Tema 3. Variables aleatorias

Estadística

Ángel Serrano Sánchez de León

Índice

- Definición.
- Variables aleatorias continuas y discretas.
- Medidas características de una variable aleatoria.
  - Esperanza matemática.
  - Varianza y desviación típica.
- Variables aleatorias bidimensionales.
  - Esperanza matemática.
  - Varianza.
  - Covarianza.

Definición

- Una **variable aleatoria** (o estocástica) X es una función definida sobre el espacio muestral S que asigna un número real a cada uno de los puntos, o resultados posibles, de dicho espacio muestral.
- Su valor depende del resultado del experimento y por tanto no se conoce de antemano.
- Notación: variable → mayúsculas, valor concreto de la variable → minúsculas.
- Ejemplo: Sea X la variable que representa el resultado de lanzar un dado. Tras lanzarlo, vemos que sale $x = 2$.

Tipos

- **Variable aleatoria discreta:** toma un conjunto finito o infinito de valores que se pueden contar (representado por números enteros).
  - Ej.: nº de accidentes al día en un tramo de carretera.
  - Ej.: resultado de un dado.

- **Variable aleatoria continua:** toma todos los valores posibles en un intervalo (no se puede contar, pero se puede medir).
  - Ej.: peso de una pieza en una fábrica.
  - Ej.: temperatura de una habitación.

Variable aleatoria discreta

- Sea una variable aleatoria discreta X, que toma los valores $S = \{x_1, x_2, x_3, \ldots\}$.
- Se llama **función de probabilidad (función de masa de probabilidad)** de la variable discreta X a la función que asocia cada valor x con la probabilidad de que X tome ese valor.

$$f(x) \equiv P(X = x)$$

- Propiedades:

$$f(x_i) \geq 0, \ \forall x_i \in S$$

$$\sum_i f(x_i) = 1$$

5

Variable aleatoria discreta

- Se llama **función de distribución (o de frecuencia acumulada)** a la función que asocia cada valor de *x* con la probabilidad de que la variable aleatoria *X* tome un valor menor o igual que *x*.

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

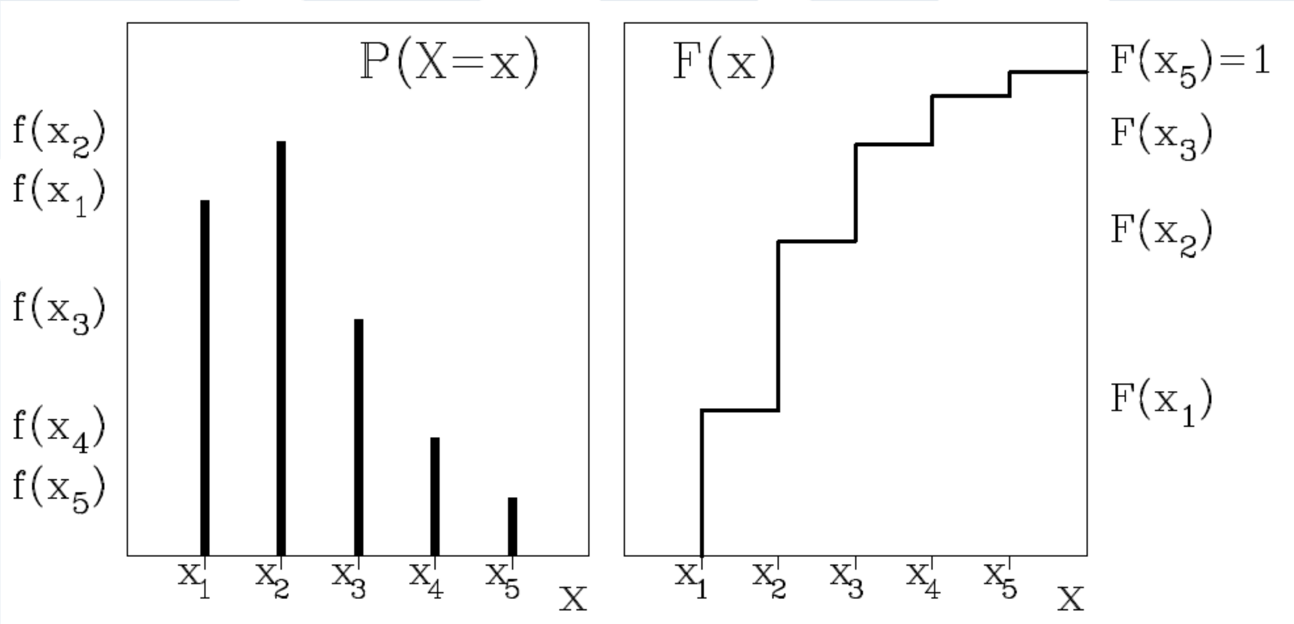
- Propiedades: $F(-\infty) = 0, \quad F(\infty) = 1$

