[g(X)] + E[h(X)].
\end{aligned}$$

$\square$

El resultado para una función lineal de X es inmediato.

Corolario 9

$E[a + bX] = a + bE[X]$.

En el ejemplo con

$$f(y) = \begin{cases} 12y^2(1-y) & si\ 0 < y < 1 \\ 0 & si\ no \end{cases}$$

supongamos que queremos calcular la esperanza de $5Y - 2$.

$$\begin{aligned} E[5Y-2] &= 5E[Y] - 2 \\ &= 5 \times \frac{3}{5} - 2 = 1 \end{aligned}$$

La varianza y la desviación típica

Una función importante es la *varianza*

$$V[X] = E\left[(X - E[X])^2\right]$$

que es una medida de dispersión de la distribución (la distancia cuadrada media de una observación de la media de la distribución).

A menudo, es más útil usar la *desviación típica*

$$DT[X] = \sqrt{V[X]}$$

que tiene las mismas unidades de X.

Se utiliza el símbolo σ^2 para representar la varianza y σ para la distribución típica.

Calculando la varianza

Teorema 10

$$V[X] = E\left[X^2\right] - E[X]^2$$

Demostración

$$\begin{aligned}
V[X] &= E\left[(X - E[X])^2\right] \\
&= E\left[X^2 - 2XE[X] + E[X]^2\right] \\
&= E\left[X^2\right] - E[2E[X]X] + E\left[E[X]^2\right] \\
&= E\left[X^2\right] - 2E[X]E[X] + E[X]^2 \\
&= E\left[X^2\right] - E[X]^2
\end{aligned}$$

□

Ejemplos

En el ejemplo sobre los pasajeros,

$$
\begin{aligned}
E[X^2] &= .05 \times 198^2 + \ldots + .02 \times 205^2 = 40580.88 \\
\sigma^2 &= E[X^2] - \mu^2 = 40580.88 - 201.44^2 \\
&= 2.8064 \\
\sigma &\approx 1.675 \ \text{ pasajeros}
\end{aligned}
$$

En el ejemplo continuo

$$
\begin{aligned}
E[Y^2] &= \int_0^1 y^2 \times 12y^2(1-y)\, dy = 12 \int_0^1 y^4 - y^5\, dy \\
&= 12 \left[\frac{y^5}{5} - \frac{y^6}{6}\right]_0^1 = \frac{12}{30} = \frac{2}{5} \\
V[Y] &= \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25} \quad \Rightarrow DT[Y] = \frac{1}{5}
\end{aligned}
$$

El coeficiente de variación

No es del todo natural tener una medida de dispersión que depende de las unidades de la variable. Por este razón, se define el coeficiente de variación de una variable X como

$$CV[X] = \frac{DT[X]}{|E[X]|}.$$

Esta medida no tiene sentido para una variable con media 0.

En el ejemplo anterior, tenemos que $CV[Y] = \frac{\frac{1}{5}}{\frac{3}{5}} = \frac{1}{3}$.

Estandarizando una variable aleatoria

Para una variable aleatoria, X, con media μ y varianza σ^2 se tiene que

$$E\left[\frac{X-\mu}{\sigma}\right] = 0 \quad \text{y} \quad V\left[\frac{X-\mu}{\sigma}\right] = 1.$$

La variable $Y = \frac{X-\mu}{\sigma}$ es una variable estandarizada.

Demostración Ejercicio. $\square$

Desigualdades

Para una variable con media μ, pensaríamos que la mayoría de observaciones que generamos a partir de la distribución estarían cerca de μ. Además, si la desviación típica es más alta, hay más probabilidad de estar más lejos de μ.

Buscamos un resultado que formaliza este idea.

La desigualdad de Markov

Teorema 11
Sea X una variable aleatoria no negativa tal que $E[X]$ existe. Entonces, para cualquier $c > 0$,

$$P(X \geq c) \leq \frac{E[X]}{c}.$$

Demostración Supongamos que X es discreta.

$$\begin{aligned}
E[X] &= \sum_{x} xP(X = x)\,dx \\
&= \sum_{x<c} xP(X = x)\,dx + \sum_{x \geq c} xP(X = x)\,dx \\
&\geq \sum_{x \geq c} xP(X = x)\,dx \\
&\geq \sum_{x \geq c} cP(X = x)\,dx = cP(X \geq c)
\end{aligned}$$

