Variables Aleatorias

Introducción

Hasta ahora nuestros sucesos han sido de varios tipos: $\{C, +\}$ en la moneda, nombres de periódicos, ángulos en una ruleta, número de veces que sale cara en el lanzamiento de una moneda etc. . . .
Necesitamos estandarizar de alguna manera todos estos sucesos. Una solución es asignar a cada suceso un cierto conjunto de números reales, es decir, convertir todos los sucesos en sucesos de números reales para trabajar con ellos de forma unificada.
Para conseguirlo utilizaremos unas funciones que transformen los elementos del espacio muestral en números; esta funciones son las variables aleatorias.

Definición de variable aleatoria

Comenzaremos dando una definición práctica de variable aleatoria.

Definición

Definición práctica de variable aleatoria (v.a.) es una aplicación que toma valores numéricos determinados por el resultado de un experimento aleatorio

$$X : \Omega \to \mathbb{R}$$

Notación:

- Normalmente representaremos las v.a. por letras mayúsculas $X, Y, Z \ldots$
- Los valores que "toman" las v.a. los representaremos por letras minúsculas (las mismas en principio) $x, y, z \ldots$

Ejemplo

Lanzamos un dado convencional de parchís el espacio muestral del experimento es

$$\Omega = \{\epsdice{1}, \epsdice{2}, \epsdice{3}, \epsdice{4}, \epsdice{5}, \epsdice{6}\}$$

y una v.a $X : \Omega \to \mathbb{R}$ sobre este espacio queda definida por

$$\begin{aligned} X(\epsdice{1}) = 1, X(\epsdice{2}) = 2, X(\epsdice{3}) = 3,\\ X(\epsdice{4}) = 4, X(\epsdice{5}) = 5, X(\epsdice{6}) = 6. \end{aligned}$$

Ahora el suceso $A = \{\epsdice{2}, \epsdice{4}, \epsdice{6}\}$, es decir "salir número par", es equivalente a $\{X = 2, X = 4, X = 6\}$.
El suceso $B = \{\epsdice{1}, \epsdice{2}, \epsdice{3}\}$, es decir "salir un número inferior o igual a 3" es en términos de la v.a. $\{X = 1, X = 2, X = 3\}$ o también $\{X \leqslant 3\}$.

Ejemplo

Consideremos el experimento lanzar una anilla al cuello de una botella. Si acertamos a ensartar la anilla en la botella el resultado del experimento es éxito y fracaso en caso contrario.
El espacio muestral asociado a este experimento será $\Omega = \{\text{éxito, fracaso}\}$. Construyamos la siguiente variable aleatoria:

$$X : \{\text{éxito, fracaso}\} \to \mathbb{R}$$

definida por

$$X(\text{éxito}) = 1 \text{ y } X(\text{fracaso}) = 0.$$

Hay dos tipos fundamentales de variables aleatorias, las discretas y las continuas. Damos a continuación una definición informal de estos tipos.

Definición

a) Una variable aleatoria es discreta si sólo puede tomar una cantidad numerable de valores con probabilidad positiva.

b) La variables aleatorias continuas toman valores en intervalos.

c) Variables aleatorias mixtas; con una parte discreta y otra continua.

Ejemplo

Son variables aleatorias discretas:

- Número de artículos defectuosos en un cargamento.
- Número de clientes que llegan a una ventanilla de un banco en una hora.
- Número de errores detectados en las cuentas de una compañía.
- Número de reclamaciones de una póliza de un seguro médico.

Son variables aleatorias continuas:

- Renta anual de una familia.
- Cantidad de petróleo importado por un país
- Variación del precio de las acciones de una compañía de telecomunicaciones.
- Porcentaje de impurezas en un lote de productos químicos.

Distribuciones de probabilidad para v.a. discretas.

- Pasamos ahora a describir el comportamiento de la v.a. Para ello utilizaremos distintas funciones que nos darán algunas probabilidades de la variable aleatoria.
- En el caso discreto estas funciones son la de probabilidad, y la función de distribución o de probabilidad acumulada.
- En el caso discreto la función de probabilidad es la que nos da las probabilidades de los sucesos elementales de la v.a. que definimos a continuación.

Definición

La función de probabilidad (*probability mass function* o incluso abusando de notación *probability density function*) de una variable aleatoria discreta X a la que denotaremos por $P_X(x)$ está definida por

$$P_X(x) = P(X = x)$$

es decir la probabilidad de que X tome el valor x. Si X no asume ese valor entonces $P_X(x) = 0$.
El conjunto

$$D_X = \{x \in \mathbb{R} \mid P_X(x) > 0\}$$

recibe el nombre de dominio de la v.a. y son los valores posibles de esta variable.
En el caso discreto lo más habitual es que $X(\Omega) = D_X$.

Ejemplo

Lanzamos un dado de parchís una vez, en esta ocasión representaremos los sucesos elementales por el número de puntos de la cara obtenida, tenemos que

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

y la variables aleatoria $X : \Omega \to \mathbb{R}$ viene definida por

$$X(i) = i \text{ para } i = 1, 2, 3, 4, 5, 6.$$

Supongamos que el dado está bien balanceado. Entonces

$$P_X(1) = P_X(2) = P_X(3) = P_X(4) = P_X(5) = P_X(6) = \frac{1}{6}$$

Concretamente:

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{si } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{en otro caso} \end{cases}$$

Su dominio es

$$D_X = \{1, 2, 3, 4, 5, 6\}.$$

Ejemplo

Sea X la v.a. asociada al lanzamiento de una moneda. SU espacio muestra es $\Omega = \{c, +\}$, la v.a. queda definida por:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = c \\ 0 & \text{si } \omega = + \end{cases}$$

entonces su función de probabilidad es:

$$P_X(x) = P(X = x) = \begin{cases} \frac{1}{2} & \text{si } x = 0, 1 \\ 0 & \text{en otro caso} \end{cases}$$

Finalmente su dominio es $D_X = \{0, 1\}$.