$$F(x_i) = \sum_{x < x_i} f(x_i) = F(x_{i-1}) + f(x_i)$$

$$P(x_i < X \leq x_j) = F(x_j) - F(x_i)$$

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

Variable aleatoria discreta



Ejemplo

- X = suma de los resultados de 2 dados.

Resultados de los dados	x_i	$f(x_i)$	$F(x_i)$
(1,1)	2	1/36	1/36
(2,1) (1,2)	3	2/36	3/36
(3,1) (2,2) (1,3)	4	3/36	6/36
(4,1) (3,2) (2,3) (1,4)	5	4/36	10/36
(5,1) (4,2) (3,3) (2,4) (1,5)	6	5/36	15/36
(6,1) (5,2) (4,3) (3,4) (2,5) (1,6)	7	6/36	21/36
(6,2) (5,3) (4,4) (3,5) (2,6)	8	5/36	26/36
(6,3) (5,4) (4,5) (3,6)	9	4/36	30/36
(6,4) (5,5) (4,6)	10	3/36	33/36
(6,5) (5,6)	11	2/36	35/36
(6,6)	12	1/36	36/36 = 1

Ejemplo (continuación)

- La probabilidad de que la suma de los dos dados esté en el rango $4 < x \leq 7$ es:

$$P(4 < x \leq 7) = F(7) - F(4) = \frac{21}{36} - \frac{6}{36} = \frac{15}{36}$$

- Otra manera:

$$P(4 < x \leq 7) = f(5) + f(6) + f(7) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} = \frac{15}{36}$$

Ejemplo (continuación)

- Probabilidad de que la suma sea mayor que 10:

$$P(x > 10) = 1 - F(10) = 1 - \frac{33}{36} = \frac{3}{36}$$

- Otra manera:

$$P(x > 10) = f(11) + f(12) = \frac{2}{36} + \frac{1}{36} = \frac{3}{36}$$

Variable aleatoria continua

- Sea una variable *X* que toma todos los valores posibles en el intervalo (–∞, ∞).
- La probabilidad de que la variable tome un valor concreto *x* es cero (hay infinitos valores posibles).
- Se define **función de densidad de probabilidad** (o **distribución de probabilidad**), *f(x)*, como aquella que cumple:

$$f(x) \geq 0, \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

Variable aleatoria continua

- *f(x)dx* representa la probabilidad de que la variable *X* tome un valor comprendido entre *x* y *x+dx*.
- La probabilidad de que *X* esté comprendida entre los valores x_1 y x_2 es:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

- En el cálculo anterior podemos sustituir el signo < por ≤, ya que:

$$P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$$

Variable aleatoria continua

- La **función de distribución** *F(x)* de una variable aleatoria continua es la probabilidad de que la variable *X* tome un valor inferior a *x*.