$\square$

La desigualdad de Chebyshev

Teorema 12

Sea $\mu = E[X]$ y $\sigma^2 = V[X]$. Entonces para $c > 0$, se tiene

$$P\left(|X-\mu| \geq c\right) \leq \frac{\sigma^2}{c^2}.$$

Demostración

$$\begin{aligned} P\left(|X-\mu| \geq c\right) &= P\left((X-\mu)^2 \geq c^2\right) \\ &\leq \frac{E\left[(X-\mu)^2\right]}{c^2} \quad \text{por la desigualdad de Markov} \\ &= \frac{V[X]}{c^2} \quad \text{por la definición de } V[X] \\ &= \frac{\sigma^2}{c^2} \end{aligned}$$

$\square$

Luego, se tiene que para una variable cualquiera con media μ y desviación típica σ, se tiene

$$\begin{aligned} P(|X-\mu| \geq 2\sigma) &\leq \frac{1}{4} \\ P(|X-\mu| \geq 3\sigma) &\leq \frac{1}{9} \end{aligned}$$

La desigualdad de Chebyshev es muy conservadora.

Volvemos al ejemplo continuo

$$f(y) = 12y^2(1-y) \quad E[Y] = \frac{3}{5} \quad V[Y] = \frac{1}{25}$$

$$
\begin{aligned}
P\left(\left|Y-\frac{3}{5}\right| \geq 2\sqrt{\frac{1}{25}}\right) &= 1-P\left(\left|Y-\frac{3}{5}\right| \leq \frac{2}{5}\right) \\
&= 1-P\left(\frac{3}{5}-\frac{2}{5} \leq Y \leq \frac{3}{5}+\frac{2}{5}\right) \\
&= 1-P\left(\frac{1}{5} \leq Y \leq 1\right) \\
&= P\left(Y \leq \frac{1}{5}\right) \\
&= \int_0^{\frac{1}{5}} 12y^2(1-y)\,dy \\
&= 12\left[\frac{y^3}{3}-\frac{y^4}{4}\right]_0^{\frac{1}{5}} = 0.0272 << \frac{1}{4}
\end{aligned}
$$

Ejemplo

Supongamos que $P(X=-1)=\frac{1}{8}$, $P(X=0)=\frac{3}{4}$ y $P(X=1)=\frac{1}{8}$. Luego,

$$\begin{aligned}
\mu &= -1\times\frac{1}{8}+0\times\frac{3}{4}+1\times\frac{1}{8}=0\\
\sigma^2 &= E\left[X^2\right]\\
&= (-1)^2\times\frac{1}{8}+0^2\times\frac{3}{4}+1^2\times\frac{1}{8}=\frac{1}{4}
\end{aligned}$$

Ahora,

$$\begin{aligned}
P\left(|X-\mu|\geq 2\sigma\right) &= P(|X|\geq 1)\\
&= P(X=-1)+P(X=1)=\frac{1}{4}
\end{aligned}$$

que es exactamente el límite de la desigualdad de Chebyshev.

Momentos, asimetría y curtosis

Además de $E[X]$ y $E[X^2]$, se puede definir esperanzas de orden más alto. Formalmente, se dice que para $k \in \mathbb{N}$, el *momento de orden* k es $E\left[X^k\right]$.

El *momento central de orden* k es $E\left[(X-\mu)^k\right]$.

En particular, se ha visto que el momento central de orden 2 es la varianza que es una medida de dispersión. El momento central de orden 3 es una medida de la *asimetría* de la distribución y el momento central de orden 4 mide el *apuntamiento* o *curtosis* de la distribución. Formalmente:

$$\frac{1}{\sigma^3}E\left[(X-\mu)^3\right] \qquad \text{es el coeficiente de asimetría}$$

$$\frac{1}{\sigma^4}E\left[(X-\mu)^4\right] \qquad \text{es el coeficiente de curtosis}$$

Funciones generadoras

En muchas situaciones, no es tan fácil calcular la media, varianza y otros momentos de una variable directamente.

Por ejemplo, para una variable binomial

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{para } x = 0, 1, \ldots, n$$

la media es

$$E[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

y tenemos que trabajar bastante para resolver el sumatorio. Además, si queremos calcular la varianza, asimetría etc., tenemos que resolver más sumatorios.

No obstante, a menudo se pueden usar funciones generadoras para derivar los momentos indirectamente.

