# Estadística

Universidad Europea de Madrid

Laureate International Universities

# Variables aleatorias

Supongamos que realizamos el experimento: tirar dos veces un dado.

Hasta ahora, hemos tratado sucesos, por ejemplo:

*A2 = "la suma de dos tiradas de un dado* es 2".

Podemos querer ahora, por ejemplo, analizar todos los sucesos asociados a los distintos valores que da la suma de las dos tiradas. Por ejemplo, A3= *"la suma de dos tiradas de un dado* es 3", A4= *"la suma de dos tiradas de un dado* es 4", etc.

# Variables aleatorias

Podríamos para ello definir una función:

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

que asocie a cada suceso elemental (en este caso los valores obtenidos en cada tirada) la suma de los valores obtenidos.

X = {suma de las dos tiradas}

Tendríamos entonces que X toma el valor 2 ( X = 2) sólo para el suceso (1,1), es decir X(1,1) = 2.

X = 4 para los sucesos elementales (1,3), (2,2), (3,1), es decir X(1,3) = 4, X(2,2) = 4, X(3,1) = 4

# Variables aleatorias

Observar que, por ejemplo:

$\{(a,b) \,/\, X(a,b) = 4\} = A4 = \{(1,3), (2,2), (3,1)\}$

$\{(a,b) \,/\, X(a,b) = 2\} = A2 = \{(1,1)\}$

$\{(a,b) \,/\, X(a,b) = 7\} = A7 = \{(1,6), (2,5), (3,4),(4,3), (5,2), (6,1)\}$

# Variables aleatorias

**Definición**

Una **variable aleatoria** es una función que asocia un número real a cada elemento del espacio muestral, de forma que la probabilidad de que la variable tome un valor sea igual a la probabilidad de que ocurra cierto suceso.

# Variables aleatorias

Ejemplo:

Consideramos el experimento de lanzar dos veces un dado y la variable aleatoria $X : \Omega \longrightarrow \mathbb{R}$, dada por

$$X = \text{suma de las dos tiradas}$$

| $x$ | *Sucesos elementales* | | | | | |
|---|---|---|---|---|---|---|
| 2 | $(1,1)$ | | | | | |
| 3 | $(1,2)$ | $(2,1)$ | | | | |
| 4 | $(1,3)$ | $(2,2)$ | $(3,1)$ | | | |
| 5 | $(1,4)$ | $(2,3)$ | $(3,2)$ | $(4,1)$ | | |
| 6 | $(1,5)$ | $(2,4)$ | $(3,3)$ | $(4,2)$ | $(5,1)$ | |
| 7 | $(1,6)$ | $(2,5)$ | $(3,4)$ | $(4,3)$ | $(5,2)$ | $(6,1)$ |
| 8 | $(2,6)$ | $(3,5)$ | $(4,4)$ | $(5,3)$ | $(6,2)$ | |
| 9 | $(3,6)$ | $(4,5)$ | $(5,4)$ | $(6,3)$ | | |
| 10 | $(4,6)$ | $(5,5)$ | $(6,4)$ | | | |
| 11 | $(5,6)$ | $(6,5)$ | | | | |
| 12 | $(6,6)$ | | | | | |

# Variables aleatorias

Observaciones:

- La variable *X* no recibe el calificativo de *aleatoria* por el hecho de que atribuya de modo imprevisible un valor cualquiera a un elemento *e de* Ω, ya que este valor está definido de forma precisa (determinística). Lo que es aleatorio en realidad, es que al hacer el experimento, no sabemos qué elemento de Ω puede ocurrir.

- La composición de una función real con una variable es también variable aleatoria, pues está definida sobre Ω y a cada elemento suyo le asocia un valor real.

Universidad Europea de Madrid

# Variables aleatorias

Si sobre los elementos de $\Omega$ existe una *distribución de probabilidad*, esta se transmite a los valores que toma la variable X. Es decir, toda v.a. conserva la estructura probabilística del experimento aleatorio que describe:

$$P(X = x) = P(\{s \in \Omega : X(s) = x\})$$

Los sucesos que nos van a interesar ahora son del tipo $X \in A$, donde A es un subconjunto de $\mathbb{R}$

$$P(X \in A) = P(\{s \in \Omega : X(s) \in A\})$$

# Ejemplo

- $P(X=2)=P((1,1))=1/36$
- $P(X=3)=P((1,2);(2,1))=2/36$
- $P(X=4)=3/36$
- $P(X=5)=4/36$
- $P(X=6)=5/36$
- $P(X=7)=6/36$
- $P(X=8)=5/36$
- $P(X=9)=4/36$
- …….

# Variables aleatorias

## Observaciones

- Los valores de una variable aleatoria cambian de un experimento a otro al cambiar los resultados del experimento.

- Una variable aleatoria está definida por:
    - Los valores que toma.
    - La probabilidad de tomar cada uno de esos valores

# Variables aleatorias

El **rango** de una variable aleatoria es el conjunto de valores que puede tomar la variable.

Atendiendo al rango las variables se pueden clasificar como:

Variables aleatorias **discretas:** aquellas en las que el rango es finito o infinito numerable

Variables aleatorias **continuas:** aquellas en las que el rango es un intervalo de números reales

# Variables aleatorias

Variables aleatorias **discretas:** frecuentemente cuentan el número de veces que ocurre algo.

Ejemplos:

• Número de caras en tres tiradas de una moneda.

• Número de llamadas que recibe un teléfono en una hora

Variables aleatorias **continuas:** frecuentemente miden una magnitud

Ejemplos:

• Tiempo que esperan los clientes para pagar en un supermercado

• Cantidad de agua consumida en un mes

# Variables aleatorias discretas

## Función de probabilidad

Sea X una variable aleatoria discreta, con posibles valores $x_1, x_2, x_3, \dots$

Se define su **función de probabilidad o de masa** como

$$p_i = P(X = x_i)$$

# Variables aleatorias discretas

Las propiedades de la función de probabilidad se deducen de forma inmediata de los axiomas de la probabilidad:

1. $0 \leq p(x_i) \leq 1$
2. $\sum_{i=1}^{n} p(x_i) = 1$

(si la variable toma infinitos valores, se tiene $\sum_{i=1}^{\infty} p(x_i) = 1$)

3. $P(a \leq X \leq b) = P(X = a) + P(X = a + 1) + \cdots + P(X = b) = \sum_{x_i=a}^{b} p(x_i)$

# Variables aleatorias discretas

## Ejemplo

Lanzamos 3 veces moneda. Espacio muestral:

$$\Omega = \{CCC,CCX,CXC,XCC,XXC,XCX,CXX,XXX\}$$

Consideramos la variable aleatoria

X = {nº de caras al lanzar tres veces una moneda}

Posibles valores de X : 0, 1, 2 y 3

La <u>función de masa</u> es:

$$p(0) = P(X = 0) = \frac{1}{8} = 0{,}125 \qquad p(1) = P(X = 1) = \frac{3}{8} = 0{,}375$$

$$p(2) = P(X = 2) = \frac{3}{8} = 0{,}375 \qquad p(3) = P(X = 3) = \frac{1}{8} = 0{,}125$$

Universidad Europea de Madrid

# Variables aleatorias discretas

## Ejemplo

¿Cuál será la probabilidad de que salgan como mucho dos caras?