Ejemplo

Tenemos una urna con tres bolas rojas,una negra y dos blancas. Realizamos una extracción y observamos el color de la bola entonces un espacio muestral es

$$\Omega = \{roja, blanca, negra\}.$$

Una variable aleatoria asociada al experimento es:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega = roja \\ 2 & \text{si } \omega = negra \\ 3 & \text{si } \omega = blanca \end{cases}$$

entonces

$$P_X(x) = \begin{cases} \frac{3}{6} & \text{si } x = 1 \\ \frac{1}{6} & \text{si } x = 2 \\ \frac{2}{6} & \text{si } x = 3 \\ 0 & \text{en otro caso} \end{cases}$$

El dominio de X es $D_X = \{1, 2, 3\}$.

Propiedades de la función de probabilidad.

Sea X una v.a. discreta $X:\Omega :\to \mathbb{R}$ con dominio D_X. Su función de probabilidad P_X verifica las siguientes propiedades:

a) $0\leqslant P_X(x)\leqslant 1$ para todo $x\in\mathbb{R}$

b) $\displaystyle\sum_{x\in D_X} P_X(x)=1$

Ejemplo

Lanzamos al aire tres veces, de forma independiente, una moneda perfecta. El espacio muestral de este experimento es

$$\Omega = \{ccc, cc+, c+c, +cc, c++, +c+, ++c, +++\}$$

(expresados en orden de aparición).
Este espacio tiene todos los sucesos elementales equiprobables.
Consideremos la variable aleatoria asociada a este experimento $X =$ número de caras en los tres lanzamientos.

Ejemplo

Entonces

$$
\begin{aligned}
P(X=0) &= P(\{+++\}) = \tfrac{1}{8} \\
P(X=1) &= P(\{c++, +c+, ++c\}) = \tfrac{3}{8} \\
P(X=2) &= P(\{cc+, c+c, +cc\}) = \tfrac{3}{8} \\
P(X=3) &= P(\{ccc\}) = \tfrac{1}{8}
\end{aligned}
$$

La función de probabilidad de X es:

$$
P_X(x) = \begin{cases} \frac{1}{8} & \text{si } x = 0, 3 \\ \frac{3}{8} & \text{si } x = 1, 2 \\ 0 & \text{en otro caso} \end{cases}
$$

Función de distribución.

Definición

La función de distribución de probabilidad (acumulada) de la v.a. X (de cualquier tipo; discreta o continua) $F_X(x)$ representa la probabilidad de que X tome un menor o igual que x es decir

$$F_X(x) = P(X \leqslant x)$$

Esta función también se denomina función de distribución de probabilidad o simplemente función de distribución de una v.a., y en inglés cumulative distribution function por lo que se abrevia con el acrónimo cdf.

Propiedades

Sea X una v.a. y F_X su función de distribución:

1) $P(X > x) = 1 - P(X \leqslant x) = 1 - F_X(x)$
2) Sea a y b tales que $a < b$,
   $P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) = F_X(b) - F_X(a)$

Demostración:

1) $\overline{\{X > x\}} = \{X \leqslant x\}$.
$P(X > x) = 1 - P(\overline{\{X > x\}}) = 1 - P(X \leqslant x) = 1 - F_X(x)$

2) $\{a < X \leqslant b\} = \{X \leqslant b\} - \{X \leqslant a\}$
$P(a < X \leqslant b) = P(\{X \leqslant b\} - \{X \leqslant a\}) =$
$P(\{X \leqslant b\}) - P(\{X \leqslant a\}) = F_X(b) - F_X(a)$.

Propiedades

Sea F_X la función de distribución de una v.a. X entonces:

a) $0 \leqslant F_X(x) \leqslant 1$.

b) La función F_X es no decreciente.

c) La función F_X es continua por la derecha.

d) Si denotamos por $F_X(x_0^-) = \lim\limits_{x \to x_0^-} F(x)$, entonces se cumple que $P(X < x_0) = F_X(x_0^-)$ y que $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$.

e) Se cumple que $\lim\limits_{x \to \infty} F_X(x) = 1$; $\lim\limits_{x \to -\infty} F_X(x) = 0$.

f) Toda función F verificando las propiedades anteriores es función de distribución de alguna v.a. X.

g) $P(X > x) = 1 - F_X(x)$

h) Dados $a, b \in \mathbb{R}$ con $a < b$

$$P(a < X \leqslant b) = F_X(b) - F_X(a).$$

Propiedades

En las propiedades anteriores no se pueden cambiar en general las desigualdades de estrictas o no estrictas, veamos que propiedades tenemos cuando se cambian estas desigualdades. Sea F_X una función de distribución de la v.a. X y denotamos por $F_X(x_0^-) = \lim_{x \to x_0^-} F_X(x)$, entonces.

a) $P(X = x) = F_X(x) - F_X(x^-)$
b) $P(a < X < b) = F_X(b^-) - F_X(a)$
c) $P(a \leqslant X < b) = F_X(b^-) - F_X(a^-)$
d) $P(X < a) = F_X(a^-)$
e) $P(a \leqslant X \leqslant b) = F_X(b) - F_X(a^-)$
f) $P(X \geqslant a) = 1 - F_X(a^-)$

