

# Probabilidades y Estadística (C)

Franco Frizzo

(Basado en el apunte de Ana M. Bianco y Elena J. Martínez)

## Índice

<b>1. Conceptos básicos de probabilidad</b>	<b>3</b>
1.1. Probabilidad condicional . . . . .	4
1.2. Independencia de eventos . . . . .	5
<b>2. Variables aleatorias</b>	<b>6</b>
2.1. Variables aleatorias discretas . . . . .	6
2.1.1. Esperanza de una v.a. discreta . . . . .	8
2.1.2. Varianza de una v.a. discreta . . . . .	8
2.2. Variables aleatorias continuas . . . . .	9
2.2.1. Esperanza de una v.a. continua . . . . .	10
2.2.2. Varianza de una v.a. continua . . . . .	11
<b>3. Distribuciones de variables aleatorias</b>	<b>11</b>
3.1. Distribuciones de variables aleatorias discretas . . . . .	11
3.1.1. Distribución binomial . . . . .	11
3.1.2. Distribución geométrica . . . . .	12
3.1.3. Distribución binomial negativa . . . . .	12
3.1.4. Distribución hipergeométrica . . . . .	13
3.1.5. Distribución Poisson . . . . .	13
3.2. Distribuciones de variables aleatorias continuas . . . . .	15
3.2.1. Distribución uniforme . . . . .	15
3.2.2. Distribución normal . . . . .	15
3.2.3. Distribución gamma . . . . .	16
3.2.4. Distribución exponencial . . . . .	17
3.3. Funciones generadoras de momentos . . . . .	17
3.3.1. Funciones generadoras de momentos de algunas distribuciones . . . . .	19
3.4. Generación de números aleatorios . . . . .	19
<b>4. Vectores aleatorios</b>	<b>20</b>
4.1. Vectores aleatorios en dos dimensiones . . . . .	20
4.1.1. Independencia de variables aleatorias . . . . .	21
4.1.2. Esperanza de una función de dos variables aleatorias . . . . .	21
4.1.3. Covarianza y correlación . . . . .	22
4.2. Extensión a más de dos dimensiones . . . . .	24
4.2.1. Vectores aleatorios discretos . . . . .	24
4.2.2. Vectores aleatorios continuos . . . . .	25
4.2.3. Independencia de variables aleatorias . . . . .	26
4.2.4. Distribución de la suma de variables aleatorias independientes . . . . .	26

4.2.5. Sumas y promedios de variables aleatorias . . . . .	27
4.3. Desigualdad de Chebyshev . . . . .	28
4.4. Ley de los Grandes Números . . . . .	29
4.5. Teorema Central del Límite . . . . .	30
<b>5. Inferencia estadística</b>	<b>32</b>
5.1. Conceptos básicos de una investigación estadística . . . . .	32
5.1.1. Medidas de resumen . . . . .	32
5.2. Estimación puntual . . . . .	33
5.2.1. Método de estimación por momentos . . . . .	33
5.2.2. Método de estimación de máxima verosimilitud . . . . .	34
5.2.3. Propiedades de los estimadores . . . . .	34
5.3. Intervalos de confianza . . . . .	35
5.3.1. Intervalos de confianza para los parámetros de una distribución normal . . . . .	35
5.3.2. Método general para obtener intervalos de confianza . . . . .	37
5.3.3. Intervalos de confianza de nivel asintótico . . . . .	37
5.4. Tests de hipótesis . . . . .	39
5.4.1. Test de hipótesis para los parámetros de una distribución normal . . . . .	40
5.4.2. Tests de hipótesis de nivel aproximado . . . . .	41
5.4.3. Relación entre tests de hipótesis bilaterales e intervalos de confianza . . . . .	42
<b>6. Cadenas de Márkov</b>	<b>42</b>

# 1. Conceptos básicos de probabilidad

**Definición.** Dado un experimento aleatorio (cualquier proceso o acción que genera observaciones y que puede ser repetible), su **espacio muestral** asociado es el conjunto de todos sus resultados posibles. Lo notaremos  $\mathcal{S}$ .

**Definición.** Se denomina **suceso** o **evento** a cualquier subconjunto del espacio muestral. Si  $\mathcal{S}$  es finito o infinito numerable, cualquier subconjunto es un evento. Si  $\mathcal{S}$  es infinito, “casi todo” subconjunto de  $\mathcal{S}$  es un evento.

Un evento es *elemental* o *simple* si consiste de un único resultado individual y *compuesto* si consiste de más de un evento elemental.

**Definición.** Dado un experimento aleatorio y un espacio muestral asociado  $\mathcal{S}$ , a cada evento  $A$  se le asociará un número que notaremos  $P(A)$  y que llamaremos **probabilidad** del evento  $A$ . Esta asignación debe satisfacer los siguientes axiomas:

A1.  $P(A) \geq 0$  para todo evento  $A$ .

A2.  $P(\mathcal{S}) = 1$ .

A3. a) Si  $A_1, A_2, \dots, A_n$  es una colección finita de sucesos mutuamente excluyentes, es decir que  $A_i \cap A_j = \emptyset \forall i \neq j$ , entonces:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

b) Si  $A_1, A_2, \dots$  es una colección infinita numerable de sucesos mutuamente excluyentes, es decir que  $A_i \cap A_j = \emptyset \forall i \neq j$ , entonces:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

## Propiedades.

1.  $P(A^C) = 1 - P(A)$ , para cualquier evento  $A$ .
2.  $P(\emptyset) = 0$ .
3. Si  $A \subseteq B \Rightarrow P(A) \leq P(B)$  y  $P(B - A) = P(B) - P(A)$ .
4. Dados dos sucesos cualesquiera  $A$  y  $B$ ,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .
5. Dados dos sucesos cualesquiera  $A$  y  $B$ ,  $P(A \cup B) \leq P(A) + P(B)$ .

*Demostración.*

1.  $1 = P(\mathcal{S}) \stackrel{A2}{=} P(A \cup A^C) \stackrel{A3a}{=} P(A) + P(A^C) \Rightarrow P(A^C) = 1 - P(A)$ .

2.  $P(\emptyset) = 1 - P(\emptyset^C) = 1 - P(\mathcal{S}) = 1 - 1 = 0$ .

3. Si  $A \subseteq B \Rightarrow B = A \cup (B - A)$  y estos dos eventos son excluyentes. Por el axioma A3a,  $P(B) = P(A) + P(B - A)$ . Dado que, por el axioma A1,  $P(B - A) \geq 0$ , resulta  $P(B) \geq P(A)$ , y despejando, se obtiene la segunda afirmación.

4.  $A \cup B = A \cup (B - A) = A \cup (B \cap A^C)$ , y estos dos eventos son excluyentes. Entonces, por el axioma A3a,

$$P(A \cup B) = P(A \cup (B \cap A^C)) = P(A) + P(B \cap A^C) \quad (1)$$

Por otra parte,  $B = (B \cap A) \cup (B \cap A^C)$ , y estos dos eventos son disjuntos, entonces

$$P(B) = P(B \cap A) + P(B \cap A^C) \Rightarrow P(B \cap A^C) = P(B) - P(B \cap A) \quad (2)$$

De (1) y (2) resulta que  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , como queríamos demostrar.

