MOMENTOS DE UNA VARIABLE ALEATORIA UNIDIMENSIONAL

TEMA 1

Xabier Erdocia
Itsaso Leceta

OBJETIVOS

- ✓ Comprender e interiorizar el concepto de variable aleatoria unidimensional
- ✓ Ser capaz de calcular los momentos y el valor medio de una variable aleatoria
- ✓ Cálculo e interpretación adecuada de medidas de posición y dispersión mediante momentos
- ✓ Ser capaz de encontrar la mínima probabilidad de que la variable aleatoria se encuentre dentro de un rango, utilizando la desigualdad de Tchebychev

ÍNDICE

1.1. Concepto de variable aleatoria unidimensional
1.2. Valor medio
1.3. Momentos de la variable aleatoria
1.4. Medidas de posición. Medidas de dispersión. Coeficientes de Fisher.
1.5. Desigualdad de Tchebychev

1.1. Concepto de variable aleatoria unidimensional

eman ta zabal zazu
UPV EHU

- Sea E un ensayo aleatorio y S su espacio muestral. La **variable aleatoria** es una función en la que a cada posible resultado en S se le asigna un número real.

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

- La variable aleatoria puede ser discreta o continua:

  - **Discreta**: cuando el número de imágenes es finito o infinito cuantificable

  - **Continua**: cuando la imagen puede tomar todos los valores de un rango

- Supongamos el siguiente ensayo: “tirar una moneda al aire 3 veces”.

- La variable aleatoria X puede definirse como “número de cruces obtenidas”. La variable aleatoria asigna un valor real a todos los posibles resultados obtenidos en el ensayo.

- Los posibles valores de la variable aleatoria (resultados) son:

  - X=0 ⇒ 3 caras ○○○

  - X=1 ⇒ 1 cruz y 2 caras

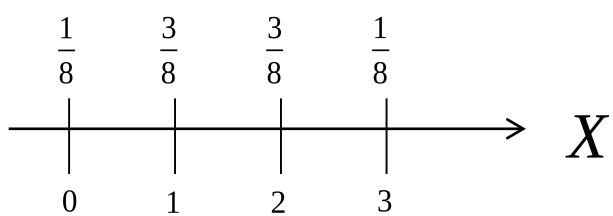
    ✕○○ ○✕○ ○○✕

  - X=2 ⇒ 2 cruces y 1 cara

    ✕✕○ ○✕✕ ✕○✕

  - X=3 ⇒ 3 cruces ✕✕✕

- La probabilidad de cada caso se puede calcular:
- La cantidad total de resultados posibles es 8:
  - X=0 ⇨ ○○○ $P(X=0)=\frac{1}{8}$
  - X=1 ⇨ ✖○○ ○✖○ ○○✖ $P(X=1)=\frac{3}{8}$
  - X=2 ⇨ ✖✖○ ○✖✖ ✖○✖ $P(X=2)=\frac{3}{8}$
  - X=3 ⇨ ✖✖✖ $P(X=3)=\frac{1}{8}$



- **Distribución de probabilidad**: representa la distribución de probabilidad entre todos los resultados posibles de una prueba para la variable aleatoria definida.

    - Para analizar distribuciones discretas, *la función de probabilidad* es:

    $$p(x) = P(X = x) \,/\, \forall x \in \mathbb{R}$$

        - Es decir, la probabilidad de que la variable tome un valor concreto.