La función generadora de probabilidades

Supongamos que tenemos una variable discreta, X con función de probabilidad P y valores posibles $0, 1, 2, \ldots$. Entonces, la *función generadora de probabilidades* de X es

$$G(s) = E\left[s^X\right] = \sum_{x=0}^{\infty} s^x P(X = x).$$

Por la definición de la función, es fácil de ver que:

- $G(1) = 1$,
- $G(0) = P(X = 0)$.

pero el resultado más útil es ...

... el siguiente teorema que nos proporciona una manera de calcular la media y otros momentos de la variable.

Teorema 13

$$\begin{aligned} E[X] &= \left.\frac{dG}{ds}\right|_{s=1} \\ E[X(X-1)] &= \left.\frac{dG}{ds^2}\right|_{s=1} \\ E[X(X-1)\cdots(X-k+1)] &= \left.\frac{d^kG}{ds^k}\right|_{s=1} \end{aligned}$$

Demostración

$$
\begin{aligned}
\frac{dG}{ds} &= \frac{d}{ds}E\left[s^X\right] = E\left[\frac{d}{ds}s^X\right] = E\left[Xs^{X-1}\right] \\
\left.\frac{dG}{ds}\right|_{s=1} &= E\left[X \times 1^{X-1}\right] = E[X] \\
\frac{d^2G}{ds^2} &= E\left[X(X-1)s^{X-2}\right] \\
\left.\frac{d^2G}{ds^2}\right|_{s=1} &= E[X(X-1)] \\
\frac{d^kG}{ds^k} &= E\left[X(X-1)\cdots(X-k+1)s^{X-k}\right] \\
\left.\frac{d^kG}{ds^k}\right|_{s=1} &= E\left[X(X-1)\cdots(X-k+1)\right]
\end{aligned}
$$

$\square$

Corolario 14

$$V[X] = G''(1) + G'(1) - G'(1)^2$$

Demostración

$$\begin{aligned} V[X] &= E\left[X^2\right] - E[X]^2 \\ &= E\left[X^2 - X + X\right] - E[X]^2 \\ &= E[X(X-1)] + E[X] - E[X]^2 \end{aligned}$$

que es la fórmula para la varianza. $\square$

Ejemplo

Supongamos que X es una variable binomial con función de probabilidad

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \textit{para } x = 0, 1, \ldots, n.$$

Calcular la función generadora de probabilidades y la media y varianza de X.

Observamos primero que para dos números, a y b, se tiene

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$

Luego:

$$
\begin{aligned}
G(s) &= E\left[s^X\right] = \sum_{x=0}^{n} s^x \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \\
&= (sp + 1 - p)^n \quad \text{y derivando,} \\
\frac{dG}{ds} &= np(sp + 1 - p)^{n-1} \\
\left.\frac{dG}{ds}\right|_{s=1} &= np = E[X] \\
\frac{d^2G}{ds^2} &= n(n-1)p^2(sp + 1 - p)^{n-2} \\
\left.\frac{d^2G}{ds^2}\right|_{s=1} &= n(n-1)p^2 = E[X(X-1)] \\
V[X] &= n(n-1)p^2 + np - (np)^2 \\
&= np(1-p)
\end{aligned}
$$

La función generadora de momentos

Sólo se utiliza la función generadora de probabilidades para variables discretas y no negativas. Una función más general, que se puede utilizar para cualquiera variable es la *función generadora de momentos* definido como

$$M(s) = E\left[e^{sX}\right].$$

Observamos que si X es una variable discreta y no negativa, entonces

$$M(s) = E\left[e^{sX}\right] = E\left[\left(e^{s}\right)^{X}\right] = G\left(e^{s}\right)$$

donde G es la función generadora de probabilidades de X.

Igual que con la función generadora de probabilidades, se puede utilizar esta función para calcular los momentos.