5. Se deduce inmediatamente de la propiedad anterior y del axioma A1.  $\square$

**Definición.** Sea un experimento aleatorio cuyo espacio muestral asociado  $\mathcal{S}$  es finito y sea  $n = \#\mathcal{S}$ , el cardinal del conjunto. Diremos que el espacio es de **equiprobabilidad** si los  $n$  sucesos elementales tienen igual probabilidad, es decir si

$$P(E_i) = p \quad \forall i$$

Como  $1 = P(\mathcal{S}) = \sum_{i=1}^n P(E_i) = \sum_{i=1}^n p = np \Rightarrow p = \frac{1}{n} = \frac{1}{\#\mathcal{S}}$ .

Dado cualquier suceso  $A$ ,  $P(A) = \sum_{E \in A} P(E) = \sum_{E \in A} \frac{1}{n} = \frac{\#A}{\#\mathcal{S}}$ .

### 1.1. Probabilidad condicional

**Definición.** Sean  $A$  y  $B$  eventos tales que  $P(B) > 0$ . La probabilidad del evento  $A$  **condicional** a la ocurrencia del evento  $B$  es

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

**Propiedades.**

1.  $P(A | B) \geq 0$  para todo suceso  $A$ .
2.  $P(\mathcal{S} | B) = 1$ .

*Demostración.*  $P(\mathcal{S} | B) = \frac{P(B \cap \mathcal{S})}{P(B)} = \frac{P(B)}{P(B)} = 1 \quad \square$

3. Regla del producto:  $P(A \cap B) = P(A | B)P(B)$ . Si además  $P(A) \neq 0$ , entonces  $P(A \cap B) = P(B | A)P(A)$ .

**Definición.** Una colección de eventos  $A_1, A_2, \dots, A_k$  constituye una **partición** del espacio muestral  $\mathcal{S}$  si

1.  $A_i \cap A_j = \emptyset \quad \forall i \neq j$ .
2.  $P(A_i) \geq 0 \quad \forall i$ .
3.  $\bigcup_{i=1}^k A_i = \mathcal{S}$ .

**Teorema (Probabilidad Total).** Sea  $A_1, A_2, \dots, A_k$  una partición del espacio muestral  $\mathcal{S}$  y sea  $B$  un suceso cualquiera, entonces

$$P(B) = \sum_{i=1}^k P(B | A_i)P(A_i)$$

*Demostración.*

$$B = B \cup \mathcal{S} = B \cup \left( \bigcup_{i=1}^k A_i \right) = \bigcup_{i=1}^k (B \cap A_i)$$

Como  $(B \cap A_i) \cap (B \cap A_j) = \emptyset \forall i \neq j$ , entonces:

$$P(B) = P\left(\bigcup_{i=1}^k (B \cap A_i)\right) = \sum_{i=1}^k P(B \cap A_i) = \sum_{i=1}^k P(B | A_i) P(A_i)$$

□

**Teorema (Bayes).** Sea  $A_1, A_2, \dots, A_k$  una partición del espacio muestral  $\mathcal{S}$  y sea  $B$  un suceso cualquiera tal que  $P(B) > 0$ , entonces

$$P(A_j | B) = \frac{P(B | A_j) P(A_j)}{\sum_{i=1}^k P(B | A_i) P(A_i)}$$

*Demostración.*

$$P(A_j | B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B | A_j) P(A_j)}{\sum_{i=1}^k P(B | A_i) P(A_i)}$$

En el numerador se aplicó la regla del producto, y en el denominador el Teorema de la Probabilidad Total. □

## 1.2. Independencia de eventos

**Definición.** Los eventos  $A$  y  $B$  son **independientes** si  $P(A \cap B) = P(A)P(B)$ . Si la igualdad no se cumple, decimos que los eventos son **dependientes**.

**Proposición.** Supongamos  $P(B) > 0$ ,  $A$  y  $B$  son independientes si y sólo si  $P(A | B) = P(A)$ .

*Demostración.*

( $\Rightarrow$ ) Si  $P(B) > 0 \Rightarrow P(A | B) = \frac{P(A \cap B)}{P(B)}$  está bien definida. Por ser  $A$  y  $B$  independientes,  $P(A \cap B) = P(A)P(B)$ , entonces

$$P(A | B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

( $\Leftarrow$ ) Aplicando la regla del producto, si  $P(B) > 0$ ,  $P(A \cap B) = P(A | B)P(B) = P(A)P(B)$ .

□

*Observación.* Si  $P(B) = 0$ , como  $A \cap B \subseteq B$ ,  $P(A \cap B) = 0$ , y por lo tanto la igualdad  $P(A \cap B) = P(A)P(B)$  siempre se satisface.

### Propiedades.

1. Si los sucesos  $A$  y  $B$  son excluyentes, es decir si  $A \cap B = \emptyset$  y si  $P(A) > 0$ ,  $P(B) > 0$ , entonces  $A$  y  $B$  no son independientes.

*Demostración.*  $0 = P(A \cap B) \neq P(A)P(B)$ . □

2. Si  $P(B) = 0$ , entonces  $B$  es independiente de cualquier suceso  $A$  tal que  $P(A) > 0$ .

*Demostración.* Como  $A \cap B \subseteq B$ ,  $P(A \cap B) = 0$ , y por lo tanto  $P(A \cap B) = P(A)P(B)$ , es decir,  $A$  y  $B$  son independientes. □

3. Si  $A \subseteq B$ ,  $P(A) > 0$  y  $P(B) < 1$ ,  $A$  y  $B$  no son independientes.

*Demostración.* Como  $A \subseteq B \Rightarrow A \cap B = A \Rightarrow P(A \cap B) = P(A) \neq P(A)P(B)$ . Luego,  $A$  y  $B$  no son independientes.  $\square$

4. Si  $A$  y  $B$  son sucesos independientes,  $A$  y  $B^C$  también lo son.

*Demostración.*  $P(A) = P(A \cap B) + P(A \cap B^C) \Rightarrow P(A \cap B^C) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^C)$ .  $\square$

**Definición.** Los eventos  $A_1, A_2, \dots, A_n$  son independientes si para todo  $k = 2, \dots, n$  y para todo conjunto de índices  $\{i_1, i_2, \dots, i_k\}$  tales que  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ , se verifica

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j})$$

Es decir, es necesario verificar  $\sum_{i=2}^n \binom{n}{i} = 2^n - n - 1$  condiciones.

*Observación.* Si los sucesos  $A_1, A_2, \dots, A_n$  son independientes entonces son independientes de a pares, pero la recíproca no es cierta.

## 2. Variables aleatorias

**Definición.** Sea  $\mathcal{S}$  un espacio muestral asociado con un experimento aleatorio. Una variable aleatoria  $X$  es una función que asocia a cada elemento  $w \in \mathcal{S}$  un número real  $X(w) = x$ , es decir

$$X : \mathcal{S} \rightarrow \mathbb{R}$$

Como se observa, en general representaremos a las v.a. con letras mayúsculas ( $X, Y, Z$ , etc.), y a sus valores con letras minúsculas, es decir,  $X(w) = x$  significa que  $x$  es el número real asociado al resultado  $w \in \mathcal{S}$  a través de  $X$ .