    - Para analizar distribuciones continuas se define la *función de densidad* de probabilidad ($\mathrm{f}(x)$) en un intervalo A :

    $$P(X \in A) = \int_A \mathrm{f}(x)dx$$

        - Es decir, representa la distribución de la probabilidad por unidad de longitud en cada punto.
        - Se debe cumplir lo siguiente: $\int_{-\infty}^{\infty} \mathrm{f}(x)dx = 1$

- Para definir variables aleatorias (discretas o continuas), también se utiliza la *función de distribución* ($F(x)$):

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

1.2. Valor medio

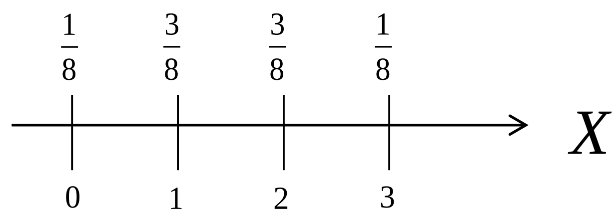
- **Cuando la variable es discreta:**

  - Sea X una variable aleatoria unidimensional que toma valores $\{x_1, x_2, \ldots, x_n\}$ con una función de probabilidad *p(x)*. Por definición, el valor medio es un valor numérico calculado como:

$$E(X) = \sum_{i=1}^{n} x_i \, p(x_i)$$

    - El valor promedio también se llama expectativa matemática o valor esperado.

- Ejemplo: tomemos el ejemplo anterior de lanzar una moneda al aire 3 veces:

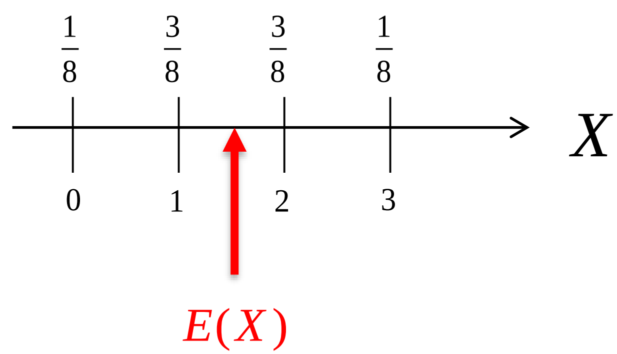


El valor medio es el siguiente:

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$

- Se puede observar que el valor medio es el **centro de gravedad** de la masa de distribución de la variable

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}$$



Si supusiéramos que el eje X es una barra sin peso en la que las masas colocadas son iguales a *p(x)* y la barra se sostiene en el punto E(X), esta barra estaría en equilibrio.

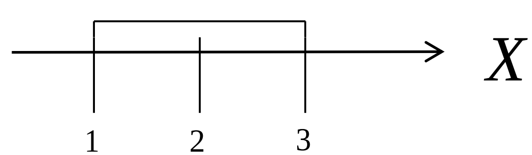
- **Cuando la variable es continua**:

  - Sea X variable unidimensional aleatoria con función de densidad f(x). Por definición, el valor medio es un valor numérico y se calcula como:

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

  - Ejemplo: la variable X está definida por la función de densidad:

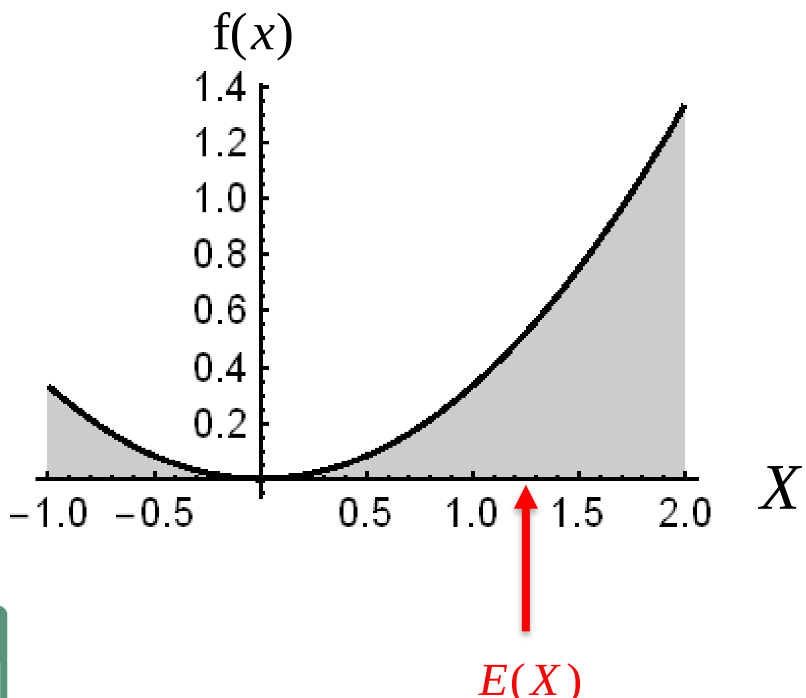
$$\mathrm{f}(x) = \begin{cases} \dfrac{x^2}{3} & -1 < x < 2 \\ 0 & \text{otros casos} \end{cases}$$