Propiedades

a) Si F_X es continua en x se tiene que $P(X = x) = 0$. Así que si la v.a. es continua $P(X \leqslant a) = P(X < a) + P(X = a) = P(X < a)$ y propiedades similares.

b) Sea X una variable aleatoria discreta que con dominio D_X y que tiene por función de probabilidad $P_X(x)$ entonces su función de distribución $F_X(x_0)$ es

$$F_X(x_0) = \sum_{x \leqslant x_0} P_X(x)$$

donde $\sum_{x \leqslant x_0}$ indica que sumamos todos los $x \in D_X$ tales que $x \leqslant x_0$

Demostración:

a) Si X es continua

$$P(X=a)=F(a)-F(a^{-})=F(a)-F(a)=0$$

por lo tanto

$$\begin{aligned} P(X\leqslant a) &= P(X<a)+P(X=a) \\ &= P(X<a)+0=P(X<a). \end{aligned}$$

b)

$$\begin{aligned} F_X(x_0) &= P(X\leqslant x_0)=P\left(\bigcup_{x\leqslant x_0;x\in D_X}\{x\}\right) \\ &= \sum_{x\leqslant x_0} P(X=x)=\sum_{x\leqslant x_0} P_X(x). \end{aligned}$$

Ejemplo

En el experimento del dado se tiene que:

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{si } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{en el resto de casos} \end{cases},$$

por lo tanto

$$F_X(x) = P(X \leqslant x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{6} & \text{si } 1 \leqslant x < 2 \\ \frac{2}{6} & \text{si } 2 \leqslant x < 3 \\ \frac{3}{6} & \text{si } 3 \leqslant x < 4 \\ \frac{4}{6} & \text{si } 4 \leqslant x < 5 \\ \frac{5}{6} & \text{si } 5 \leqslant x < 6 \\ 1 & \text{si } 6 \leqslant x \end{cases}$$

Ejemplo

Calculemos más detalladamente algún valor de F_X, por ejemplo:

$$\begin{aligned}
F_X(3.5) =& P(X \leqslant 3.5) = P(\{X = 1\} \cup \{X = 2\} \cup \{X = 3\}) \\
=& P(\{X = 1\}) + P(\{X = 2\}) + P(\{X = 3\}) \\
=& \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2},
\end{aligned}$$

o de otra forma

$$\begin{aligned}
F_X(3.5) &= \sum_{x \leqslant 3.5} P_X(x) = \sum_{x=1}^{3} P(X = x) \\
&= \sum_{x=1}^{3} \frac{1}{6} = 3 \cdot \frac{1}{6} = \frac{1}{2}.
\end{aligned}$$

Propiedades de la función de distribución

Sea X una variable con función de distribución F_X entonces:

a) $0 \leqslant F_X(x) \leqslant 1$ para todo x

b) Si $x < x'$ entonces

$$F_X(x) \leqslant F_X(x').$$

Es una función creciente, no necesariamente estrictamente creciente.

c) $\lim\limits_{x\to-\infty} F_X(x) = 0$ y $\lim\limits_{x\to+\infty} F_X(x) = 1$

d) Es continua por la derecha $\lim\limits_{x\to x_0^+} F_X(x) = F_X(x_0)$.

Momentos de variables aleatorias discretas

- Al igual que en estadística descriptiva utilizamos distintas medidas para resumir los valores centrales y para medir la dispersión de una muestra, podemos definir las correspondiente medidas para variables aleatorias.
- Estas medidas se les suele añadir el adjetivo poblacionales mientras que a las que provienen de la muestra se las adjetiva como muestrales.

Por ejemplo podemos buscar un valor que resuma toda la variable. Este valor es el que "esperamos" que se resuma la v.a. o esperamos que las realizaciones de la v.a. queden cerca de él. Veamos su definición formal.

Definición

El valor esperado o esperanza (*expected value* en inglés) $E(X)$ de una v.a. discreta X, se define como

$$E(X) = \sum_{x \in X(\Omega)} x P_X(x)$$

En ocasiones se le domina media (*mean* en inglés *mitjana* en catalán) poblacional o simpemente media y muy frecuentemente se la denota $\mu_X = E(X)$ o simplemente $\mu = E(X)$.

Iterpretación de la media aritmética

Supongamos que lanzamos un dado n veces y obtenemos unas frecuencias absolutas n_i para el resultado i con $i=1,\ldots,6$. Sea X la v.a. que nos representa el valor de una tirada del dado.
Calculemos la media aritmética (o media muestral) de los datos

$$\overline{x}=\frac{1\cdot n_1+2\cdot n_2+3\cdot n_3+4\cdot n_4+5\cdot n_5+6\cdot n_6}{n}=\sum_{x=1}^{6} x\frac{n_x}{n}.$$

Iterpretación de la media artimética

Si $n \to \infty$ se tiene que

$$\lim_{n\to\infty} \frac{n_x}{n} = P_X(x).$$

Por lo tanto

$$E(X) = \lim_{n\to\infty} \sum_{x=1}^{6} x \frac{n_x}{n}.$$

Entonces el valor esperado en una v.a. discreta puede entenderse como el valor promedio que tomaría una v.a. en un número grande de repeticiones.

Ejemplo

Sea X= número de erratas en una página de un texto con dominio $D_X = \{0, 1, 2\}$, y resulta que

$$P(X = 0) = 0.42, P(X = 1) = 0.4, P(X = 2) = 0.18.$$

entonces

$$E(X) = 0 \cdot 0.42 + 1 \cdot 0.4 + 2 \cdot 0.18 = 0.76.$$

Elegida una página del texto al azar esperamos encontrar 0.76 errores.