$$F(x) = P(X < x) = \int_{-\infty}^{x} f(t)dt$$

- Propiedades:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx = \int_{-\infty}^{x_2} f(x)dx - \int_{-\infty}^{x_1} f(x)dx = F(x_2) - F(x_1)$$

$$F(-\infty) = 0, \quad F(\infty) = 1 \qquad\qquad f(x) = \frac{dF(x)}{dx}$$

Variable aleatoria continua

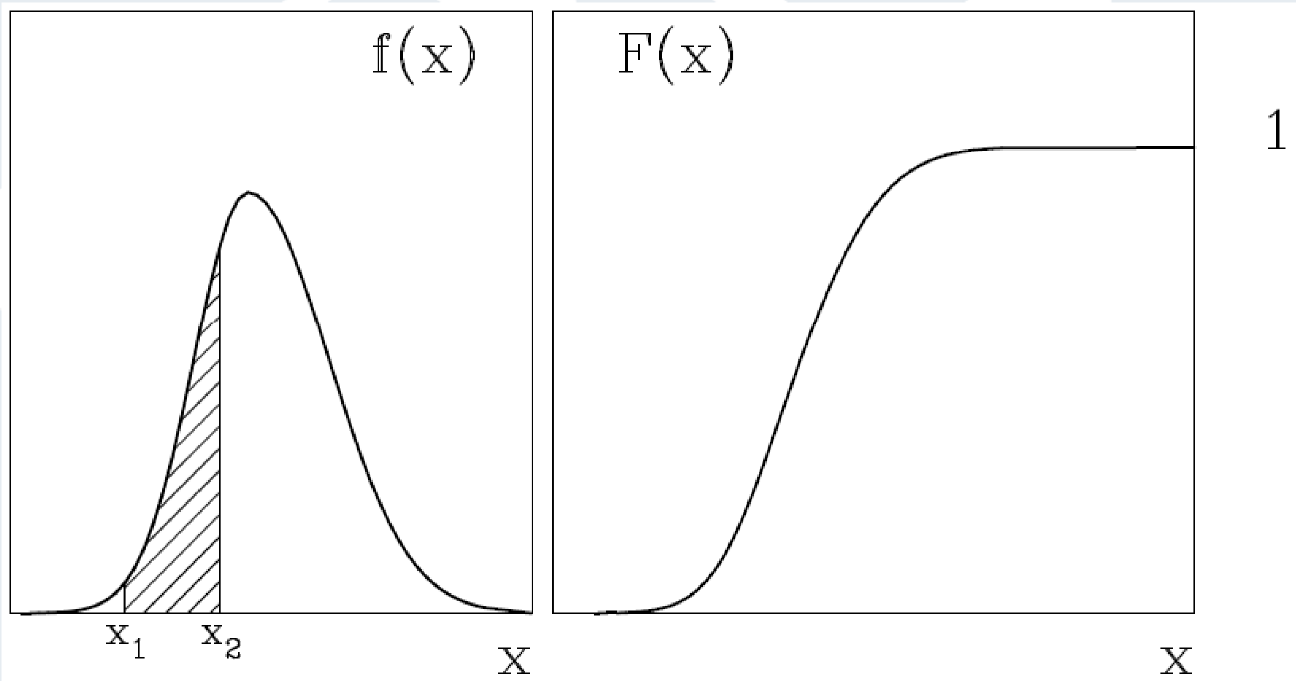
- En el caso de que la variable X solo tome valores en el intervalo (a,b), entonces las integrales definidas anteriores serán entre esos dos valores.

$$\int_a^b f(x)dx = 1$$

- La función de distribución sería:

$$F(x) = \begin{cases} 0, & x < a \\ \int_a^x f(t)dt & a < x < b \\ 1, & x > b \end{cases}$$

Variable aleatoria continua



Medidas características de una variable aleatoria

- Igual que para las distribuciones de frecuencias, definimos **medidas características de las variables aleatorias**.
  - **Centralización**: valor esperado o esperanza matemática.
  - **Dispersión**: desviación típica, varianza.
- Por convenio, se usan **letras griegas** para las distribuciones de variables aleatorias y letras latinas para las distribuciones de frecuencias (tomadas de una muestra).

Esperanza matemática o valor esperado

- Sea una variable aleatoria discreta *X* que toma los valores x_1, x_2, …, y sea *f(x)* su función de probabilidad.
- La **media**, **esperanza matemática** o **valor esperado** μ, también representado por *E(X)*, de *X* es:

$$\mu \equiv E(X) = \sum_i x_i f(x_i)$$

- En el caso de una variable aleatoria continua:

$$\mu \equiv E(X) = \int_{-\infty}^{\infty} x \; f(x) dx$$

- Representa un valor típico de la variable aleatoria.