- El valor medio es el siguiente:

$$f(x) = \begin{cases} \dfrac{x^2}{3} & -1 < x < 2 \\ 0 & \text{otros casos} \end{cases}$$

$$E(X) = \int_{-1}^{2} x \frac{x^2}{3} dx = \frac{5}{4}$$



En variables continuas la interpretación es la misma, la masa cambia continuamente y E(X) es el centro de gravedad

1.3. Momentos de la variable aleatoria

- El k-ésimo momento estadístico de una variable aleatoria está definido por el orden y el punto de aplicación.

$$\alpha_{k,a}$$

- donde k es el orden y a es el punto de aplicación.

- El momento estadístico es el valor medio de la diferencia entre la variable y el punto de aplicación, elevado a la potencia k .

$$\alpha_{k,a} = E(X-a)^k$$

- Cuando se desconoce una distribución, las diferentes características pueden dar una visión general de la misma. Los momentos son estas características de la distribución.

Por lo tanto,

- Para variables discretas:

$$\alpha_{k,a} = E(X-a)^k = \sum_{i=1}^{n} (x_i - a)^k \, p(x_i)$$

- Para variables continuas :

$$\alpha_{k,a} = E(X-a)^k = \int_{-\infty}^{\infty} (x-a)^k \, \mathrm{f}(x) dx$$

- Los momentos estadísticos se pueden clasificar según el origen:

  - $a=0$, momentos no centrados

$$a=0 \Rightarrow \alpha_{k,0}=\alpha_k \qquad \alpha_k=E(X)^k$$

  - $E(X)$, momento centrado en la media

$$a=E(X) \Rightarrow \alpha_{k,E(X)}=\mu_k \qquad \mu_k=E(X-E(X))^k$$

  - $a\neq 0$, momento centrado en a

$$a\neq 0 \ \wedge \ a\neq E(X) \Rightarrow \alpha_{k,a} \qquad \alpha_{k,a}=E(X-a)^k$$

1.4. Medidas de posición. Medidas de dispersion. Coeficientes de Fisher.

- Cuantificando los valores de los momentos podemos obtener medidas de posición, dispersión o forma.

MEDIDAS DE TENDENCIA CENTRAL:

- **Media**

La media es el momento no centrado (o centrado en el origen) de orden uno.

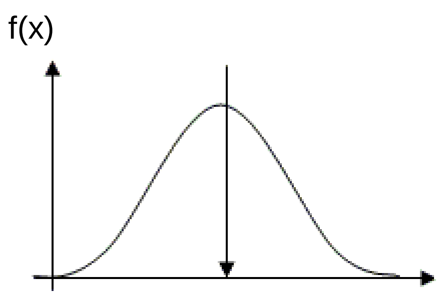
$$k = 1 \Rightarrow \alpha_{1,0} = \alpha_1 \qquad \alpha_1 = E(X)^1 = E(X)$$

Como se indica, la media representa la posición del centro de gravedad de la masa de distribución.