Ejemplo

Supongamos que en el ejemplo anterior el editor nos paga 2 euros por cada página que encontremos con 1 error y 3 euros por cada página con dos errores (y nada por las páginas correctas) ¿Cuánto esperamos cobrar si analizamos una página?

$$0 \cdot 0.42 + 2 \cdot 0.4 + 3 \cdot 0.18 = 1.34$$

Esperanzas de funciones de variables aleatorias discretas

Definición

Sea X una v.a. discreta con función de probabilidad P_X y de distribución F_X. Entonces el valor esperado de una función $g(x)$ es :

$$E(g(X)) = \sum_{x} g(x) P_X(x).$$

Propiedades de los valores esperados

a) $E(k) = k$. para cualquier constante k.

b) Si $a \leqslant X \leqslant b$ entonces $a \leqslant E(X) \leqslant b$.

c) Si X es una v.a. discreta que toma valores enteros no negativos entonces $E(X) = \sum_{x=0}^{+\infty} (1 - F_X(x))$.

Ejemplo

Supongamos que estamos sentados delante de nuestro ordenador con un amigo y le decimos que en dos minutos podemos programar una paleta para poner colores a unos gráficos.
Queremos la que paleta tenga dos botones con las opciones color rojo y color azul. Como hemos programado a gran velocidad resulta que el programa tiene un error; cada vez que se abre la paleta los colores se colocan al azar (con igual probabilidad) en cada botón, así que no sabemos en que color hemos de pinchar.
Además, como nos sobraron 15 segundos para hacer el programa y pensando en la comodidad del usuario, la paleta se cierra después de haber seleccionado un color y hay que volverla a abrir de nuevo.
La pregunta es ¿cuál es el valor esperado del número de veces que hemos pinchar el botón de color azul antes de obtener este color?

Ejemplo

Llamemos X al número de veces que pinchamos en el botón azul (y nos sale rojo) hasta obtener el primer azul. La variable X toma valores en los enteros no negativos. Su función de probabilidad queda determinada por

$$P_X(x) = P(X = x) = P(\overbrace{rojo, rojo, \ldots, rojo}^{x \text{veces}}, azul) = \left(\frac{1}{2}\right)^{x+1}$$

Propiedades de las series geométricas

Recordemos conceptos básicos de las series geométricas.

- Una progresión geométrica es una sucesión de la forma
$$r^0, r^1, \ldots, r^n, \ldots.$$
El valor r recibe el nombre de razón de la progresión geométrica.
- La serie geométrica es la suma de todos los valores de la progresión geométrica $\displaystyle\sum_{k=0}^{+\infty} r^k$.
- Las sumas parciales desde el término n_0 al n de una progresión geométrica son $\sum_{k=n_0}^{n} r^k = \frac{r^{n_0} - r^n r}{1-r}$.
- Si $|r| < 1$ la serie geométrica es convergente y $\displaystyle\sum_{k=0}^{+\infty} r^k = \frac{1}{1-r}$. En el caso en que se comience en n_0 se tiene que $\displaystyle\sum_{k=n_0}^{+\infty} r^k = \frac{r^{n_0}}{1-r}$.

Conceptos básicos de series geométricas

- Si $|r| < 1$ también son convergentes las derivadas, respecto de r, de la serie geométrica y convergen a la derivada correspondiente. Así tenemos que

$$\left(\sum_{k=0}^{+\infty} r^k\right)' = \sum_{k=1}^{+\infty} k r^{k-1} = \left(\frac{1}{1-r}\right)' = \frac{1}{(1-r)^2}.$$

$$\left(\sum_{k=0}^{+\infty} r^k\right)'' = \sum_{k=2}^{+\infty} k(k-1) r^{k-2} = \left(\frac{1}{1-r}\right)'' = \frac{2}{(1-r)^3}.$$

Ejemplo

Su esperanza es

$$E(X) = \sum_{x=0}^{+\infty} xP(X=x) = \sum_{x=0}^{+\infty} x\left(\frac{1}{2}\right)^{x+1}$$

$$= \left(\frac{1}{2}\right)^2 \sum_{x=1}^{+\infty} x\left(\frac{1}{2}\right)^{x-1} = \left(\frac{1}{2}\right)^2 \frac{1}{\left(1-\frac{1}{2}\right)^2} = 1.$$

Ahora calculemos su función de distribución

$$F_X(x) = P(X \leqslant x) = \sum_{k=0}^{x} P(X=k) = \sum_{k=0}^{x} \left(\frac{1}{2}\right)^{k+1}$$

$$= \frac{\frac{1}{2} - \frac{1}{2}^{x+1}\frac{1}{2}}{1-\frac{1}{2}} = 1 - \left(\frac{1}{2}\right)^{x+1}.$$

Ejemplo

Como la variable toma valores enteros positivos, podemos calcular su valor esperado de esta otra manera

$$E(X) = \sum_{x=0}^{+\infty} (1 - F_X(x)) = \sum_{x=0}^{+\infty} (\frac{1}{2})^{x+1} = \frac{1}{2} \frac{1}{1 - \frac{1}{2}} = 1.$$

Como ejercicio calculad el valor esperado de la variable

$$Y = \text{número de intentos para conseguir el color azul.}$$

Momentos de una variable aleatoria

Definición

- Llamaremos momento de orden r respecto al punto C a $E\left((X-C)^m\right)$.
- Cuando $C=0$ los momentos reciben el nombre de momentos respecto al origen.
- Cuando $C=E(X)$ reciben el nombre de momentos centrales o respecto de la media

Luego la esperanza es el momento de orden 1 respecto al origen. Estos momentos son la versión poblacional de los momentos que vimos en el capítulo de estadística descriptiva, recibiendo estos último el nombre de momentos muestrales.