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}125 + 0{,}375 + 0{,}375$$
$$= 0{,}875$$

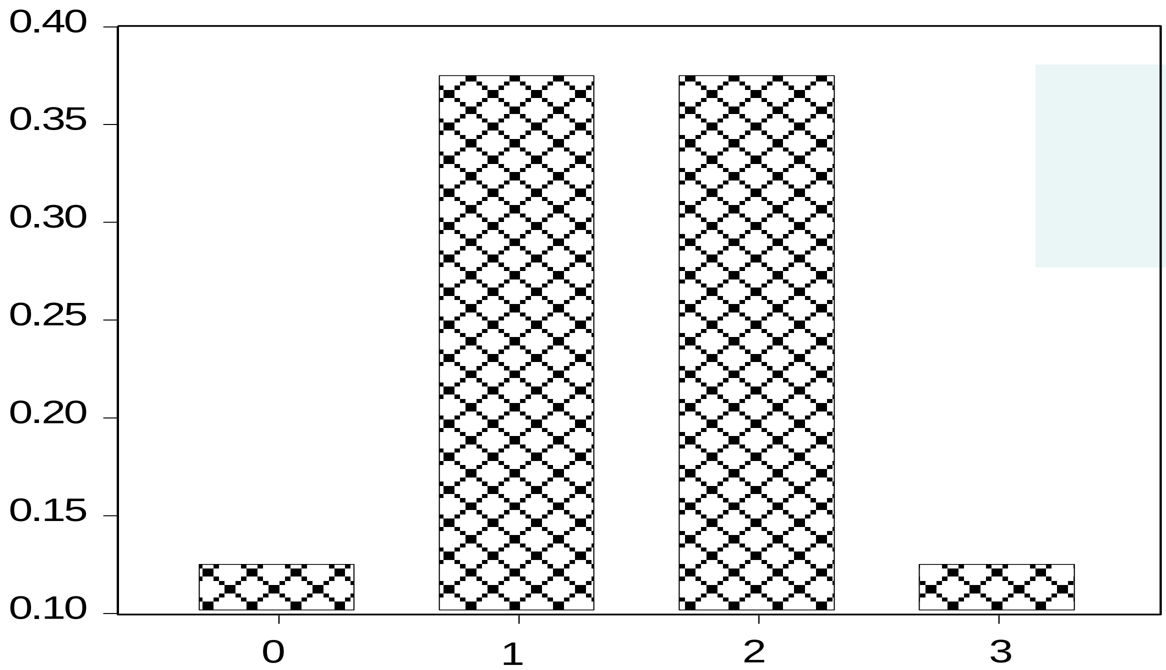
¿y la probabilidad de que el número de caras esté entre 1 y 2?

$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = 0{,}375 + 0{,}375 = 0{,}75$$

# Variables aleatorias discretas

## Ejemplo

Gráfica de la función de probabilidad de X:



# Variables aleatorias discretas

## Ejercicio 1:

Una vendedora ha concertado dos citas para vender enciclopedias. Cree que en la primera cita puede realizar una venta con una probabilidad 0,3; que en la segunda lo puede hacer con una probabilidad de 0,6, y que los resultados de las dos citas son independientes. ¿Cuál es la función de probabilidad de X, el número de ventas realizadas?

# Variables aleatorias discretas

## Solución

La variable X puede tomar cualquiera de los valores 0, 1 ó 2. Será igual a cero si no vende en ninguna de las citas; por tanto:

$$
\begin{aligned}
P(X = 0) &= P(\{\text{no vende en la primera, no vende en la segunda}\}) \\
&= P(\{\text{no vende en la primera}\})P(\{\text{no vende en la segunda}\}) \\
&= (1-0{,}3)(1-0{,}6) = 0{,}28
\end{aligned}
$$

# Variables aleatorias discretas

## Solución

La variable X será igual a uno si consigue vender en la primera pero no en la segunda, o si no vende en la primera pero si en la segunda. Se tiene entonces:

P(X = 1) = P({vende en la primera, no vende en la segunda}) +
P({no vende en la primera, vende en la segunda}) +
= P({vende en la primera})P({no vende en la segunda}) +
P({no vende en la primera})P({vende en la segunda})
= 0,3 · (1-0,6) + (1-0,3) · 0,6 = 0,54

Universidad Europea de Madrid

# Variables aleatorias discretas

Solución

La variable X será igual a 2 si consigue vender en la primera y en la segunda. Se tiene entonces:

$$P(X = 2) = P(\{\text{vende en la primera, vende en la segunda}\}) +$$
$$= P(\{\text{vende en la primera}\})P(\{\text{vende en la segunda}\})$$
$$= 0{,}3 \cdot 0{,}6 = 0{,}18$$

# Variables aleatorias discretas

## Función de distribución

En ocasiones nos puede interesar la probabilidad de que una variable tome un valor menor o igual que una cantidad. Definimos entonces la **función de distribución**

$$F(x_o) = P(X \leq x_o)$$

# Variables aleatorias discretas

## Función de distribución

La función de distribución, definida sobre todo $\mathbb{R}$, cumple:

$$0 \leq F(x) \leq 1$$

Es no decreciente.

$$P(X > x) = 1 - F(x)$$

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$$

# Variables aleatorias discretas

## Función de distribución

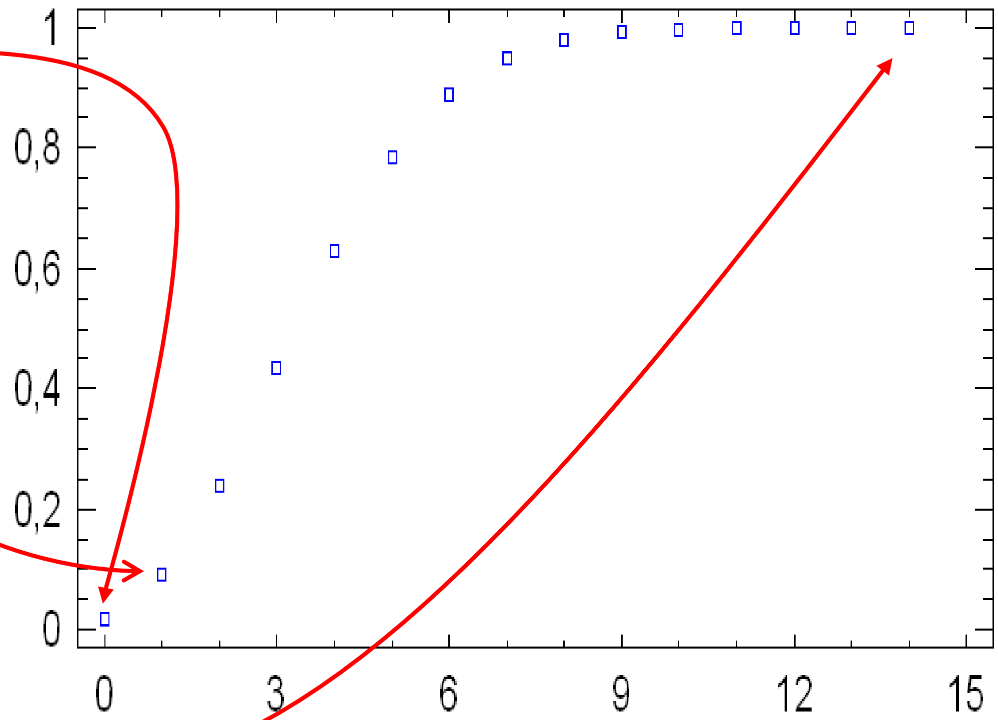
Si X toma valores $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$