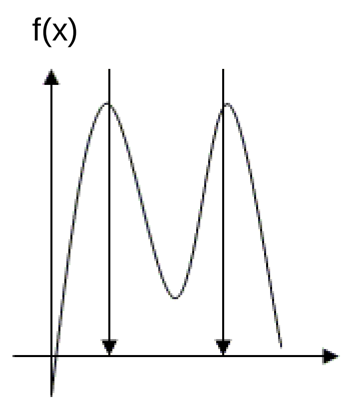
➢ Moda

La moda es el valor que aparece con mayor frecuencia en una distribución.

Es decir, el valor de la variable que hace máxima la función de probabilidad o la función de densidad (discreta o continua según la variable aleatoria). Por ejemplo en un caso continuo:



Un valor para la moda



Bimodal

Mediana

La mediana es el valor de la variable aleatoria que divide la distribución en dos. La mediana es el valor de cualquier x que satisfaga:

$$P(X \leq x) = 0,5 \quad \text{y} \quad P(X \geq x) = 0,5$$

MEDIDAS DE DISPERSION:

Las medidas de tendencia central describen de alguna manera el centro de la distribución. Sin embargo, las medidas de dispersión describen qué tan dispersos son los valores.

- **Varianza**

La varianza de una variable aleatoria es el momento de segundo orden centrado en la media

$$a = E(X) \Rightarrow \alpha_{2,E(X)} = \mu_2 \qquad \mu_2 = \sigma^2 = E(X - E(X))^2$$

- Si la variable es continua:

$$\sigma^2 = E(X - E(X))^2 = \int_{-\infty}^{\infty} (x - E(X))^2 \, \mathrm{f}(x)dx$$

- Si la variable es discreta:

$$\sigma^2 = E(X - E(X))^2 = \sum_{i=1}^{n} (x_i - E(X))^2 \, p(x_i)$$

La varianza se puede interpretar como el momento de inercia de la distribución con respecto a la media. Los valores bajos de la varianza indican que la masa de la variable se concentra alrededor de la media. Sin embargo, valores altos de varianza indican que la masa está dispersa.

➢ **Desviación estándar**

La desviación estándar es la raíz cuadrada positiva de la varianza.

$$\sigma = \sqrt{\sigma^2}$$

COEFICIENTE DE FISHER

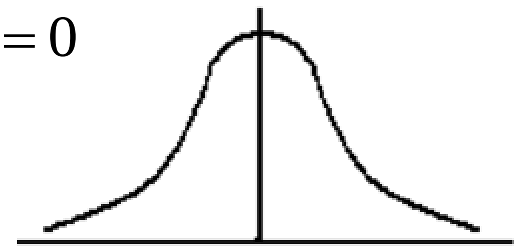
Describen la forma de la distribución.

- Coeficiente de asimetría: asimetría

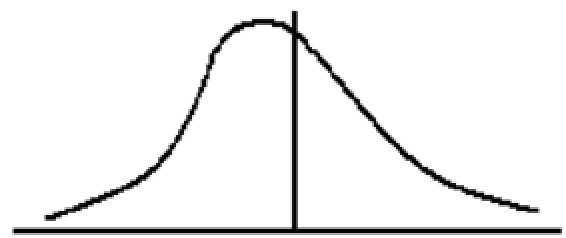
  Da información sobre la simetría de la distribución.

$$\gamma_1 = \frac{1}{\sigma^3} E(X - E(X))^3 = \frac{\mu_3}{\sigma^3}$$

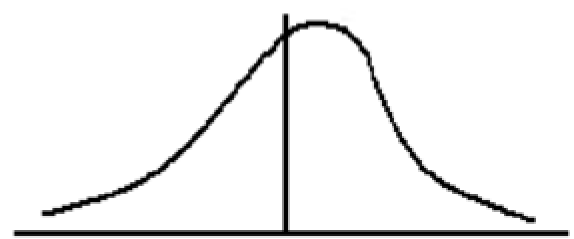
- Cuando la distribución es simétrica: $\gamma_1 = 0$