Hemmos descrito el comportamiento aleatorio de una v.a. discreta mediante P_X y F_X.También tenemos un valor central; el valor esperado $E(X)$.Como medida básica nos queda definir una medida de lo lejos que están los datos del valor central $E(X)$ una de estas medidas es la varianza de X.

Definición

Sea X una v.a. Llamaremos varianza de X a

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

Por lo tanto la varianza es el momento central de orden 2. De forma frecuente se utiliza la notación

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

A la raíz cuadrada positiva de la varianza

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

se la denomina desviación típica o estándar de X.

En el caso discreto:

$$\sigma_X^2 = Var(X) = E((X - E(X))^2) = \sum_x (x - E(X))^2 P_X(x).$$

Propiedades

Sea X una v.a.

$$Var(X) = E(X^2) - (E(X))^2 = \sum_x x^2 P_X(X) - (E(X))^2$$

Demostración:

$$Var(X) = \sum_x (x - E(X))^2 P_X(x) = \sum_x (x^2 - 2xE(X) + (E(X)^2) P$$

$$\sum_x x^2 P_X(x) - E(X) \sum_x 2x P_X(x) + (E(X)^2) \sum_x P_X(x) =$$

$$E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2.$$

Ejemplo

Vamos a calcular en el ejemplo anterior la varianza del número de errores. Recordemos que:

$$P(X=0)=0.42,\quad P(X=1)=0.4,\quad P(X=2)=0.18$$

y $E(X)=0.76$
Entonces:

$$Var(X)=E(X^2)-(E(X))^2=E(X^2)-(0.76)^2$$

Ahora

$$E(X^2)=0^2(0.41)+1^2(0.4)+2^2(0.18)=0.4+0.72=1.12$$

y por lo tanto

$$Var(X)=E(X^2)-(0.76)^2=1.12-0.5776=0.542$$

y

$$\sqrt{Var(X)}=\sqrt{0.542}$$

En resumen $\sigma_X^2=0.542$ y $\sigma_X=\sqrt{0.542}$

Propiedades

- $Var(X) \geqslant 0$
- $Var(cte) = E(cte^2) - (E(cte))^2 = cte^2 - cte^2 = 0$
- *El mínimo de* $E((X - C)^2)$ *se alcanza cuando* $C = E(X)$ *y es* $Var(X)$. *Esta propiedad es una de las que hace útil a la varianza como medida de dispersión.*

Demostración:(ejercicio)

Un cambio lineal o transformación lineal de una v.a. X es otra v.a. $Y = a + bX$ donde $a, b \in \mathbb{R}$. Sea X una v.a. con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ y $a, b \in \mathbb{R}$. Entonces si $Y = a + bX$:

- $E(Y) = E(a + bX) = a + bE(X) = a + b\mu_X$.
- $Var(Y) = Var(a + bX) = b^2 Var(X) = b^2\sigma_X^2$
- $\sigma_Y = \sqrt{Var(Y)} = \sqrt{b^2 Var(X)} = |b|\sigma_X$

Demostración:

$$E(Y) = E(a+bX) = \sum_x (a+bX)P_X(x) = a\sum_x P_X(x)+b\sum_x xP$$

Las demostración de las demás propiedades queda como ejercicio.

Como ya hemos dicho las variables aleatorias continuas toman valores en intervalos. Lo más habitual es que estas variables tengan función de distribución continua y derivable (salvo a los más en una cantidad finita o numerable de puntos). En lo que sigue supondremos que la función de distribución de variables aleatorias continuas cumplen estas propiedades.

Notemos que si X es una v.a. con función de distribución continua se tiene que $P(X=x_0)=F_X(x_0)-F(x_0^-)=0$. Por lo que no tiene sentido definir ''*función de probabilidad*". Más en general tendremos que $P(X<x_0)=P(X\leqslant x_0)$. Por otra parte podemos utilizar una regla parecida del cociente entre casos favorables y casos posibles de Laplace pero en este caso el conteo se hace por la "medida" de los casos posibles partida por la "medida" de los casos favorables. Veamos un ejemplo de v.a. continua, que ampliaremos en el tema siguiente, en el que se utilizan todos estos conceptos.

Ejemplo: Distribución uniforme en el intervalo unidad.

Supongamos que lanzamos un dardo a una diana de radio 1, de forma que sea "*equiprobable*" cualquier distancia al centro[1]. Consideremos la v.a. continua X =distancia al centro.

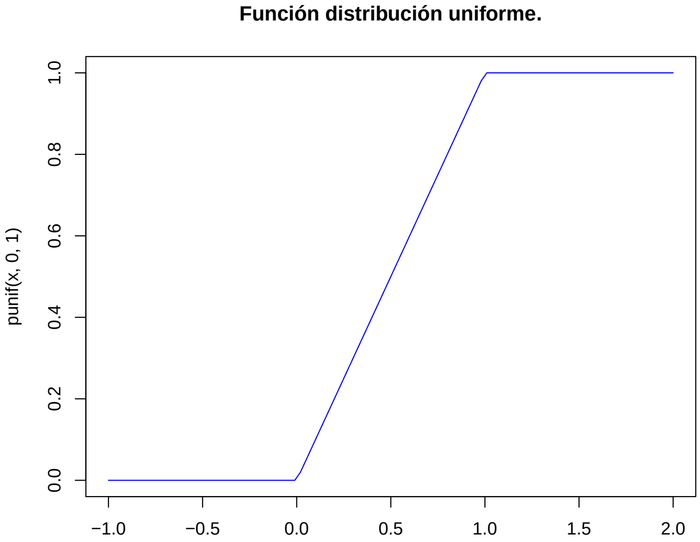
$$F_X(x) = \begin{cases} 0 & \text{si } x \leqslant 0 \\ x & \text{si } 0 < x < 1 \\ 1 & \text{si } x \geqslant 1 \end{cases}$$

[1]!Cuidado! esto no es equivalente a que cualquier punto de la diana sea "*equiprobable*".