Ejemplo

- X = suma de los resultados de 2 dados.

Resultados	x_i	$f(x_i)$	$x_i\, f(x_i)$
(1,1)	2	1/36	2/36
(2,1) (1,2)	3	2/36	6/36
(3,1) (2,2) (1,3)	4	3/36	12/36
(4,1) (3,2) (2,3) (1,4)	5	4/36	20/36
(5,1) (4,2) (3,3) (2,4) (1,5)	6	5/36	30/36
(6,1) (5,2) (4,3) (3,4) (2,5) (1,6)	7	6/36	42/36
(6,2) (5,3) (4,4) (3,5) (2,6)	8	5/36	40/36
(6,3) (5,4) (4,5) (3,6)	9	4/36	36/36
(6,4) (5,5) (4,6)	10	3/36	30/36
(6,5) (5,6)	11	2/36	22/36
(6,6)	12	1/36	12/36
			$252/36 = 7 = E(X)$

Valor esperado de una función

- Sea *g(X)* una función de una variable aleatoria *X* y que, por tanto, tomará también valores aleatorios.
- Se define el valor esperado de *g(X)* como:

$$\mu_{g(X)} \equiv E(g(X)) = \sum_i g(x_i) f(x_i)$$

$$\text{(en el caso continuo)} \quad = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Linealidad

- El valor esperado cumple las propiedades de **linealidad**.
- Sea $g(X) = a\,X + b$, con a y b constantes.

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

Varianza

- La **varianza** de una variable aleatoria es una medida de dispersión, representada por σ^2 o *Var(X)*.

$$\sigma^2 \equiv Var(X) = E\left((X-\mu)^2\right) = \sum_i (x_i-\mu)^2 f(x_i)$$

(en el caso continuo) $$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

- Propiedad: $$\sigma^2 = E(X^2) - \mu^2$$

Varianza

- La varianza de una función de una variable aleatoria se define como:

$$\sigma^2_{g(X)} = E\left((g(X) - \mu_{g(X)})^2\right) = \sum_i \left(g(x_i) - \mu_{g(X)}\right)^2 f(x_i)$$

(en el caso continuo) $$= \int_{-\infty}^{\infty} \left(g(x) - \mu_{g(X)}\right)^2 f(x)dx$$

Varianza

- Sea $g(X) = aX + b$, su varianza es:

$$\sigma^2_{aX+b} = E\left((aX+b-\mu_{aX+b})^2\right) = E\left((aX+b-a\mu_X-b)^2\right) =$$
$$= E\left((aX-a\mu_X)^2\right) = E\left(a^2(X-\mu_X)^2\right) = a^2E\left((X-\mu_X)^2\right) = a^2\sigma^2_X$$

- De aquí se deduce que la varianza de una función aleatoria constante es cero:

$$\sigma^2_b = 0$$

Ejemplo

- X = suma de los resultados de 2 dados.

Resultados	x_i	$f(x_i)$	x_i^2	$x_i^2 f(x_i)$
(1,1)	2	1/36	4	4/36
(2,1) (1,2)	3	2/36	9	18/36
(3,1) (2,2) (1,3)	4	3/36	16	48/36
(4,1) (3,2) (2,3) (1,4)	5	4/36	25	100/36
(5,1) (4,2) (3,3) (2,4) (1,5)	6	5/36	36	180/36
(6,1) (5,2) (4,3) (3,4) (2,5) (1,6)	7	6/36	49	294/36
(6,2) (5,3) (4,4) (3,5) (2,6)	8	5/36	64	320/36
(6,3) (5,4) (4,5) (3,6)	9	4/36	81	324/36
(6,4) (5,5) (4,6)	10	3/36	100	300/36
(6,5) (5,6)	11	2/36	121	242/36
(6,6)	12	1/36	144	144/36