Indicaremos con  $R_X$  el rango de la v.a.  $X$ , es decir, el conjunto de sus valores posibles.

### 2.1. Variables aleatorias discretas

**Definición.** Una variable aleatoria es **discreta** si toma un número finito o infinito numerable de valores.

**Definición.** La **función de probabilidad puntual** o de **masa** de una v.a. discreta  $X$  se define para todo  $x \in \mathbb{R}$  como

$$p_X(x) = P(X = x) = P(\{w \in \mathcal{S} / X(w) = x\})$$

**Propiedades.**

1.  $p_x(x) \geq 0 \forall x$ .
2.  $\sum_{x \in R_X} p_x(x) = 1$ .

**Definición.** La **función de distribución acumulada** de una v.a. discreta  $X$ , con función de probabilidad puntual  $p_X(x)$ , se define para todo  $x \in \mathbb{R}$  como

$$F_X(x) = P(X \leq x) = \sum_{y \leq x, y \in R_X} p_X(y)$$

### Propiedades.

1.  $\forall x \in \mathbb{R}, F_X(x) \in [0, 1]$ .

*Demostración.*  $F_X(X) = P(X \leq x)$ , es decir,  $F_X(x)$  es una probabilidad, y toda probabilidad toma valores entre 0 y 1.  $\square$

2.  $F_X(x)$  es monótona no decreciente, es decir, si  $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$ .

*Demostración.* Consideremos el suceso

$$A = \{w/X(w) \leq x_2\} = \{w/X(w) \leq x_1\} \cup \{w/x_1 < X(w) \leq x_2\} = A_1 \cup A_2$$

Como  $A_1 \cap A_2 = \emptyset$ ,  $P(A) = P(A_1) + P(A_2)$ , es decir

$$P(X \leq x_2) = P(X \leq x_1) + P(x_1 < X \leq x_2) \geq P(X \leq x_1)$$

y, por lo tanto,

$$F_X(x_2) \geq F_X(x_1)$$

$\square$

3.  $F_X(x)$  es continua a derecha, es decir,  $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$ .
4.  $\lim_{x \rightarrow +\infty} F_X(x) = 1$  y  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .

*Demostración.*

- $\lim_{x \rightarrow +\infty} F_X(x) = \lim_{x \rightarrow +\infty} P(X \leq x) = \lim_{x \rightarrow +\infty} P(w/X(w) \leq x) = P(\mathcal{S}) = 1$ .
- $\lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} P(X \leq x) = \lim_{x \rightarrow -\infty} P(w/X(w) \leq x) = P(\emptyset) = 0$ .

$\square$

5. En cada punto  $x$ , el valor del salto es la probabilidad puntual, es decir

$$p_X(x) = F_X(x) - F_X(x^-)$$

donde  $x^- = \lim_{h \rightarrow 0^+} (x - h)$ .

*Demostración.*  $p_X(x) = P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - F_X(x^-)$ .  $\square$

**Proposición.** Sean  $a$  y  $b$  tales que  $a \leq b$ . Entonces

1.  $P(a < X \leq b) = F_X(b) - F_X(a)$
2.  $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$
3.  $P(a < X < b) = F_X(b^-) - F_X(a)$
4.  $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$

### 2.1.1. Esperanza de una v.a. discreta

**Definición.** Sea  $X$  una v.a. discreta que toma valores en  $R_X$ , con función de probabilidad puntual  $p_X(x)$ . La **esperanza** o **valor esperado** de  $X$  se define como

$$E(X) = \mu_X = \sum_{x \in R_X} x p_X(x)$$

siempre que  $\sum_{x \in R_X} |x| p_X(x) < \infty$ . Si la serie de los valores absolutos diverge, la esperanza no puede definirse y decimos que no existe.

**Proposición.** Si  $X$  es discreta y toma valores  $x_1, x_2, \dots$ , entonces  $h(X)$  es discreta con valores  $y_1, y_2, \dots$ , siendo  $y_j = h(x_i)$  para al menos un valor de  $i$ .

**Proposición.** Si la v.a.  $X$  tiene función de probabilidad puntual  $p_X(x)$  para todo  $x \in R_X$ , entonces la esperanza de cualquier función real  $h(X)$  está dada por

$$E(h(X)) = \sum_{x \in R_X} h(x) p_X(x)$$

si la serie es absolutamente convergente, es decir, si  $\sum_{x \in R_X} |h(x)| p_X(x) < \infty$ .

*Demostración.* Sea  $Y = h(X)$ . Entonces

$$E(Y) = \sum_j y_j p_Y(y_j) = \sum_j y_j \left[ \sum_{i|h(x_i)=y_j} p_X(x_i) \right] = \sum_j \sum_{i|h(x_i)=y_j} y_j p_X(x_i) = \sum_i h(x_i) p_X(x_i)$$

□

### Propiedades.

1. (*Linealidad*) Si  $a$  y  $b$  son constantes reales,  $E(aX + b) = aE(X) + b$ .

*Demostración.* Sea  $h(X) = aX + b$ . Entonces,

$$E(h(X)) = E(aX + b) = \sum_{x \in R_X} (ax + b) p_X(x) = a \sum_{x \in R_X} x p_X(x) + b \sum_{x \in R_X} p_X(x) = aE(X) + b$$

□

2. Si  $X$  es una v.a. tal que  $P(X = c) = 1$ , entonces  $E(X) = c$ .

*Demostración.*  $E(X) = c p_X(c) = c$ .

□

### 2.1.2. Varianza de una v.a. discreta

**Definición.** Sea  $X$  una v.a. discreta con función de probabilidad puntual  $p_X(x)$ . La **varianza** de  $X$ , que se denotará  $V(X)$ ,  $\sigma_X^2$  o  $\sigma^2$ , es

$$V(X) = \sigma_X^2 = \sum_{x \in R_X} (x - E(X))^2 p_X(x) = E[(X - E(X))^2]$$

y el **desvío estándar** de  $X$  es

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

**Proposición.**

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

*Demostración.*

$$\begin{aligned} V(X) &= E \left[ (X - E(X))^2 \right] = \sum_{x \in R_X} (x - E(X))^2 p_X(x) = \sum_{x \in R_X} (x^2 - 2E(X)x + (E(X))^2) p_X(x) = \\ &= \sum_{x \in R_X} x^2 p_X(x) - 2E(X) \sum_{x \in R_X} x p_X(x) + (E(X))^2 \sum_{x \in R_X} p_X(x) = \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2 \end{aligned}$$

□

### Propiedades.

1.  $V(aX + b) = a^2 V(X)$  y  $\sigma_{aX+b} = |a|\sigma_X$ .

*Demostración.* En general,  $V(h(X)) = \sum_{x \in R_X} (h(x) - E(h(X)))^2 p_X(x)$ . Entonces,

$$\begin{aligned} V(aX + b) &= \sum_{x \in R_X} (ax + b - E(aX + b))^2 p_X(x) = \sum_{x \in R_X} (ax + b - aE(X) - b)^2 p_X(x) \\ &= \sum_{x \in R_X} (ax - aE(X))^2 p_X(x) = a^2 \sum_{x \in R_X} (x - E(X))^2 p_X(x) = a^2 V(X) \end{aligned}$$

y por lo tanto,  $\sigma_{aX+b} = |a|\sigma_X$ .