Ya que

- C.F. ''*longitud favorable*'' $x - 0$
- C.P. ''*longitud posible*'' $1 - 0$
- Luego $P(X \leqslant x) = \frac{x-0}{1-0} = x$

```r
curve(punif(x,0,1),xlim=c(-1,2),col="blue",
      main="Función distribución uniforme.")
```



En las variables continuas los sucesos del tipo $\{X \leqslant x\}$ y $\{X < x\}$ tendrán la misma probabilidad, y otros tipos de sucesos similares también, algunas de estas propiedades se explicitan en la siguiente proposición.

Propiedades

Dada una v.a. continua X se tiene que:

a) $P(X \leqslant b) = P(X < b)$

b) $P(X < b) = P(X < a) + P(a < X < b)$

c) $P(a < X < b) = P(X < b) - P(X < a)$

Demostración:

b) $\{X < a\} \cap \{a < X < b\} = \emptyset$
$\{X < a\} \cup \{a < X < b\} = \{X < a\}$ entonces

$$P(X \leqslant b) = P(\{X < a\} \cup \{a < X < b\}) = P(X < a) + P(a$$

a) $P(X < b) = P(X < b) + P(X = b) = P(X < b)$

c) Ídem que b) aplicando a).

Las propiedades anteriores y combinaciones de ellas se pueden escribir utilizando la función de distribución de X:

Propiedades

Dada una variable aleatoria continua se tiene que:

a) $F_X(b) = F_X(a) + P(a < X < b)$

b) $P(a < X < b) = F_X(b) - F_X(a)$

c) $P(a \leqslant X \leqslant b) = F_X(b) - F_X(a)$

Ejemplo

Demostración: ejercicio.
En los dardos:

$$P(0.25 < X < 0.3) = F_X(0.3) - F_X(0.25) =$$

$$= 0.3 - 0.25 = 0.05$$

Una función $f:\mathbb{R}\to\mathbb{R}$ es una función de densidad sobre $\mathbb{R}$ si cumple que

a) $f_X(x)\geqslant 0$ para todo $x\in\mathbb{R}$.
b) f es continua salvo a lo más en una cantidad finita de puntos sobre cada intervalo acotado de $\mathbb{R}$.
c) $\displaystyle\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

Definición

Sea X una v.a. con función de distribución F_X. Sea $f:\mathbb{R}\to\mathbb{R}$ una función de densidad tal que

$$F_X(x)=\int_{-\infty}^{x} f_X(t)dt. \text{ para todo } x\in\mathbb{R}.$$

Entonces X es una variable aleatoria continua y f_X es la densidad de la v.a. X.
El conjunto $D_X=\{x\in\mathbb{R}|f_x(x)>0\}$ recibe el nombre de soporte o dominio de la variable aleatoria continua y se interpreta su conjunto de resultados posibles.

En nuestra diana la función f es una densidad

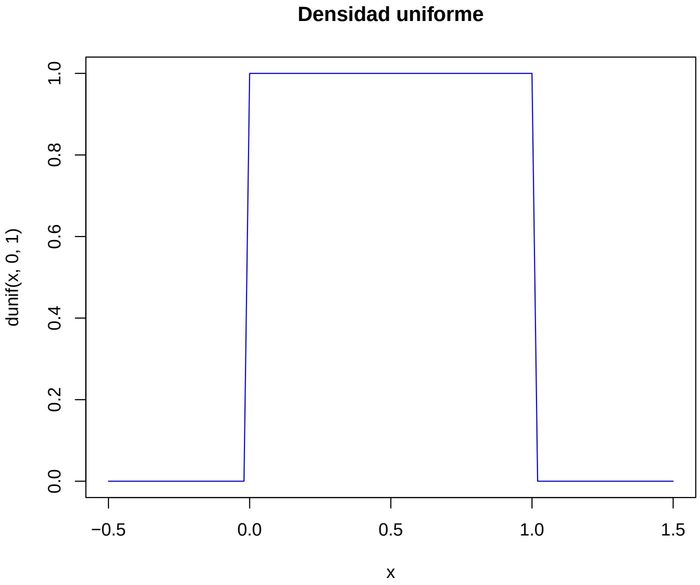
$$f_X(x) = \begin{cases} 0 & \text{si } x \leqslant 0 \\ 1 & \text{si } 0 < x < 1 \\ 0 & \text{si } 1 \leqslant x \end{cases}$$

que es la densidad de X, en efecto:

- Si $x \leqslant 0$ entonces $\int_{-\infty}^{x} f_X(t)dt = 0$.
- Si $0 \leqslant x \leqslant 1$ entonces $\int_{-\infty}^{x} f_X(t)dt = \int_0^x 1dt = x$.
- Si $x \geqslant 1$ entonces $\int_{-\infty}^{x} f_X(t)dt = \int_0^1 1dt = 1$.

uego $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ para todo $x \in \mathbb{R}$.

```
curve(dunif(x,0,1),xlim=c(-0.5,1.5),col="blue",main
```



La función de densidad nos permite calcular diversas probabilidades.