$$F(x_1) = P(X \leq x_1) = p(x_1)$$

$$F(x_2) = P(X \leq x_2) = p(x_1) + p(x_2)$$

$$\vdots$$

$$F(x_n) = P(X \leq x_n) = \sum_{i=1}^{n} p(x_i) = 1$$



# Variables aleatorias discretas

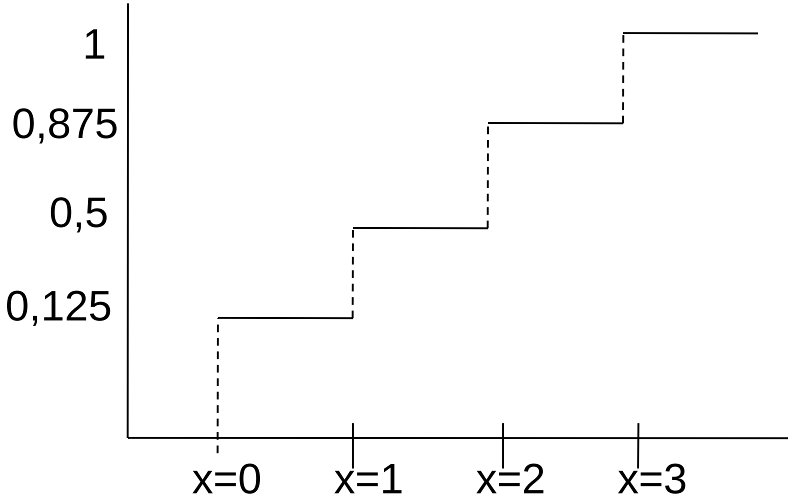
Ejemplo 1

Continuando con el ejemplo anterior: nº caras al lanzar tres veces una moneda. Calculemos la función de distribución

$P(X \leq 0) = P(X = 0) = 0{,}125$ $\qquad$ $P(X \leq 1) = P(X = 0) + P(X = 1) = 0{,}125 + 0{,}375 = 0{,}5$

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}5 + 0{,}375$

$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0{,}875 + 0{,}125$



# Variables aleatorias discretas

Ejemplo 2:

En ocasiones, algunas líneas aéreas venden más pasajes que los disponibles en un vuelo. Una compañía ha vendido 205 billetes que corresponden a un avión con 200 plazas.

Sea X la variable aleatoria que expresa el número de viajeros que se presentan en el aeropuerto para viajar en el avión. La distribución de X es

| $x$ | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 |
|---|---|---|---|---|---|---|---|---|
| P(X = x) | 0,05 | 0,09 | 0,15 | 0,20 | 0,23 | 0,17 | 0,09 | 0,02 |

Universidad Europea de Madrid

# Variables aleatorias discretas

1. Hallar la probabilidad de que todos los pasajeros que llegan a tomar el vuelo tengan plaza.

Queremos calcular $P(X \leq 200)$.

$$P(X \leq 200) = P(X = 198) + P(X = 199) + P(X = 200)$$
$$= 0{,}05 + 0{,}09 + 0{,}15$$
$$= 0{,}29$$

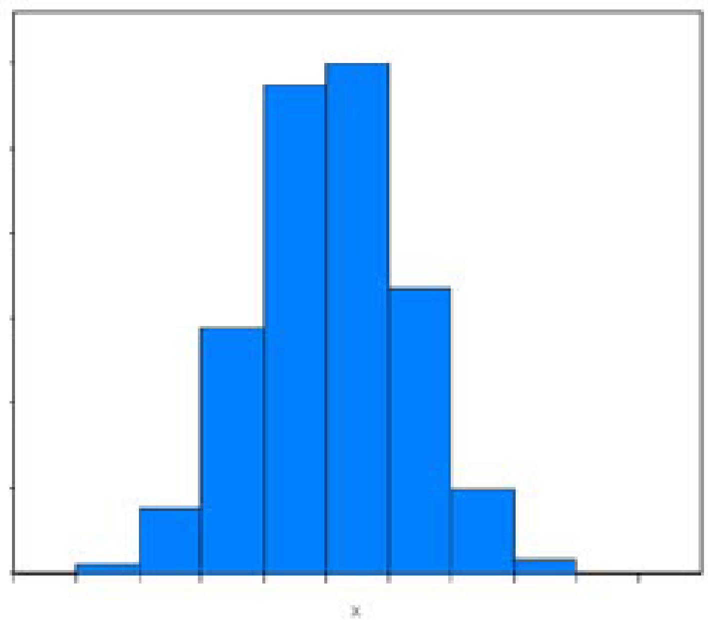
2. ¿Cuál es la probabilidad de que se quede sin plaza alguno de los pasajeros que se presentan en el aeropuerto?

$$P(X > 200) = 1 - P(X \leq 200) = 0{,}71.$$

Universidad Europea de Madrid

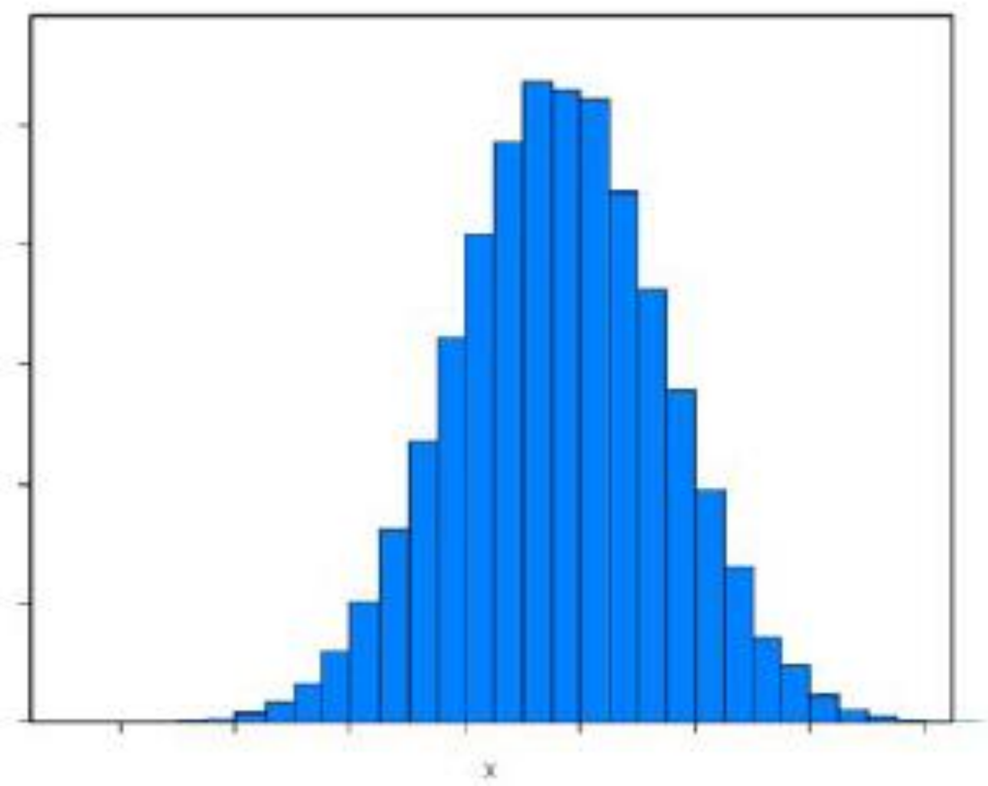
# Variables aleatorias continuas

Si medimos una variable continua y la representamos en un histograma:



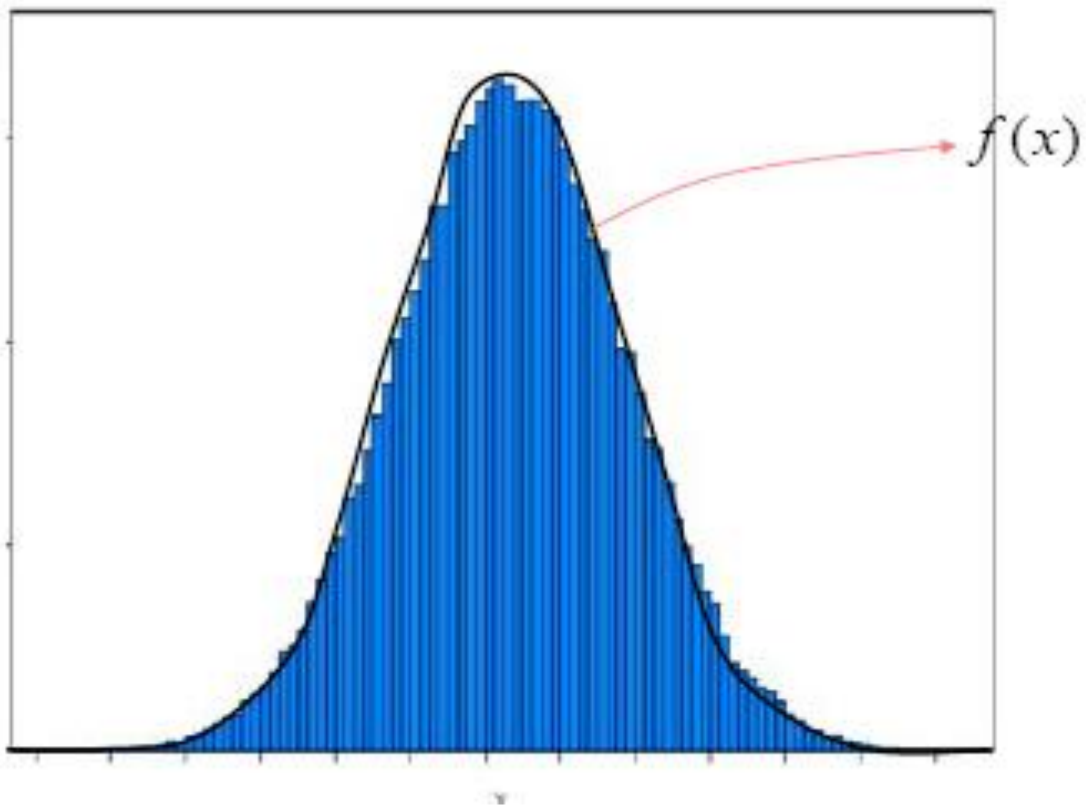
# Variables aleatorias continuas

Y hacemos las clases cada vez más pequeñas:



# Variables aleatorias continuas

El polígono de frecuencias tenderá a una curva



# Variables aleatorias continuas

Diferencias:

• Histograma o polígono:

Describe sólo a los n datos observados (MUESTRA)

La altura suele ser la frecuencia (absoluta o relativa) de cada intervalo

• Función de densidad:

Describe a TODA LA POBLACIÓN

Es la densidad de probabilidad o probabilidad por unidad de medida en cada punto (unidades distintas que el histograma. Para que representaran las mismas unidades en este último deberíamos elegir la altura de tal forma que el área represente a la frecuencia)

# Variables aleatorias continuas

Cuando una variable es continua, no tiene sentido hacer la suma:

$$\sum_{x_i \in V} p(x_i)$$

ya que el conjunto de valores V que toma la variable es no numerable.

Lo natural es generalizar $\sum \rightarrow \int$

Introducimos un nuevo concepto que sustituye en variables continuas al de función de probabilidad en variables discretas

# Variables aleatorias continuas

## Función de densidad

La **función de densidad** describe la distribución de probabilidad de una variable continua. Es una función que verifica:

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

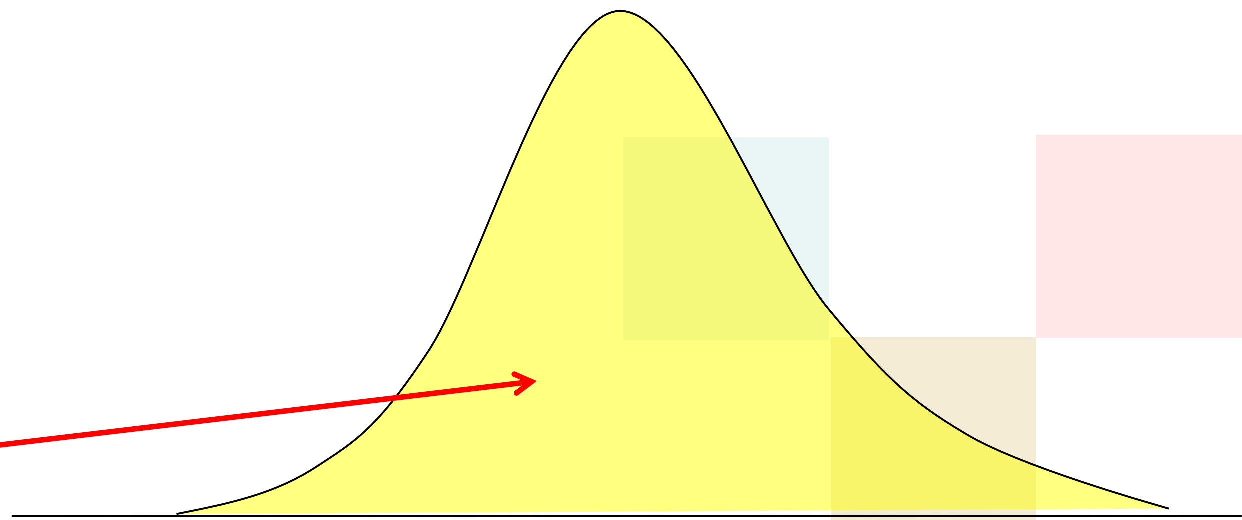
# Variables aleatorias continuas

## Función de densidad

Gráficamente

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$



$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

# Variables aleatorias continuas

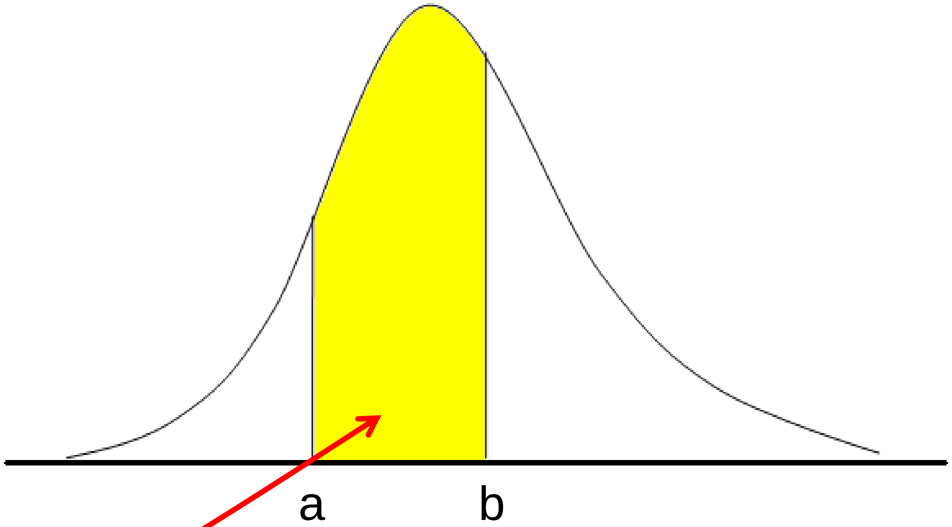
## Función de densidad

Gráficamente

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

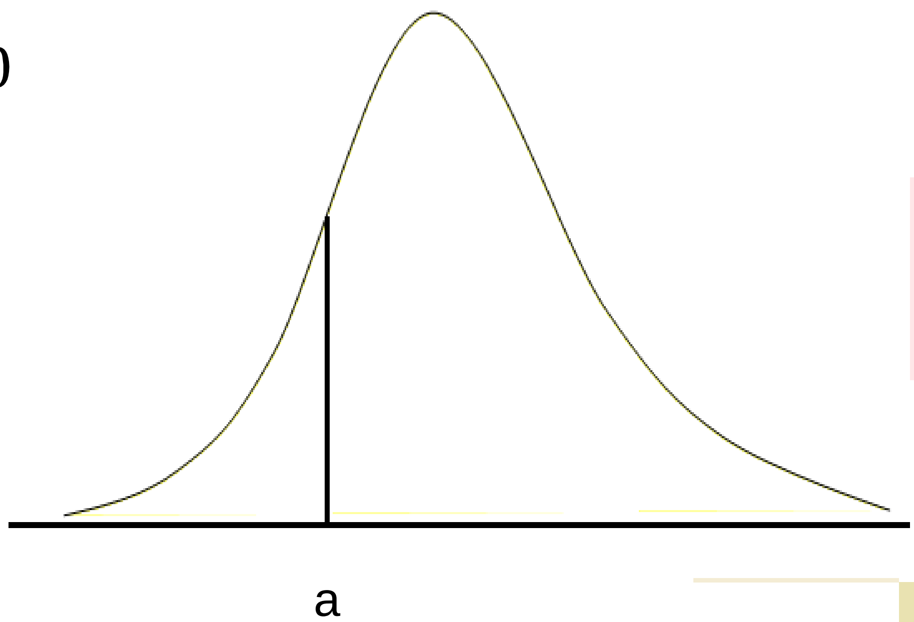



Universidad Europea de Madrid

# Variables aleatorias continuas

**Observación**

$$P(X = a) = \int_a^a f(x)dx = 0$$




Y por lo tanto se verifica

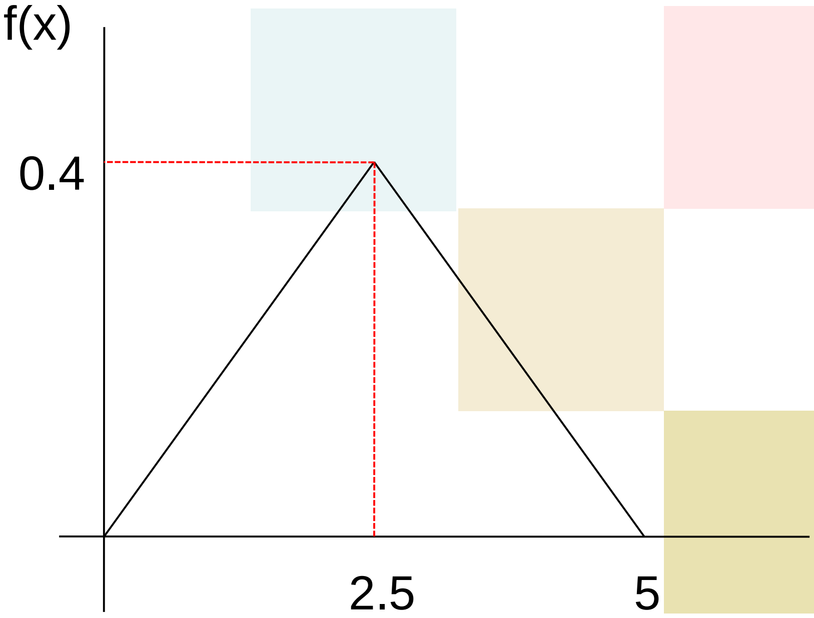
$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

# Variables aleatorias continuas

Ejemplo:

La función de densidad para el tiempo de uso de un tipo de máquinas durante un año (en horas x100) es:

$$f(x) = \begin{cases} \dfrac{0.4}{0.25}x & si\ 0 < x < 2.5 \\ 0.8 - \dfrac{0.4}{0.25}x & si\ 2.5 < x < 5 \\ 0 & en\ \ otro\ caso \end{cases}$$



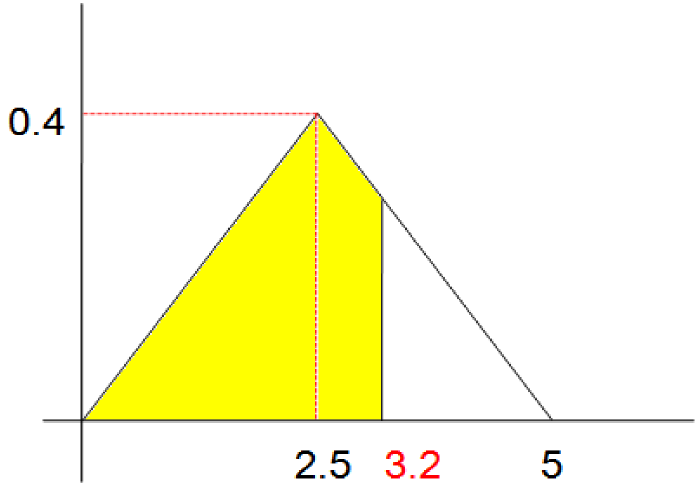
# Variables aleatorias continuas

Ejemplo:

¿Cuál es la probabilidad de que una máquina elegida al azar haya funcionado durante menos de 320 horas?

P(X < 3.2) =

$$\int_0^{3.2} f(x)dx = \int_0^{2.5} \frac{0.4}{0.25} x \, dx + \int_{2.5}^{3.2} \left(8 - \frac{0.4}{0.25} x\right) dx = 0{,}74$$

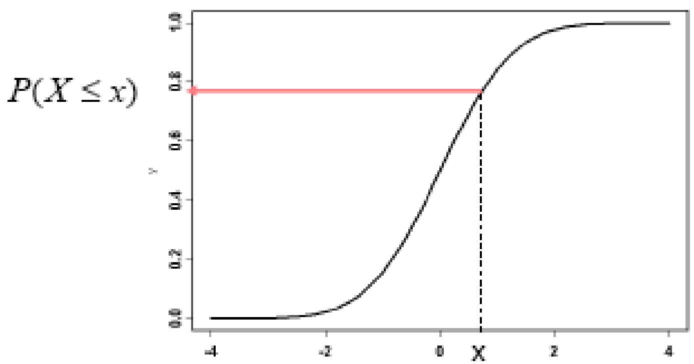


# Variables aleatorias continuas

## Función de distribución

Al igual que en el caso de variables discretas, podemos describir la distribución de una variable aleatoria continua mediante la **función de distribución**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$