En particular, observemos que  $\sigma_{aX}^2 = a^2 \sigma_X^2$  y  $\sigma_{X+b}^2 = \sigma_X^2$ , es decir, un cambio de escala afecta a la varianza, pero una traslación no la afecta. □

2. Si  $X$  es una v.a. tal que  $P(X = c) = 1$ , entonces  $V(X) = 0$ .

*Demostración.*  $V(X) = E(X^2) - (E(X))^2 = c^2 p_X(c) - (c p_X(c))^2 = c^2 - c^2 = 0$ . □

## 2.2. Variables aleatorias continuas

**Definición.** Una variable aleatoria  $X$  es **continua** si existe una función  $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$ , llamada *función de densidad* de la v.a.  $X$ , tal que

$$P(X \in A) = \int_A f_X(x) dx \quad \forall A \subseteq \mathbb{R}$$

En particular, si  $A = [a, b]$ , entonces

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

y  $P(X = a) = P(a \leq X \leq a) = 0 \quad \forall a \in \mathbb{R}$ .

**Propiedad.** Para que una función  $f(x)$  sea una función de densidad, debe satisfacer

- $f(x) \geq 0 \quad \forall x \in \mathbb{R}$ .
- $\int_{-\infty}^{+\infty} f(x) dx = 1$ .

**Definición.** La **función de distribución acumulada** de una v.a. continua  $X$  con función de densidad  $f_X(x)$  se define, para todo  $x \in \mathbb{R}$ , como

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt$$

**Propiedades.** Sea  $X$  una v.a. continua.

1.  $\forall x \in \mathbb{R}, F_X(x) \in [0, 1]$ .
2.  $F_X(x)$  es monótona no decreciente, es decir, si  $x_1 < x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$ .
3.  $F_X(x)$  es continua en todo punto.
4.  $\lim_{x \rightarrow +\infty} F_X(x) = 1$  y  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ .

**Proposición.** Sean  $a$  y  $b$  tales que  $a \leq b$ , entonces

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

*Demostración.* Se deduce inmediatamente del hecho de que, si  $X$  es continua,  $P(X = x) = 0$   $\square$

**Proposición.** Si  $X$  es una v.a. continua con función de densidad  $f_X(x)$  y función de distribución acumulada  $F_X(x)$ , entonces, en todo punto donde  $F_X(x)$  es derivable

$$F'(x) = \frac{\partial F_X(x)}{\partial x} = f_X(x)$$

*Demostración.* Resulta del Teorema Fundamental del Cálculo Integral y de la definición de  $F_X(X)$ .  $\square$

**Definición.** Sea  $X$  una v.a. continua con función de densidad  $f_X(x)$  y función de distribución acumulada  $F_X(x)$ , y sea  $0 < p < 1$ . El **percentil**  $(100p)$ -ésimo de la distribución de  $X$  es el valor  $x_p$ , tal que

$$F_X(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(t) dt = p$$

El 50-percentil de una distribución se denomina *mediana* de la distribución.

### 2.2.1. Esperanza de una v.a. continua

**Definición.** Sea  $X$  una v.a. continua con función de densidad  $f_X(x)$ . La **esperanza** o **valor esperado** de  $X$  se define como

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

siempre que  $\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty$ . Si esta integral no converge, la esperanza no puede definirse y decimos que no existe.

**Proposición.** Si la v.a. continua  $X$  tiene función de densidad  $f_X(x)$ , entonces la esperanza de cualquier función real  $h(X)$  está dada por

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

si  $\int_{-\infty}^{+\infty} |h(x)| f_X(x) dx < \infty$ .

**Propiedad** (Linealidad). Si  $a$  y  $b$  son constantes reales,  $E(aX + b) = aE(X) + b$ .

*Demostración.* Sea  $h(X) = aX + b$ , entonces

$$\begin{aligned} E(h(X)) &= \int_{-\infty}^{+\infty} h(x) f_X(x) dx = \int_{-\infty}^{+\infty} (ax + b) f_X(x) dx = \\ &= a \int_{-\infty}^{+\infty} x f_X(x) dx + b \int_{-\infty}^{+\infty} f_X(x) dx = aE(X) + b \end{aligned}$$

$\square$

### 2.2.2. Varianza de una v.a. continua

**Definición.** Sea  $X$  una v.a. continua con densidad  $f_X(x)$ . La **varianza** de  $X$ , que se denotará  $V(X)$ ,  $\sigma_X^2$  o  $\sigma^2$ , es

$$V(X) = \sigma_X^2 = E[(X - E(X))^2] = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx$$

y el **desvío estándar** de  $X$  es  $\sigma_X = \sqrt{V(X)}$ .

**Proposición.**

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

*Demostración.*

$$\begin{aligned} V(X) &= E[(X - E(X))^2] = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx = \\ &= \int_{-\infty}^{+\infty} (x^2 - 2xE(X) + (E(X))^2) f_X(x) dx = \\ &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - 2E(X) \int_{-\infty}^{+\infty} x f_X(x) dx + (E(X))^2 \int_{-\infty}^{+\infty} f_X(x) dx = \\ &= E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2 \end{aligned}$$

□

**Propiedad.** Sea  $X$  una v.a. continua con densidad  $f_X(x)$ , entonces

$$V(aX + b) = a^2 V(X) \quad \text{y} \quad \sigma_{aX+b} = |a|\sigma_X$$

*Demostración.* En general,  $V(h(X)) = \int_{-\infty}^{+\infty} (h(X) - E(h(X)))^2 f_X(x) dx$ . Entonces,

$$\begin{aligned} V(aX + b) &= \int_{-\infty}^{+\infty} (ax + b - E(aX + b))^2 f_X(x) dx = \int_{-\infty}^{+\infty} (ax + b - aE(X) - b)^2 f_X(x) dx \\ &= \int_{-\infty}^{+\infty} (ax - aE(X))^2 f_X(x) dx = a^2 \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx = a^2 V(X) \end{aligned}$$

y por lo tanto,  $\sigma_{aX+b} = |a|\sigma_X$ .

□

## 3. Distribuciones de variables aleatorias

### 3.1. Distribuciones de variables aleatorias discretas

#### 3.1.1. Distribución binomial

Un *experimento binomial* es un experimento aleatorio que satisface las siguientes condiciones

- Consta de  $n$  repeticiones, siendo  $n$  fijo.
- Las repeticiones son idénticas, y en cada una de ellas hay solo dos resultados posibles, que denominaremos éxito (E) y fracaso (F) (a un experimento de este tipo se lo denomina *ensayo de Bernoulli*).
- Las repeticiones son independientes, es decir, el resultado de una no influye sobre el de las demás.
- La probabilidad de éxito  $p = P(E)$  se mantiene constante durante las repeticiones.