Propiedades

Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces

1. $P(a < X < b) = P(a < X \leqslant b) = P(a \leqslant X < b) = P(a \leqslant X \leqslant b) = \int_a^b f_X(x)dx.$
2. *Si A es un recinto adecuado*[a] *de* $\mathbb{R}$ *entonces* $P(X \in A) = \int_A f(x)dx = \int_{A\cap D_X} f(x)dx.$

[a] Un boreliano

La siguiente proposición fija algunas propiedades de la función de densidad para v.a. continuas y nos da un método de cálculo; la densidad es la derivada de la función de distribución.

Propiedades

Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces:

a) *F_X es continua.*

b) *Si f_x es continua en un punto x, F_X es derivable en ese punto y $F'_X(x) = f_X(x)$.*

c) *$P(X = x) = 0$ para todo $x \in \mathbb{R}$.*

Ejemplo

Sea $X =$ tiempo de ejecución de un proceso. Se supone que X sigue una distribución uniforme en dos unidades de tiempo, si tarda más el proceso se cancela. Entonces

$$F_X(x) = P(X \leqslant x) = \frac{CF}{CP} = \frac{x}{2}$$

Luego su función de distribución es:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leqslant 0 \\ \frac{x}{2} & \text{si } 0 < x < 2 \\ 1 & \text{si } 2 \leqslant x \end{cases}$$

mientras que su función de densidad es:

$$f_X(x) = F'_X(x) = \begin{cases} 0 & \text{si } x \leqslant 0 \\ \frac{1}{2} & \text{si } 0 < x \leqslant 2 \\ 0 & \text{si } 2 \leqslant x \end{cases}$$

Efectivamente

- $f_X(x) \geqslant 0$, y tiene un conjunto finito de discontinuidades.
- $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$. para todo $x \in \mathbb{R}$ (ejercicio, resolverlo gráficamente.)
- $\int\limits_{-\infty}^{+\infty} f_X(x)dx = \int\limits_{0}^{2} \frac{1}{2}dx = \frac{x}{2} \left.\right|_0^2 == \frac{2}{2} - \frac{0}{2} = 1.$

Ejercicio: Calcular la probabilidad de que uno de nuestros procesos tarde más de una unidad de tiempo en ser procesado. Calcular también la probabilidad de que dure entre 0.5 y 1.5 unidades de tiempo.

Los mismos comentarios y definiciones que se dieron en la sección correspondiente del tema de estadística descriptiva son aplicables aquí. Así que sólo daremos las definiciones, la forma de cálculo y algunos ejemplos.

Sea X una v.a. continua con función de densidad $f_X(x)$ entonces:

- su esperanza es : $E(X) = \int\limits_{-\infty}^{+\infty} x f_X(x) dx.$
- Si $g(x)$ es una función de la variable X entonces

$$E(g(X)) = \int\limits_{-\infty}^{+\infty} g(x) f_X(x) dx.$$

- $Var(X) = \sigma_X^2 = E((X-\mu_X)^2) = \int_{-\infty}^{+\infty} (x-\mu_X)^2 f_X(x) dx.$
- A $\sigma_X = +\sqrt{\sigma_X^2}$ se le denomina desviación típica de X.

Propiedades

- $\sigma_X^2 \geqslant 0$
- $Var(cte) = E(cte^2) - (E(cte))^2 = cte^2 - cte^2 = 0$
- $Var(x) = E(X^2) - \mu_X^2 = \int\limits_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu_X^2.$
- El mínimo de $E((X - C)^2)$ se alcanza cuando $C = E(X)$ y es $Var(X)$.

Ejemplos Calcular μ_X y σ_X^2 en el dardo.
Resultado $\mu_X = \frac{1}{2}$, $E(X^2) = \frac{1}{3}$, $Var(X) = \frac{1}{12}$.

Sea X una v.a. continua con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ sea $Y = a + bX$, donde $a, b \in \mathbb{R}$, es una nueva v.a. continua obtenida mediante una transformación lineal de X. Se verifican las mismas propiedades que en el caso discreto:

- $E(Y) = E(a + bX) = a + bE(X)$
- $Var(Y) = Var(a + bX) = b^2 Var(X)$
- $\sigma_Y = |b|\sigma_X$
- $Z = \frac{X-\mu_X}{\sigma_X}$ es una transformación lineal de X de forma que

$$E(Z) = 0 \text{ y } Var(Z) = 1$$

Ejemplo En una empresa de venta de vinos por internet, sea $X =$ número de litros de vino del país vendidos en un año. Supongamos que sabemos que $E(X) = 10000$ y que $Var(X) = 100$ Supongamos que los gastos fijos de distribución son 50000 y el beneficio por litro es de 10 pts por botella. Definimos $T = 10X - 50000$ que será el beneficio después de gastos entonces:

$$E(T) = 10E(X) - 50000 = 50000$$

y

$$Var(T) = 10^2 VAR(X) = 10000$$

Muchas variables aleatorias son funciones de otras v.a. En lo que sigue resumiremos diversas técnicas para dada una v.a. X y una transformación $Y = h(X)$ encontrar F_Y a partir de F_X.

Propiedades

Sea X una v.a. discreta con $X(\Omega)=\{x_1,x_2,\ldots,x_n,..\}$ y sea $h:\mathbb{R}\to\mathbb{R}$ una aplicación. Entonces $Y=h(X)$ es también una v.a. discreta. Además si P_X y F_X son las funciones de probabilidad y de distribución de X entonces

a) $$P_Y(y)=\sum_{x_i|h(x_i)=y} P_X(x_i).$$

b) $$F_Y(y)=\sum_{x_i|h(x_i)\leqslant y} P_X(x_i).$$

Desafortunadamente este caso no es tan sencillo como el anterior, pues la transformación de una v.a. continua puede ser continua, discreta, mixta ...