# Variables aleatorias continuas

En el caso discreto la diferencia entre dos valores consecutivos de $F(x)$ proporciona la función de probabilidad. En el caso de variables continuas se cumple que:

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

# Variables aleatorias continuas

La función de distribución de una variable continua cumple las mismas propiedades que la de las variables discretas:

1. $\lim_{x \to -\infty} F(x) = 0$
2. $\lim_{x \to +\infty} F(x) = 1$
3. $F$ es no decreciente, es decir $F(x) \leq F(x + \epsilon)$, para $\epsilon > 0$
4. $F$ es continua por la derecha

Universidad Europea de Madrid

# Variables aleatorias continuas

Ejemplo:

La función de distribución para el tiempo de uso de un tipo de máquinas durante un año (en horas x100):

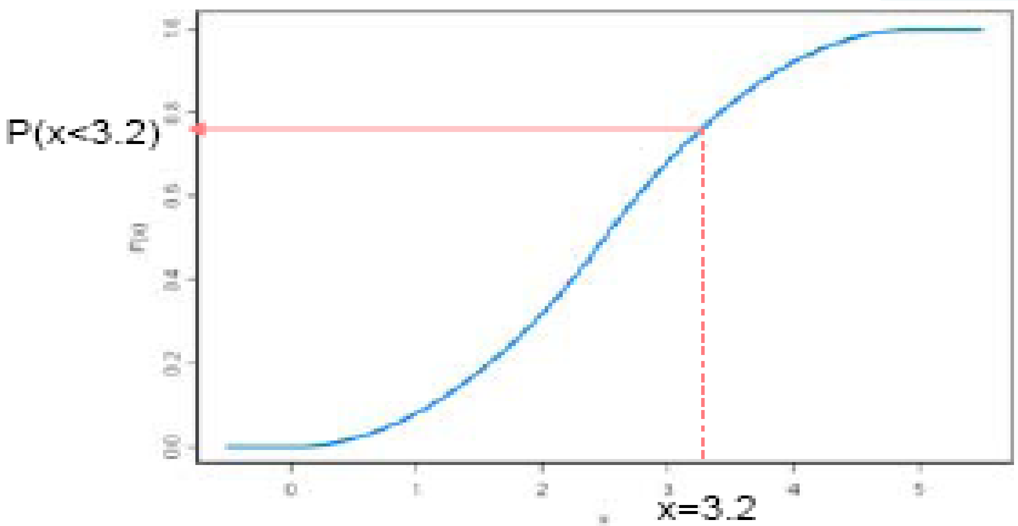
$$f(x) = \begin{cases} \dfrac{0.4}{0.25}x & si\ 0 < x < 2.5 \\ 0.8 - \dfrac{0.4}{0.25}x & si\ 2.5 < x < 5 \\ 0 & en\ \ otro\ caso \end{cases}$$

$$F(x) = \begin{cases} \displaystyle\int_0^x \frac{0.4}{0.25}u\, du & si\ 0 < x < 2.5 \\ \displaystyle\int_0^{2.5} \frac{0.4}{0.25}u\, du + \int_{2.5}^x \left(0.8 - \frac{0.4}{0.25}u\right) du & si\ 2.5 < x < 5 \\ 1 & si\ x \geq 5 \end{cases}$$

# Variables aleatorias continuas

Ejemplo:

$$F(x) = \begin{cases} 0.08x^2 & si\ 0 < x < 2.5 \\ -1 + 0.8x - 0.08x^2 & si\ 2.5 < x < 5 \\ 1 & si\ x \geq 5 \end{cases}$$



# Medidas características de una variable

## Medidas de centralización

### MEDIA

En el caso de una muestra de datos la media muestral:

$$\bar{x} = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \cdots + \frac{1}{n}x_n$$

a cada valor se le asigna un peso 1/n

La media μ o **esperanza** de una v.a. utiliza la probabilidad como peso:

$$\boxed{\begin{gathered}\mu = E[X] = \sum_i x_i p(x_i) \\ \mu = E[X] = \int_{-\infty}^{+\infty} x f(x) dx\end{gathered}}$$

$\mu = E[X] = \sum_i x_i p(x_i)$ → v.a. discreta

$\mu = E[X] = \int_{-\infty}^{+\infty} x f(x) dx$ → v.a. continua

# Medidas características de una variable

Algunas propiedades que cumple la esperanza

$$1.\ E\big(g(X)\big) = \begin{cases} \displaystyle\sum_{i=1}^{k} g(x_i)p(x_i) & si\ X\ es\ discreta \\ \displaystyle\int_{-\infty}^{+\infty} g(x)f(x)dx & si\ X\ es\ continua \end{cases}$$

$$2.\ E(aX + bY) = aE(X) + bE(Y) \longrightarrow$$ Linealidad de la esperanza

# Medidas características de una variable

## Medidas de centralización

### MEDIANA

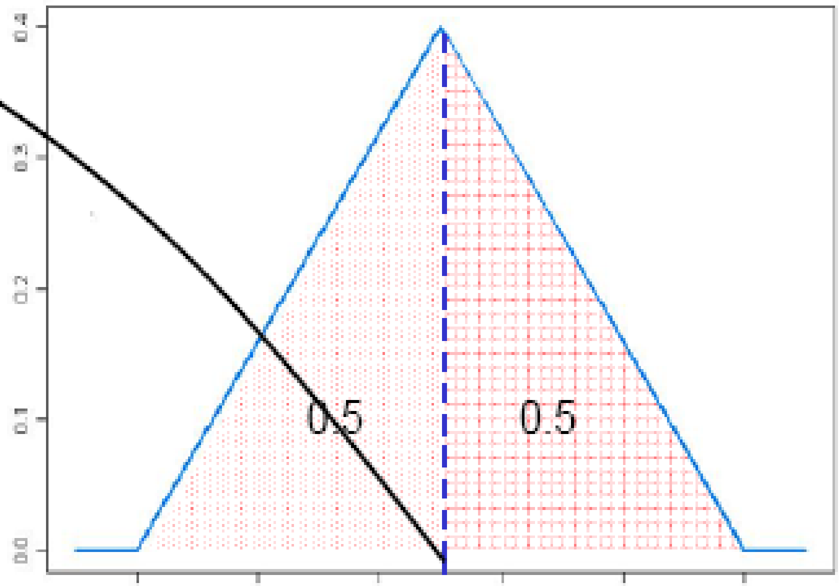
Intuitivamente: **Mediana** = valor que divide a la probabilidad total en dos partes iguales.

Definición: la mediana es el valor *m que cumple:*

$$F(m) = P(X \leq m) \geq ½ \;\; y \;\; P(X \geq m) \geq ½$$

$$P(X \leq m) = 0.5$$

$$F(m) \geq 0.5$$



# Medidas características de una variable

**Ejemplo** :En un puesto de feria se ofrece la posibilidad de lanzar a ciegas un dardo a unos globos. Si se consigue reventar un globo, se recibe un premio igual a una cantidad oculta tras el globo.

Supongamos que la probabilidad de acertar con algún globo es 1/3. Los premios se distribuyen de la siguiente manera:

- 40% de premios de 0.30 Euros
- 30% de premios de 0.6 Euros
- 20% de premios de 1.5 Euros
- 10% de premios de 6.0 Euros

Si cada lanzamiento cuesta 0.9 Euros, ¿Cúal es la "ganancia" esperada del dueño del puesto en cada lanzamiento?.