**Definición.** Dado un experimento binomial que conste de  $n$  repeticiones, con  $P(E) = p$ , consideramos la v.a.

$X$  : número de éxitos en las  $n$  repeticiones

Decimos que  $X$  tiene **distribución binomial** con parámetros  $n$  y  $p$ , y lo notamos  $X \sim \text{Bi}(n, p)$ .

**Función de probabilidad puntual.** Si  $X \sim \text{Bi}(n, p)$ , entonces

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \forall x \in 0, 1, \dots, n$$

**Función de distribución acumulada.** Si  $X \sim \text{Bi}(n, p)$ , entonces

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k} & \text{si } 0 \leq x \leq n \\ 1 & \text{si } x > n \end{cases}$$

donde  $\lfloor x \rfloor$  denota la parte entera de  $x$ .

**Esperanza y varianza.** Sea  $X \sim \text{Bi}(n, p)$ .

$$E(X) = np \quad \text{y} \quad V(X) = np(1-p)$$

### 3.1.2. Distribución geométrica

**Definición.** Supongamos que se realizan repeticiones independientes de un ensayo de Bernoulli con probabilidad de éxito  $P(E) = p$  constante. Definimos la v.a.

$X$  : número de repeticiones hasta obtener el primer éxito

Decimos que  $X$  tiene **distribución geométrica** con parámetro  $p$ , y lo notamos  $X \sim G(p)$ .

**Función de probabilidad puntual.** Si  $X \sim G(p)$ , entonces

$$p_X(x) = (1-p)^{x-1} \quad \forall x \in \mathbb{N}$$

**Función de distribución acumulada.** Si  $X \sim G(p)$ , entonces

$$F_X(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - (1-p)^{\lfloor x \rfloor} & \text{si } x \geq 1 \end{cases}$$

donde  $\lfloor x \rfloor$  denota la parte entera de  $x$ .

**Esperanza y varianza.** Sea  $X \sim G(p)$ .

$$E(X) = \frac{1}{p} \quad \text{y} \quad V(X) = \frac{1-p}{p^2}$$

**Proposición** (Propiedad de falta de memoria). Sea  $X \sim G(p)$ , y sean  $n$  y  $m$  números naturales cualesquiera.

$$P(X > n + m \mid X > n) = P(X > m)$$

### 3.1.3. Distribución binomial negativa

**Definición.** Supongamos que se realizan repeticiones independientes de un ensayo de Bernoulli con probabilidad de éxito  $P(E) = p$  constante. Sea  $r \geq 1$ . Definimos la v.a.

$X$  : número de repeticiones hasta obtener el  $r$ -ésimo éxito

Decimos que  $X$  tiene **distribución binomial negativa** con parámetros  $r$  y  $p$ , y lo notamos  $X \sim \text{BN}(r, p)$ .

**Función de probabilidad puntual.** Si  $X \sim \text{BN}(r, p)$ , entonces

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \forall x \in \{r, r+1, r+2, \dots\}$$

**Función de distribución acumulada.** Si  $X \sim \text{BN}(r, p)$ , entonces

$$F_X(x) = \begin{cases} 0 & \text{si } x < r \\ \sum_{k=r}^{\lfloor x \rfloor} \binom{k-1}{r-1} p^r (1-p)^{k-r} & \text{si } x \geq r \end{cases}$$

donde  $\lfloor x \rfloor$  denota la parte entera de  $x$ .

**Esperanza y varianza.** Sea  $X \sim \text{BN}(r, p)$ .

$$E(X) = \frac{r}{p} \quad \text{y} \quad V(X) = \frac{r(1-p)}{p^2}$$

### 3.1.4. Distribución hipergeométrica

**Definición.** Consideremos una población de  $N$  elementos, en la que cada elemento puede ser clasificado como éxito o fracaso, y entre los cuales hay  $D$  éxitos. Supongamos que se extrae una muestra de  $n$  elementos de la población, de forma tal que cualquier subconjunto de  $n$  elementos tiene igual probabilidad de ser elegido. Definimos la v.a.

$X$  : número de éxitos en la muestra de tamaño  $n$

Decimos que  $X$  tiene **distribución hipergeométrica** con parámetros  $n$ ,  $N$  y  $D$ , y lo notamos  $X \sim \text{H}(n, N, D)$ .

**Función de probabilidad puntual.** Si  $X \sim \text{H}(n, N, D)$ , entonces

$$p_X(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad \forall \text{máx}(0, n - (N - D)) \leq x \leq \text{mín}(n, D)$$

**Esperanza y varianza.** Sea  $X \sim \text{H}(n, N, D)$ .

$$E(X) = n \frac{D}{N} \quad \text{y} \quad V(X) = \left( \frac{N-n}{N-1} \right) n \frac{D}{N} \left( 1 - \frac{D}{N} \right)$$

*Observación.* Si  $n$  es pequeño en relación a  $N$ , la distribución hipergeométrica puede ser aproximada por una distribución binomial de parámetros  $n$  y  $p = \frac{D}{N}$ . Observemos que, en este caso, el factor  $\left( \frac{N-n}{N-1} \right)$  que aparece en la varianza (denominado *factor de corrección por población finita*) es aproximadamente 1.

### 3.1.5. Distribución Poisson

**Definición.** Decimos que la variable aleatoria  $X$  tiene **distribución de Poisson** con parámetro  $\lambda > 0$ , y lo notamos  $X \sim \mathcal{P}(\lambda)$ , si su función de probabilidad puntual está dada por

$$p_X(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \forall x \in \mathbb{N} \cup \{0\}$$

**Esperanza y varianza.** Sea  $X \sim \mathcal{P}(\lambda)$ .

$$E(X) = V(X) = \lambda$$

**Proposición** (Aproximación de la distribución binomial por la distribución Poisson). Sea  $X \sim \text{Bi}(n, p)$ , y supongamos que  $n \rightarrow \infty$  y  $p \rightarrow 0$  de forma tal que el producto  $np = \lambda$  se mantenga constante. Entonces

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \rightarrow \frac{e^{-\lambda} \lambda^x}{x!} \quad \forall x \in \mathbb{N} \cup \{0\}$$

Esto quiere decir que, para valores de  $n$  grandes y de  $p$  pequeños, la distribución binomial es bien aproximada por una distribución Poisson de parámetro  $\lambda = np$ .

Algunos autores sugieren que la aproximación es buena para  $n \geq 100$ ,  $p \leq 0,01$  y  $\lambda \leq 20$ .

*Demostración.*

$$\begin{aligned} p_X(x) &= \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \\ &= \frac{n(n-1)\dots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \frac{\lambda^x}{x!} = \\ &= \left[ \frac{n}{n} \frac{n-1}{n} \dots \frac{n-x+1}{n} \right] \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \frac{\lambda^x}{x!} \end{aligned}$$

Observemos que:

- $\lim_{n \rightarrow \infty} 1 \frac{n-1}{n} \dots \frac{n-x+1}{n} = 1$
- $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$
- $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1$

Entonces,  $p_X(x) \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}$ , como queríamos demostrar. □

### Procesos de Poisson.

**Definición.** Supongamos que observamos la ocurrencia de un evento que se repite a lo largo del tiempo, y que existe una cantidad  $\theta > 0$ , tal que:

1. Dado un intervalo pequeño de longitud  $\Delta t$ , la probabilidad de que ocurra exactamente un evento es aproximadamente igual a  $\theta \Delta t$ , es decir,

$$P(\text{ocurra un evento en } \Delta t) = \theta \Delta t + o(\Delta t)$$

donde  $o(h)$  es una función  $g(h)$  tal que  $\lim_{h \rightarrow 0} \frac{g(h)}{h} = 0$ .