Propiedades

Sea X una v.a. continua cuya función de densidad es f_X. Sea $h:\mathbb{R}\to\mathbb{R}$ una aplicación estrictamente monótona y derivable, tal que $h'(x)\neq 0$ para todo $x\in\mathbb{R}$. Sea $Y=h(X)$ la transformación de X por h. Entonces Y es una v.a. continua con función de densidad

$$f_Y(y)=\left.\frac{f_X(x)}{|h'(x)|}\right|_{x=h^{-1}(y)}$$

Propiedades

Sea X una v.a. continua cuya función de densidad es f_X. Si $h:\mathbb{R}\to\mathbb{R}$ es una aplicación, no necesariamente monótona, pero sí derivable con derivada no nula, y si la ecuación $h(x)=y$ tiene un número finito de soluciones $x_1,x_2,..,x_n$ entonces:

$$f_Y(y)=\sum_{k=1}^{n}\left.\frac{f_X(x)}{|h'(x)|}\right|_{x=x_k}$$

Cuando no podamos aplicar las propiedades anteriores intentaremos calcular primero la función de distribución de la transformación y luego su densidad.
Notemos que en general si $Y = g(X)$ es una v.a. transformación de la v.a. X entonces

$$F_Y(y) = P(Y \leqslant y) = P(g(X) \leqslant y)$$

Por ejemplo si g es estrictamente creciente y cont.

$$F_Y(y) = P(g(X) \leqslant y) = P(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y))$$

y si g es estrictamente decreciente y cont.

$$F_Y(y) = P(g(X) \leqslant y) = P(X \geqslant g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

Veremos en esta sección distintas desigualdades que acotan determinadas probabilidades de una variable aleatoria. Estas desigualdades sirven en algunos casos para acotar probabilidades de determinados sucesos, también son interesantes desde el punto de vista teórico y por ejemplo para justificar que la varianza es una mediada de la dispersión de los datos

Propiedades

Sea X una v.a. positiva con $E(X)$ finita. Entonces $P(X \geqslant a) \leqslant \frac{E(X)}{a}$ para todo $a > 0$.

Demostración:

Si X es continua y sólo toma valores positivos

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{0}^{+\infty} x f_X(x) dx =$$

$$\int_{0}^{a} x f_X(x) dx + \int_{a}^{+\infty} x f_X(x) dx \geqslant \int_{a}^{+\infty} x f_X(x) dx \geqslant$$

$$a \int_{a}^{+\infty} f_X(x) dx = aP(X \geqslant a)$$

de donde se sigue que $P(X \geqslant a) \leqslant \frac{E(X)}{a}$

Corollary

Sea X una v.a. con $E(X)$ finita entonces $P(|X| \geqslant a) \leqslant \frac{E(|X|)}{a}$. Para todo $a > 0$

Propiedades

Sea X una v.a.con $E(X)=\mu$ y $Var(X)=\sigma^2$ entonces

$$P(|X-\mu|\geqslant a)\leqslant \frac{\sigma^2}{a^2}$$

para todo $a>0$

Demostración:
Apliquemos la consecuencia de la desigualdad de Markov a la v.a. no negativa $Y^2 = (X-\mu)^2$ entones
$P(Y^2 \geqslant a^2) \leqslant \frac{E(Y^2)}{a^2} = \frac{E((X-\mu)^2)}{a^2} = \frac{Var(X)}{a^2} = \frac{\sigma^2}{a^2}$

Por otra parte

$$P(Y^2 \geqslant a^2) = P(|Y| \geqslant a) = P(|X - \mu| \geqslant a)$$

hecho que, junto con la desigualdad anterior, demuestra el resultado.

Observación: Supongamos que X es una v.a. con $Var(X) = 0$ entonces aplicando la desigualdad anterior $P(|X - E(X)| \geqslant a) = 0$ para todo $a > 0$ lo que implica que $P(X = E(X)) = 1$ luego la probabilidad de que X sea constantemente $E(X)$ es 1. Lo que nos confirma la utilidad de la varianza es una medida de la dispersión de los datos.

Ejemplo

Se sabe que el tiempo de respuesta medio y la desviación típica de un sistema multiusuario son 15 y 3 u.t. respectivamente. Entonces:

$P(|X-15|\geqslant 5)\leqslant \frac{9}{25}=0.36.$

Si substituimos a por $a\sigma$ en la desigualdad de Chebyshef.

Entonces nos queda:

$P(|X-\mu|\geqslant a\sigma)\leqslant \frac{\sigma^2}{(a\sigma)^2}=\frac{1}{a^2}$

Que es otra manera de expresar la desigualdad de Chebyshef.

La desigualdad de Chebyshef también se puede escribir de al menos dos maneras más:

$P(\mu-a\leqslant X\leqslant \mu+a)\geqslant 1-\frac{\sigma^2}{a^2}$

$P(\mu-a\cdot\sigma\leqslant X\leqslant \mu+a\cdot\sigma)$

Tomando la segunda expresión que hemos visto para la desigualdad de Chebyshef para distintos valores de $a > 0$ tenemos la siguiente tabla.

a	$P(\|X - E(X)\| \geqslant a\sigma)$
1	$\leqslant 1$
2	$\leqslant 0.25$
3	$\leqslant 0.111$
4	$\leqslant 0.0025$

Lo que se interpreta, por ejemplo para $a = 2$, como que dada una v.a. X con cualquier distribución que tenga $E(X)$ y $Var(X)$ finitos la probabilidad de que un valor se aleje de la media μ más de 2 desviaciones típicas es menor o igual que 0.25. Es decir sólo el 25 % de los valores estarán alejados de la media más de 2σ !sea cual sea la distribución de la v.a.!