- Sesgo positivo: $\gamma_1 > 0$



- Sesgo negativo: $\gamma_1 < 0$

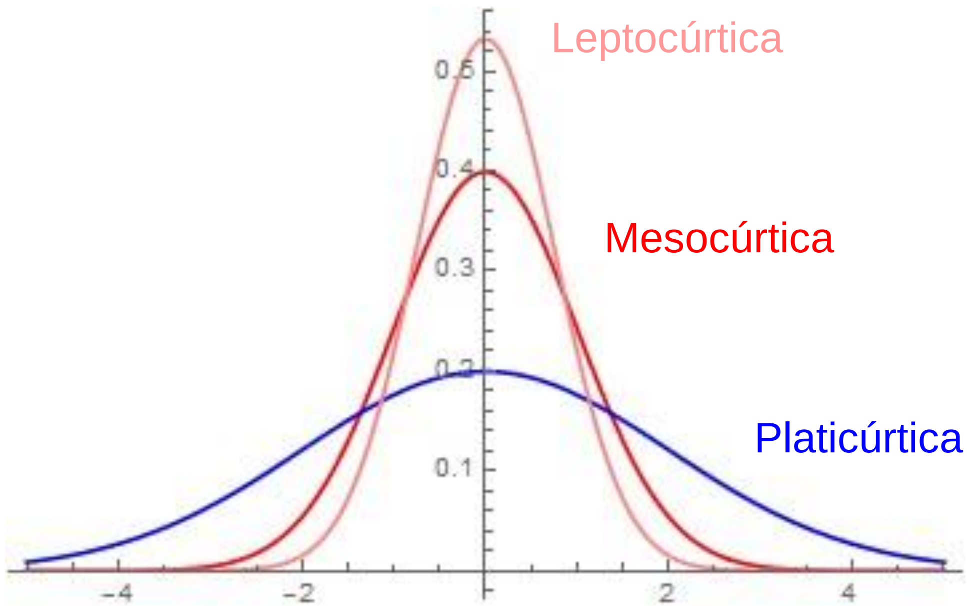


- Coeficiente de **curtosis**

Da información sobre el "pico" de la distribución.

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

- Cuando la distribución es mesocúrtica: $\gamma_2 = 0$
- Cuando la distribución es leptocúrtica: $\gamma_2 > 0$
- Cuando la distribución es platicúrtica: $\gamma_2 < 0$



1.5. Desigualdad de Tchebychev

- Sea X la variable aleatoria con valores finitos de media E(X) y varianza σ^2. La desigualdad de Tchebychev basada en la diferenciación de Markov es:

$$P(E(X)-k\sigma < X < E(X)+k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k>0$$

$$P(|X-E(X)|<k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k>0$$

- Cuando se desconoce la distribución de la variable, utilizando la desigualdad de Tchebychev podemos calcular la probabilidad mínima de que la variable aleatoria se encuentre dentro de un rango. Este valor de probabilidad se puede calcular. Para ello, necesitamos conocer la media y la varianza de dicha variable aleatoria.

- A través de esta desigualdad obtendremos información sólo cuando $\forall k>1$

- Ejemplo: Sea una variable aleatoria con media 50 y varianza 16. Calcular la mínima probabilidad de que la variable esté dentro del rango (42, 58).

$$P(E(X)-k\sigma < X < E(X)+k\sigma) \geq 1-\frac{1}{k^2} \quad \forall k>0$$

$$P(42 < X < 58) \geq 1-\frac{1}{k^2} \quad \forall k>0$$

$$42 = E(X)-k\sigma \qquad 42 = 50-k4 \qquad k=2 \qquad \text{o}$$

$$58 = E(X)+k\sigma \qquad 58 = 50+k4 \qquad k=2$$

$$P(42 < X < 58) \geq 1-\frac{1}{2^2} = 0,75$$