2. La probabilidad de que ocurra más de un evento en un intervalo pequeño de longitud  $\Delta t$  es pequeña en comparación con la probabilidad de que ocurra un evento, es decir,

$$P(\text{ocurra más de un evento en } \Delta t) = o(\Delta t)$$

3. El número de eventos que ocurren en un intervalo es independiente del número de eventos que ocurren en otro intervalo disjunto.

Un experimento que cumple con estas características se conoce como **proceso de Poisson**, con *tasa media de ocurrencia* o *intensidad*  $\theta$ . Se cumple que el número de ocurrencias del evento en un período de longitud  $t$  tiene distribución de Poisson de parámetro  $\theta t$ , es decir, la v.a.

$X_t$  : número de ocurrencias del evento en el intervalo de longitud  $t$

satisface  $X \sim \mathcal{P}(\theta t)$ .

## 3.2. Distribuciones de variables aleatorias continuas

### 3.2.1. Distribución uniforme

**Definición.** Sea  $X$  una v.a. continua. Decimos que  $X$  tiene **distribución uniforme** con parámetros  $A$  y  $B$ , y lo notamos  $X \sim \mathcal{U}(A, B)$ , si su función de densidad es

$$f_X(x) = \frac{1}{B-A} \mathbb{1}_{[A,B]}(x)$$

es decir, si su densidad es constante dentro del intervalo  $[A, B]$  y 0 fuera de él.

**Función de distribución acumulada.** Sea  $X \sim \mathcal{U}(A, B)$ , entonces

$$F_X(x) = \begin{cases} 0 & \text{si } x < A \\ \frac{x-A}{B-A} & \text{si } A \leq x \leq B \\ 1 & \text{si } x > B \end{cases}$$

**Esperanza y varianza.** Sea  $X \sim \mathcal{U}(A, B)$ , entonces

$$E(X) = \frac{A+B}{2} \quad \text{y} \quad V(X) = \frac{(B-A)^2}{12}$$

### 3.2.2. Distribución normal

**Definición.** Sea  $X$  una v.a. continua. Decimos que  $X$  tiene **distribución normal** con parámetros  $\mu$  y  $\sigma^2$  ( $\mu \in \mathbb{R}, \sigma > 0$ ), y lo notamos  $X \sim \mathcal{N}(\mu, \sigma^2)$ , si su función de densidad es

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

El gráfico de la función de densidad normal tiene forma de campana, con eje de simetría en  $x = \mu$  y puntos de inflexión en  $x = \mu \pm \sigma$ .

La importancia de la distribución normal radica no sólo en que frecuentemente en la práctica se hallan variables que tienen esta distribución (por ejemplo, los errores de medición) sino porque, bajo ciertas condiciones, suele ser una buena aproximación a la distribución de otras variables aleatorias.

**Esperanza y varianza.** Sea  $X \sim \mathcal{N}(\mu, \sigma^2)$ , entonces

$$E(X) = \mu \quad \text{y} \quad V(X) = \sigma^2$$

### Distribución normal estándar.

**Definición.** Sea  $Z$  una v.a. continua con distribución normal. Decimos que  $Z$  tiene **distribución normal estándar** si sus parámetros son  $\mu = 0$  y  $\sigma^2 = 1$ , es decir,  $Z \sim \mathcal{N}(0, 1)$ . Su función de densidad es

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

**Función de distribución acumulada.** Dada  $Z \sim N(0, 1)$ , su función de distribución acumulada se denota  $\Phi(z)$ , y está dada por

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Esta función no tiene una expresión analítica conocida, por lo que está tabulada.

**Propiedades.**

1. Si  $X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ .
2. Si  $Z \sim N(0, 1)$  y  $\sigma > 0 \Rightarrow X = \sigma Z + \mu \sim N(\mu, \sigma^2)$ .
3. Sean  $X \sim N(\mu, \sigma^2)$  y  $Z \sim N(0, 1)$ . Si denotamos  $x_p$  y  $z_p$  a los  $100p$ -ésimos percentiles de  $X$  y  $Z$  respectivamente, entonces  $x_p = \sigma z_p + \mu$ .

### 3.2.3. Distribución gamma

**Definición.** Dado  $\alpha > 0$ , definimos la **función gamma** o **función factorial** como

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

**Propiedades.**

1. Si  $\alpha > 1$ ,  $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ .
2. Si  $\alpha \in \mathbb{N}$ ,  $\Gamma\alpha = (\alpha - 1)!$ .
3.  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .

**Definición.** Sea  $X$  una v.a. continua. Decimos que  $X$  tiene **distribución gamma** con parámetros  $\alpha$  y  $\lambda$ , y lo notamos  $X \sim \Gamma(\alpha, \lambda)$ , si su función de densidad es

$$f_X(x) = \frac{e^{-\lambda x} x^{\alpha-1} \lambda^\alpha}{\Gamma(\alpha)} \mathbb{1}_{(0, +\infty)}(x)$$

**Esperanza y varianza.** Sea  $X \sim \Gamma(\alpha, \lambda)$ , entonces

$$E(X) = \frac{\alpha}{\lambda} \quad \text{y} \quad V(X) = \frac{\alpha}{\lambda^2}$$

**Definición.** Sea  $X$  una v.a. continua con distribución gamma. Decimos que  $X$  tiene **distribución gamma estándar** con parámetro  $\alpha$  si su parámetro  $\lambda = 1$ , es decir,  $X \sim \Gamma(\alpha, 1)$ . Su función de densidad está dada por

$$f_X(x) = \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} \mathbb{1}_{(0, +\infty)}(x)$$

Esta función de densidad es estrictamente decreciente si  $\alpha \leq 1$ , y si  $\alpha > 1$  alcanza un máximo y después decrece. La distribución gamma estándar está tabulada para diferentes valores de  $\alpha$ .

**Propiedad.** Si  $X \sim \Gamma(\alpha, \lambda)$  y  $a > 0$ ,  $aX \sim \Gamma(\alpha, \frac{\lambda}{a})$ .

Esta propiedad permite obtener probabilidades para una v. a. con distribución gamma a partir de una distribución gamma estándar. Supongamos que  $X \sim \Gamma(\alpha, \lambda)$ , entonces  $\lambda X \sim \Gamma(\alpha, 1)$  y, por ejemplo,

$$P(X \leq x) = P(\lambda X \leq \lambda x) = F_{\lambda x}(\lambda x)$$

### 3.2.4. Distribución exponencial

**Definición.** Sea  $X$  una v.a. continua. Decimos que  $X$  tiene **distribución exponencial** con parámetro  $\lambda$ , y lo notamos  $X \sim \mathcal{E}(\lambda)$ , si su función de densidad es

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{(0,+\infty)}(x)$$

*Observación.* Notemos que una distribución exponencial es un caso particular de la distribución gamma con parámetro  $\alpha = 1$ .