Desviación típica

- La **desviación típica** de una variable aleatoria se define como la raíz cuadrada positiva de la varianza.
- Caso discreto:

$$\sigma = \sqrt{\sum_i (x_i - \mu)^2 f(x_i)}$$

- Caso continuo:

$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx}$$

Ejemplo

- Luego la varianza de *X* es:

$$\sigma^2 = \sum_i x_i^2 f(x_i) - \mu^2 = \frac{1974}{36} - 7^2 = 5{,}83$$

- La desviación típica de *X* es:

$$\sigma = 2{,}42$$

Variables aleatorias bidimensionales

- Denotaremos una **variable aleatoria bidimensional** de un experimento aleatorio por (X, Y), de forma que tomará valores (x, y) en un espacio bidimensional real.
- Una variable bidimensional es **discreta** cuando las dos variables que la componen lo sean. Asimismo será **continua** cuando tanto X como Y sean continuas.

Variables aleatorias bidimensionales

- Sea una variable aleatoria bidimensional (X, Y) **discreta** (con k valores diferentes de X y l valores diferentes de Y).
- Su **función de probabilidad conjunta** es:

$$f(x, y) = P(X = x, Y = y)$$

- Propiedades:

$$\sum_{i=1}^{k} \sum_{j=1}^{l} f(x_i, y_j) = 1$$

$$f(x, y) \geq 0,\ \ \forall x \in \{x_1, \ldots, x_k\}, \forall y \in \{y_1, \ldots, y_l\}$$

Variables aleatorias bidimensionales

- Sea una variable aleatoria bidimensional (*X*, *Y*) **continua**.
- Su **función de densidad conjunta** es la función *f*(*x*,*y*) tal que:

$$P(x_1 < x < x_2, y_1 < y < y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x, y) dxdy$$

- Propiedades:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$$

$$f(x, y) \geq 0$$

Variables aleatorias bidimensionales: Media

- Las **medias** o **esperanzas matemáticas** de cada una de las dos componentes de una variable aleatoria bidimensional serían:

Caso discreto:

$$\mu_X = E(X) = \sum_{i=1}^{k}\sum_{j=1}^{l} x_i f(x_i, y_i)$$

$$\mu_Y = E(Y) = \sum_{i=1}^{k}\sum_{j=1}^{l} y_j f(x_i, y_i)$$

Caso continuo:

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\ f(x, y)dxdy$$

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\ f(x, y)dxdy$$

30

Variables aleatorias bidimensionales: Varianza

- Las **varianzas** de cada una de las dos componentes de una variable aleatoria bidimensional serían:

Caso discreto:

$$\sigma_X^2 = Var(X) = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_i - \mu_X)^2 f(x_i, y_i)$$

$$\sigma_Y^2 = Var(Y) = \sum_{i=1}^{k}\sum_{j=1}^{l}(y_j - \mu_Y)^2 f(x_i, y_i)$$

Caso continuo:

$$\sigma_X^2 = Var(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_X)^2 f(x, y)dxdy$$

$$\sigma_Y^2 = Var(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - \mu_Y)^2 f(x, y)dxdy$$

Variables aleatorias bidimensionales: Covarianza

- La **covarianza** entre las dos componentes de una variable aleatoria bidimensional serían:

$$\sigma_{XY}^2 = Cov(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right)$$

Caso discreto:

$$\sigma_{XY}^2 = Cov(X,Y) = \sum_{i=1}^{k}\sum_{j=1}^{l}(x_i-\mu_X)(y_j-\mu_Y)f(x_i,y_i)$$

Caso continuo:

$$\sigma_{XY}^2 = Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy$$

32

Variables aleatorias bidimensionales: Covarianza

- Se puede demostrar (¡ejercicio!) que la covarianza se puede reescribir como:

$$\sigma_{XY}^2 = E(XY) - \mu_X \mu_Y = \mu_{XY} - \mu_X \mu_Y$$

- Si las dos variables *X* e *Y* son **independientes** entre sí, se puede demostrar (¡ejercicio!) que su covarianza es nula:

$$E(XY) = \mu_{XY} = \mu_X \mu_Y \;\Rightarrow\; \sigma_{XY}^2 = 0$$