Universidad Europea de Madrid

# Medidas características de una variable

- **Solución:** Caracterizamos la Ganancia del  Dueño:

$$G = \begin{cases} -\mathbf{5.1} & \text{premio de } 6€ \\ -\mathbf{0.60} & \text{premio de } 1.50€ \\ \mathbf{0.30} & \text{premio de } 0.60€ \\ \mathbf{0.60} & \text{premio de } 0.30€ \\ \mathbf{0.90} & \text{no premio} \end{cases}$$

# Medidas características de una variable

Calculamos la función de masa:

- P(G=-5.1)=(1/3)(10/100)
- P(G=-0.60)= (1/3)(20/100)
- P(G=0.30)= (1/3)(30/100)
- P(G=0.60)= (1/3)(40/100)
- P(G=0.90)=2/3

E[G]=(- 5.1)(10/300)+(-0.60)(20/300)+(0.30)(30/300)+(0.60)(40/300)+(0.90)(2/3)=0.5€

Universidad Europea de Madrid

# Esperanza matemática

**Ejemplo**: Una empresa de refrescos lanza una oferta para anunciar un nuevo producto. Asegura que en cada 1000 chapas hay 500 con *inténtelo otra vez*, 300 con premio de 0.30 euros, 150 con premio de 0.60 euros, 40 con premio de 3.00 euros y 10 con premio de 6.00 euros.

Un individuo, al que no le gusta el refresco, decide comprar una botella cuyo coste es de 0.6 euros.

Caracterizar su ganancia mediante una variable aleatoria. ¿Es razonable su decisión?. Calcular la probabilidad de perder dinero

# Esperanza matemática

- **Solución:** Caracterizamos la Ganancia de la persona que compra la botella:

$$G = \begin{cases} -0.60 & \text{si no tiene premio} \\ -\mathbf{0.}30 & \text{premio de } 0.30€ \\ \mathbf{0} & \text{premio de } 0.60€ \\ 2\mathbf{.}4\mathbf{0} & \text{premio de } 3€ \\ 5\mathbf{.}4\mathbf{0} & \text{premio de } 6€ \end{cases}$$

# Esperanza matemática

- **¿Es razonable su decisión?**. La decisión de comprar la botella, solo por el premio (recordemos que no le gusta su contenido) será razonable si el valor esperado de la ganancia es positivo

Calculamos la función de masa:

- P(G=-0.60)=500/1000=0.5
- P(G=-0.30)= 300/100=0.3
- P(G=0)= 150/1000=0.15
- P(G=2.40)= 40/1000=0.04
- P(G=5.40)=10/1000=0.01

E[X]=(-0.60)·0.5+(-0.30) ·0.3

+0·0.15+2.40·0.04+5.40·0.01

= - 0.24

Universidad Europea de Madrid

# Medidas características de una variable

Ejemplo:

¿Cuál es el tiempo medio de funcionamiento de las máquinas?

$$f(x) = \begin{cases} \dfrac{0.4}{0.25}x & si\ 0 \leq x < 2.5 \\ 0.8 - \dfrac{0.4}{0.25}x & si\ 2.5 \leq x < 5 \\ 0 & en\ \ otro\ caso \end{cases}$$

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{2.5} \frac{0.4}{2.5}x^2 dx + \int_{2.5}^{5} \left(0.8x - \frac{0.4}{2.5}x^2\right) dx$$

$$= 2.5$$

Universidad Europea de Madrid

# Momentos de una variable aleatoria

- Los **momentos** de una variable aleatoria son los valores esperados de ciertas funciones de la variables. Constituyen un conjunto de medidas descriptivas que pueden emplearse para caracterizar la distribución y especificarla si todos los datos son conocidos.

- Se pueden calcular momentos respecto al origen y momentos respecto a la media.

Universidad Europea de Madrid

# Momentos de una variable aleatoria

- Momento de orden *r* respecto al origen

$$E[x^r] = \begin{cases} \sum x^r p(x) & \text{si X es discreta} \\ \int \mathrm{x}^{\mathrm{r}} f(x)dx & \text{si X es continua} \end{cases}$$

- Momento de orden *r* respecto a la media

$$E[(x-\mu)^r] = \begin{cases} \sum (x-\mu)^r p(x) & \text{si X es discreta} \\ \int (\mathrm{x} - \mu)^{\mathrm{r}} f(x)dx & \text{si X es continua} \end{cases}$$

Universidad Europea de Madrid

# Momentos de una variable aleatoria

## Observaciones:

- El momento respecto al origen de orden 1 es la Esperanza.

$$E[x] = \mu$$

- El momento respecto a la media de orden 0 es 1.
- El momento respecto a la media de orden 1 es 0.
- El momento respecto a la media de orden 2 es la VARIANZA.

# Medidas características de una variable

## Medidas de dispersión

### VARIANZA

En el caso de una muestra de datos la varianza muestral es una medida de la dispersión de los datos:

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

La **Varianza** de una v.a. utiliza la probabilidad como peso**:**

$$\boxed{\begin{array}{l} \sigma^2 = Var[X] = \sum_i (x_i - \mu)^2 p(x_i) \\ \sigma^2 = Var[X] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \end{array}}$$

→ v.a. discreta

→ v.a. continua

Universidad Europea de Madrid

# Medidas características de una variable

Forma alternativa para el cálculo de la varianza

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

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

**Demostración:**

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

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

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

$E[X]$ es una constante, no depende de $X$

Es un operador lineal

# Medidas características de una variable

Algunas propiedades que cumple la varianza

$$Var[a+bX]=0+b^2Var[X]=b^2Var[X]$$

$$Var[X+Y]=Var[X]+Var[Y]+2Cov[X,Y]$$

Universidad Europea de Madrid

# Medidas características de una variable

## Ejemplo:

Una centralita telefónica tiene 5 líneas. Sea

X = {número de líneas ocupadas en una hora}.

Espacio muestral: $\Omega$ = {0,1,2,3,4,5} .

Supongamos que su función de probabilidad viene dada por:

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| P(X = x) | 0.14 | 0.27 | 0.27 | 0.18 | 0.09 | 0.05 |

**Calculemos E[X] y Var[X]**

Universidad Europea de Madrid

# Medidas características de una variable

**Calculamos E[X] y Var[X]**

$$E[X] = 0 \cdot 0{,}4 + 1 \cdot 0{,}27 + 2 \cdot 0{,}27 + 3 \cdot 0{,}18 + 4 \cdot 0{,}09 + 5 \cdot 0{,}05 = 1{,}96$$

$$Var[X] = (0 - 1{,}96)^2 \cdot 0{,}4 + (1 - 1{,}96)^2 \cdot 0{,}27 + (2 - 1{,}96)^2 \cdot 0{,}27 +$$
$$(3 - 1{,}96)^2 \cdot 0{,}18 + (4 - 1{,}96)^2 \cdot 0{,}09 + (5 - 1{,}96)^2 \cdot 0{,}05 = 1{,}82$$

**Otra forma**:

$$E[X^2]=0^2 \cdot 0.4+1^2 \cdot 0.27+2^2 \cdot 0.27+3^2 \cdot 0.18+4^2 \cdot 0.09+5^2 \cdot 0.05=5.66$$
$$Var[X]=E[X^2]-(E[X])^2 =5.66 -(1.96)^2= 1.82$$

# Medidas características de una variable

**Ejemplo:** Calcular la varianza del tiempo de funcionamiento de las máquinas

$$f(x) = \begin{cases} \frac{0.4}{0.25}x & si\ 0 \leq x < 2.5 \\ 0.8 - \frac{0.4}{0.25}x & si\ 2.5 \leq x < 5 \\ 0 & en\ \ otro\ caso \end{cases}$$