**Función de distribución acumulada.** Sea  $X \sim \mathcal{E}(\lambda)$ , entonces

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

**Esperanza y varianza.** Sea  $X \sim \mathcal{E}(\lambda)$ , entonces

$$E(X) = \frac{1}{\lambda} \quad \text{y} \quad V(X) = \frac{1}{\lambda^2}$$

**Proposición** (Propiedad de falta de memoria). Sea  $X \sim \mathcal{E}(\lambda)$ , y sean  $s$  y  $t$  números reales positivos cualesquiera.

$$P(X > s + t \mid X > s) = P(X > t)$$

**Proposición** (Relación de la distribución exponencial con los procesos de Poisson). Dado un proceso de Poisson de intensidad  $v$ , si se define la v.a.

$T$  : tiempo hasta la ocurrencia del primer evento

entonces  $T \sim \mathcal{E}(v)$ .

*Demostración.* Si  $t \leq 0$ ,  $F_T(t) = 0$ . Sea  $t > 0$ :

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P(X_t = 0)$$

dado que el tiempo hasta la primera ocurrencia es mayor que  $t$  si y solo si no ha ocurrido ningún evento en el intervalo  $(0, t)$ . Entonces,

$$F_T(t) = 1 - P(X_t = 0) = 1 - \frac{e^{-vt}(vt)^0}{0!} = 1 - e^{-vt}$$

y, por lo tanto,

$$F_T(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-vx} & \text{si } x > 0 \end{cases}$$

es decir,  $T \sim \mathcal{E}(v)$ . □

### 3.3. Funciones generadoras de momentos

**Definición.** Sea  $X$  una variable aleatoria. El **momento** de orden  $k$  de  $X$  se define como  $E(X^k)$ , siempre que la esperanza exista.

**Definición.** La **función generadora de momentos** de una variable aleatoria  $X$  es una función a valores reales  $M_X(t)$ , definida como

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x \in R_X} e^{tx} p_X(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx & \text{si } X \text{ es continua} \end{cases}$$

siempre que la esperanza exista para todo  $t \in (-h, h)$ ,  $h > 0$ . Esta última es una condición técnica necesaria para que  $M_X(t)$  sea diferenciable en  $t = 0$ .

Enunciaremos, sin demostración, el siguiente lema de cálculo avanzado, a partir del cual puede demostrarse un importante teorema acerca de la función generadora de momentos de una variable aleatoria.

**Lema.** Si la función  $g(t)$ , definida por

$$g(t) = \sum_x e^{tx} p(x) \quad \text{o} \quad g(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

converge para todo  $t \in (-h, h)$  para algún  $h > 0$ , entonces existen las derivadas de orden  $n$  de  $g(t)$  para todo  $t \in (-h, h)$  y para todo  $n \in \mathbb{N}$ , y se obtienen como

$$\frac{\partial^n g(t)}{\partial t^n} = \sum_x \frac{\partial^n e^{tx}}{\partial t^n} p(x) \quad \text{o} \quad \frac{\partial^n g(t)}{\partial t^n} = \int_{-\infty}^{+\infty} \frac{\partial^n e^{tx}}{\partial t^n} f(x) dx$$

**Teorema.** Sea  $X$  una v.a. para la cual existe la función generadora de momentos  $M_X(t)$ . Entonces

$$E(X^n) = \left. \frac{\partial^n M_X(t)}{\partial t^n} \right|_{t=0}$$

*Demostración.* Si la función generadora de momentos existe para todo  $t \in (-h, h)$  para algún  $h > 0$ , aplicando el lema anterior

$$\frac{\partial^n M_X(t)}{\partial t^n} = \sum_x \frac{\partial^n e^{tx}}{\partial t^n} p(x) \quad \text{o} \quad \frac{\partial^n M_X(t)}{\partial t^n} = \int_{-\infty}^{+\infty} \frac{\partial^n e^{tx}}{\partial t^n} f(x) dx$$

$$\frac{\partial^n M_X(t)}{\partial t^n} = \sum_x x^n e^{tx} p(x) \quad \text{o} \quad \frac{\partial^n M_X(t)}{\partial t^n} = \int_{-\infty}^{+\infty} x^n e^{tx} f(x) dx$$

Evaluando estas derivadas en 0,

$$\left. \frac{\partial^n M_X(t)}{\partial t^n} \right|_{t=0} = \sum_x x^n p(x) = E(X^n) \quad \text{o} \quad \left. \frac{\partial^n M_X(t)}{\partial t^n} \right|_{t=0} = \int_{-\infty}^{+\infty} x^n f(x) dx = E(X^n)$$

□

**Propiedad.** Sea  $X$  una v.a. con función generadora de momentos  $M_X(t)$ . Entonces, si  $Y = aX + b$ ,  $M_Y(t) = e^{bt} M_X(at)$ .

**Teorema** (Unicidad de la función generadora de momentos). Si existe la función generadora de momentos de una variable aleatoria, es única. Además, la función generadora de momentos determina a la función de probabilidad puntual o densidad de la v.a., salvo a lo sumo en un conjunto de probabilidad 0.

### 3.3.1. Funciones generadoras de momentos de algunas distribuciones

Distribución	$M_X(t)$
$\text{Bi}(n, p)$	$(e^t p + 1 - p)^n$
$G(p)$	$\frac{pe^t}{1 - (1 - p)e^t}$
$\text{BN}(r, p)$	$\left(\frac{pe^t}{1 - (1 - p)e^t}\right)^r$
$\mathcal{P}(\lambda)$	$e\lambda(e^t - 1)$
$\mathcal{U}(A, B)$	$\frac{e^{tb} - e^{ta}}{t(b - a)}$
$N(\mu, \sigma^2)$	$e^{\frac{\sigma^2 t^2}{2} + \mu t}$
$\Gamma(\alpha, \lambda)$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$
$\mathcal{E}(\lambda)$	$\frac{\lambda}{\lambda - t}$

### 3.4. Generación de números aleatorios

En diversas aplicaciones, es útil disponer de números elegidos al azar. Existen algoritmos que permiten generar números aleatorios que se comportan como si proviniesen de una distribución  $\mathcal{U}(0, 1)$ . El siguiente teorema nos permite afirmar que esto es suficiente para obtener números aleatorios que respeten alguna otra distribución determinada.

**Teorema.** Sean  $U$  una variable aleatoria tal que  $U \sim \mathcal{U}(0, 1)$ , y  $G$  una función de distribución acumulada continua y estrictamente creciente. Si  $X = G^{-1}(U)$ , entonces la función de distribución acumulada de  $X$  es  $G$ , es decir  $F_X = G$ .