Teorema 15

$$\begin{aligned} \left.\frac{dM}{ds}\right|_{s=0} &= E[X] \\ \left.\frac{d^2M}{ds^2}\right|_{s=0} &= E\left[X^2\right] \\ \left.\frac{d^kM}{ds^k}\right|_{s=0} &= E\left[X^k\right] \end{aligned}$$

Demostración

$$
\begin{aligned}
\frac{dM}{ds} &= \frac{d}{ds} E\left[e^{sX}\right] \\
&= E\left[\frac{d}{ds} e^{sX}\right] \\
&= E\left[X e^{sX}\right] \\
\left.\frac{dM}{ds}\right|_{s=0} &= E\left[X e^{0}\right] = E[X]
\end{aligned}
$$

Los otros resultados siguen de la misma manera. □

Ejemplo

Sea X una variable exponencial con función de densidad

$$f(x) = \lambda e^{-\lambda x} \quad \textit{para } x > 0.$$

Calcular su función generadora de momentos y su media y varianza.

$$\begin{aligned} M_X(s) &= E\left[e^{sX}\right] \\ &= \int_0^\infty e^{sx} \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{-(\lambda - s)x}\, dx \\ &= \frac{\lambda}{\lambda - s} \quad \text{para } s < \lambda. \end{aligned}$$

Luego:

$$\begin{aligned}
\frac{dM}{ds} &= \frac{\lambda}{(\lambda - s)^2} \\
\left.\frac{dM}{ds}\right|_{s=0} &= \frac{\lambda}{\lambda^2} \\
E[X] &= \frac{1}{\lambda} \\
\frac{d^2M}{dM^2} &= \frac{2\lambda}{(\lambda - s)^3} \\
E\left[X^2\right] &= \frac{2}{\lambda^2} \\
V[X] &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}
\end{aligned}$$

Transformaciones de variables

Si X es una variable discreta e $Y = g(X)$ es una transformación, donde la función g es bijectiva, se calcula la función de probabilidad de Y mediante

$$P(Y = y) = P(g(X) = y) = P\left(X = g^{-1}(y)\right).$$

Si g no es bijectiva, entonces tendremos que obtener todas las soluciones de $g(x) = y$.

Por ejemplo, si X toma valores en $\mathbb{Z}$ y $g(x) = x^2$ entonces tendremos que $x = \sqrt{y}$ y $x = -\sqrt{y}$ que son soluciones de $x^2 = y$ y luego,

$$P(Y = y) = P(X = \sqrt{y}) + P(X = -\sqrt{y}).$$

Variables continuas

Para variables continuas, hallar la densidad de la variable transformada es más complicada. En primer lugar, si $Y = g(X)$ es una *transformación monótona creciente*, se calcula la función de distribución de Y mediante

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P\left(X \leq g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right).$$

Si g es una *función monótona decreciente*, tenemos

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P\left(X \geq g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right).$$

Derivando, se puede obtener una expresión para la densidad de Y. En el primer caso,

$$f_Y(y) = \frac{dg^{-1}(y)}{dy} f_X\left(g^{-1}(y)\right) = f_X\left(g^{-1}(y)\right)\frac{dx}{dy}$$

y en el segundo caso,

$$f_Y(y) = -\frac{dg^{-1}(y)}{dy} f_X\left(g^{-1}(y)\right) = -f_X\left(g^{-1}(y)\right)\frac{dx}{dy}$$

es decir que para cualquier función monótona, g, entonces se tiene

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X\left(g^{-1}(y)\right).$$

Ejemplo

Sea X *una variable con distribución exponencial,*

$$f_X(x) = e^{-x}$$

y definimos $Y = \log X$*. ¿Cuál es la densidad de* Y*?*

$$y = \log x \Rightarrow x = e^y \Rightarrow \frac{dx}{dy} = e^y$$

y luego, la densidad de Y es

$$f_Y(y) = e^{-e^y} \times |e^y| = e^{y-e^y}$$

para $-\infty < y < \infty$.

Calcular la probabilidad de que $Y < 0$.

Suena horrible, ya que necesitamos calcular

$$\int_{-\infty}^{0} e^{y-e^{y}}\, dy$$

pero podemos convertir la pregunta en una pregunta sobre X.

$$\begin{aligned} P(Y<0) &= P(\log X < 0) \\ &= P\left(X < e^{0}\right) = P(X<1) \\ &= 1 - e^{-1} \end{aligned}$$

Transformaciones lineales

Teorema 16
Sea $Y = a + bX$ donde X es una variable continua. Luego,

$$
\begin{aligned}
f_Y(y) &= \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right) \\
F_Y(y) &= \begin{cases} F_X\left(\frac{y-a}{b}\right) & \text{si } b > 0 \\ 1 - F_X\left(\frac{y-a}{b}\right) & \text{si } b < 0 \end{cases} \\
E[Y] &= a + bE[X] \\
V[Y] &= b^2 V[X]
\end{aligned}
$$

Demostración Ejercicio. $\square$