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{0}^{2.5} \frac{0.4}{2.5}x^2dx + \int_{2.5}^{5} \left(0.8x - \frac{0.4}{2.5}x^2\right) dx$$

$$= 2.5$$

Universidad Europea de Madrid

# Medidas características de una variable

- Aplicamos las formula : Var[X]=E[X²]-(E[X])²

$$E[X^2] = \int x^2 f(x)dx = \int_0^{2.5} x^2 \frac{0.4}{0.25} x dx + \int_{2.5}^{5} x^2 \left(0.8 - \frac{0.4}{0.25} x\right) dx =$$

$$= \int_0^{2.5} 1.6x^3 dx + \int_{2.5}^{5} (0.8x^2 - 1.6x^3) dx = \left.\frac{1.6}{4} x^4\right|_0^{2.5} + \left.\frac{0.8}{3} x^3 - \frac{1.6}{4} x^4\right|_{2.5}^{5} =$$

Var[X]= - ( 2.5 )$^2$

# Medidas características de una variable

## Desigualdad de Chebyshev

Si X es una variable aleatoria con:

$$E[X] = \mu \quad y \quad Var[X] = \sigma^2$$

Se puede demostrar que gran parte de la distribución está situada en un intervalo centrado en μ y que tiene amplitud varias veces σ.

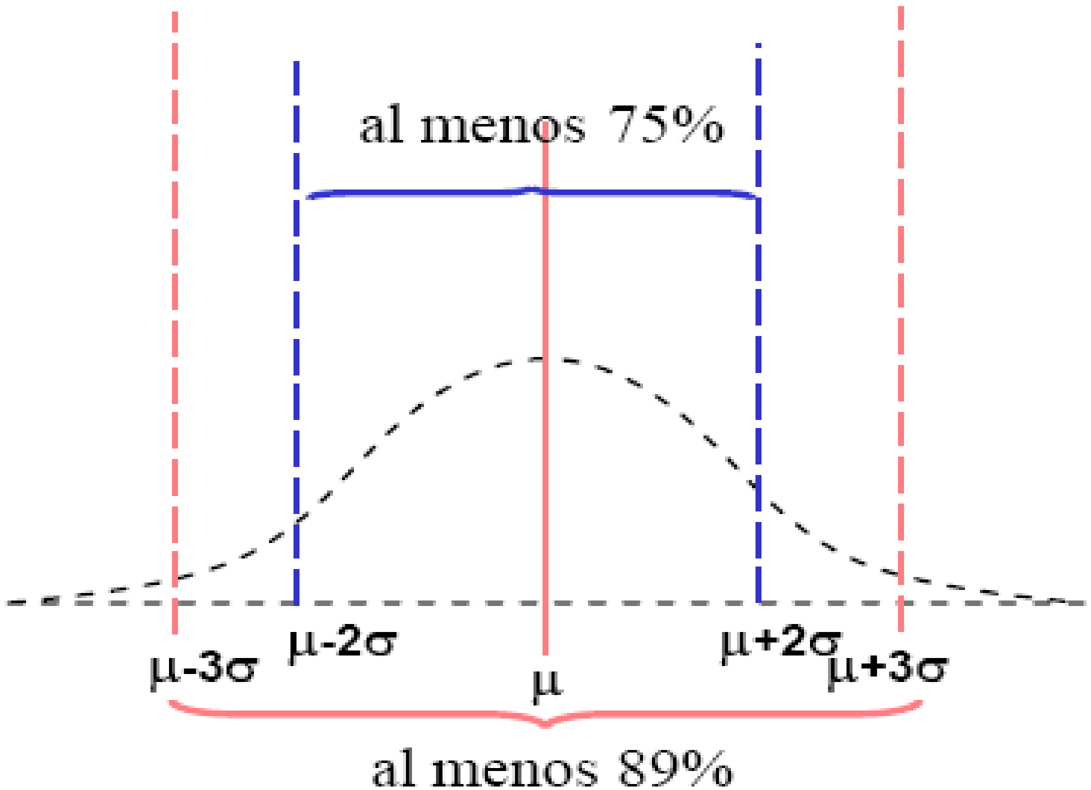
En concreto:

$$\forall\, c > 0 \qquad P(\mu - c\sigma \leq X \leq \mu + c\sigma) \geq 1 - \frac{1}{c^2}$$

Es decir, la probabilidad de realizar una observación de una variable y que esté en ese intervalo es mayor o igual que $1 - \frac{1}{c^2}$

# Medidas características de una variable

## Desigualdad de Chebyshev



# Medidas características de una variable

## Desigualdad de Chebyshev

| $c$ | Prob dentro $\pm c\ \sigma$ | Prob. fuera de $\pm c\ \sigma$ |
|---|---|---|
| 2 | 75% | 25% |
| 3 | 89% | 11% |
| 4 | 94% | 6% |
| 5 | 96% | 4% |

# Medidas características de una variable

Ejemplo 1:

El coste diario para usar una herramienta tiene una media de 3 euros y una varianza de 4 euros. ¿Con que frecuencia el coste diario será mayor de 8 euros?

$$P(X > 8) = P\left(\frac{X-3}{2} > \frac{5}{2}\right) \leq P\left(\left|\frac{X-3}{2}\right| > \frac{5}{2}\right) \leq \frac{1}{\left({}^{5}/_{2}\right)^{2}} = \frac{4}{25} = 0{,}16$$

Es decir, menos del 16% de las veces el coste diario será de más de 8 euros.

# Medidas características de una variable

Ejemplo 2:

Un sistema informático tiene una media de 4 fallos por semana, con una desviación estándar de 0,8 fallos por semana.

1. Determinar un intervalo que incluya al menos al 90% de los resultados semanales del número de fallos.

2. Se garantiza que el número de descomposturas rara vez será mayor que 8 en un período de una semana. ¿Es correcta esta afirmación?. ¿Por qué?

# Medidas características de una variable

Solución:

1. Determinar un intervalo que incluya al menos al 90% de los resultados semanales del número de descomposturas.

   Si X = número de fallos en una semana, entonces, por Chebyshev,

   buscamos *c* tal que

   $P(|X - 4| > c \cdot 0{,}8) \leq 1 - 1/c^2 = 0{,}9 \Rightarrow c = 3{,}16$

   El intervalo es por tanto: [2 , 7]

2. Se garantiza que el número de descomposturas rara vez será mayor que 8 en un período de una semana. ¿Es correcta esta afirmación?. ¿Por qué?

   Si es correcta, ya que

$$P(X > 8) = P\left(\frac{X-4}{0{,}8} > \frac{4}{0{,}8}\right) \leq P\left(\left|\frac{X-4}{0{,}8}\right| > \frac{4}{0{,}8}\right) \leq \frac{1}{\left({}^{4}/_{0{,}8}\right)^{2}} = 0{,}04$$

# Independencia de variables

Dos variables X e Y se dicen **independientes** si se verifica que

$$\forall\, A, B \subset \mathbb{R} \qquad P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\})$$

En particular, si X e Y son independientes se tiene

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x)P(Y \leq y)$$

**Propiedad**:

si X e Y son independientes, entonces

1. $E[X \cdot Y] = E[X] \cdot E[Y]$
2. $Var[X + Y] = Var[X] + Var[Y]$