*Demostración.* Recordemos que si  $U \sim \mathcal{U}(0, 1)$ , entonces su función de distribución es de la forma

$$F_U(u) = \begin{cases} 0 & \text{si } u < 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u > 1 \end{cases}$$

Por lo tanto, como  $G$  es una función estrictamente creciente y su imagen pertenece al intervalo  $(0, 1)$ , entonces

$$F_X(x) = P(X \leq x) = P(G^{-1}(U) \leq x) = P(U \leq G(x)) = F_U(G(x)) = G(x) \quad \square$$

Si la distribución  $G$  tiene saltos o es constante de a trozos, no existirá su inversa. Sin embargo, se puede demostrar que existe una función  $H$  con las propiedades requeridas en el teorema anterior, lo que permite enunciar siguiente resultado.

**Teorema.** Sean  $U$  una variable aleatoria tal que  $U \sim \mathcal{U}(0, 1)$ , y  $G$  una función de distribución acumulada. Existe una función  $H$  tal que  $H(U)$  tiene distribución acumulada  $G$ .

## 4. Vectores aleatorios

### 4.1. Vectores aleatorios en dos dimensiones

**Definición.** Sean  $X$  e  $Y$  v.a. discretas definidas sobre un espacio muestral  $\mathcal{S}$ . La **función de probabilidad conjunta** del par  $(X, Y)$  se define como

$$p_{XY}(x, y) = P(X = x, Y = y)$$

El conjunto  $R_{XY} = \{(x, y)/x \in R_X, y \in R_Y\}$  es el **rango** o **recorrido** del vector aleatorio  $(X, Y)$ .

Dado cualquier conjunto  $A \subseteq \mathbb{R}^2$ ,

$$P((X, Y) \in A) = \sum_{(x,y) \in A} p_{XY}(x, y)$$

Una función de probabilidad conjunta satisface

- $p_{XY}(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$
- $\sum_x \sum_y p_{XY}(x, y) = 1$

**Definición.** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad conjunta  $p_{XY}(x, y)$ . Las **funciones de probabilidad marginal** de  $X$  e  $Y$  están dadas por

$$p_X(x) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \sum_x p_{XY}(x, y)$$

**Definición.** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad conjunta  $p_{XY}(x, y)$ . La **función de distribución acumulada conjunta** de  $(X, Y)$  está dada por

$$F_{XY}(x, y) = \sum_{s \leq x} \sum_{t \leq y} p_{XY}(s, t) \quad \forall (x, y) \in \mathbb{R}^2$$

**Definición.** Sea  $(X, Y)$  un vector aleatorio discreto con función de probabilidad conjunta  $p_{XY}(x, y)$  y marginales  $p_X(x)$  y  $p_Y(y)$ , y sea  $x$  tal que  $p_X(x) > 0$ . La **función de probabilidad condicional** de  $Y$  dado  $X = x$  está dada por

$$p_{Y|X=x}(y) = \frac{p_{XY}(x, y)}{p_X(x)}$$

Análogamente, si  $y$  es tal que  $p_Y(y) > 0$ , la función de probabilidad condicional de  $X$  dado  $Y = y$  está dada por

$$p_{X|Y=y}(x) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

**Definición.** Sean  $X$  e  $Y$  variables aleatorias continuas definidas sobre un espacio muestral  $\mathcal{S}$ . El vector aleatorio  $(X, Y)$  es continuo si existe una función, denominada **función de densidad conjunta**,  $f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ , tal que

$$P((X, Y) \in A) = \iint_A f_{XY}(x, y) dx dy \quad \forall A \subseteq \mathbb{R}^2$$

En particular, si  $A = [a, b] \times [c, d]$ ,

$$P((X, Y) \in A) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$$

Una función de densidad conjunta satisface

- $f_{XY}(x, y) \geq 0 \forall (x, y) \in \mathbb{R}^2$ .
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1$ .

**Definición.** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad conjunta  $f_{XY}(x, y)$ . La **función de distribución acumulada conjunta** de  $(X, Y)$  está dada por

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) dt ds \quad \forall (x, y) \in \mathbb{R}^2$$

**Definición.** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad conjunta  $f_{XY}(x, y)$ . Las **funciones de densidad marginal** de  $X$  e  $Y$  están dadas por

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dx$$

**Definición.** Sea  $(X, Y)$  un vector aleatorio continuo con función de densidad conjunta  $f_{XY}(x, y)$  y marginales  $f_X(x)$  y  $f_Y(y)$ , y sea  $x$  tal que  $f_X(x) > 0$ . La **función de densidad condicional** de  $Y$  dado  $X = x$  está dada por

$$f_{Y|X=x}(y) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Análogamente, si  $y$  es tal que  $f_Y(y) > 0$ , la función de densidad condicional de  $X$  dado  $Y = y$  está dada por

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

#### 4.1.1. Independencia de variables aleatorias

**Definición.** Las variables aleatorias  $X$  e  $Y$  son **independientes** si y solo si para todo  $a < b$  y  $c < d$  se satisface

$$P(\{a < X < b\} \cap \{c < Y < d\}) = P(a < X < b) P(c < Y < d)$$

Si esta condición no se satisface, diremos que  $X$  e  $Y$  son dependientes.

En el caso de un vector  $(X, Y)$  discreto, la condición de independencia equivale a la siguiente:  $p_{XY}(x, y) = p_X(x)p_Y(y) \forall (x, y) \in \mathbb{R}^2$ . En el caso continuo,  $X$  e  $Y$  son independientes si y solo si  $f_{XY}(x, y) = f_X(x)f_Y(y) \forall (x, y) \in \mathbb{R}^2$ .

#### 4.1.2. Esperanza de una función de dos variables aleatorias

**Proposición.** Sean  $X$  e  $Y$  dos variables aleatorias discretas con función de probabilidad conjunta  $p_{XY}$  y sea  $h(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Entonces  $h(X, Y)$  es una variable aleatoria, y

$$E(h(X, Y)) = \sum_x \sum_y h(x, y)p_{XY}(x, y)$$

siempre que esta esperanza exista.

**Proposición.** Sean  $X$  e  $Y$  dos variables aleatorias continuas con función de densidad conjunta  $f_{XY}$  y sea  $h(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Entonces  $h(X, Y)$  es una variable aleatoria, y

$$E(h(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) f_{XY}(x, y) dx dy$$

siempre que esta esperanza exista.

**Proposición.** Sean  $X$  e  $Y$  dos variables aleatorias discretas o continuas, y sean  $a$  y  $b$  números reales. Entonces

$$E(aX + bY) = a E(X) + b E(Y)$$

*Demostración.* Haremos la demostración para el caso continuo (el caso discreto es similar). Sea  $h(X, Y) = aX + bY$ . Entonces

$$\begin{aligned} E(h(X, Y)) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) f_{XY}(x, y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (ax + by) f_{XY}(x, y) dx dy = \\ &= a \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{XY}(x, y) dx dy + b \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{XY}(x, y) dx dy = \\ &= a \int_{-\infty}^{+\infty} x \left( \int_{-\infty}^{+\infty} f_{XY}(x, y) dy \right) dx + b \int_{-\infty}^{+\infty} y \left( \int_{-\infty}^{+\infty} f_{XY}(x, y) dx \right) dy = \\ &= a \int_{-\infty}^{+\infty} x f_X(x) dx + b \int_{-\infty}^{+\infty} y f_Y(y) dy = a E(X) + b E(Y) \end{aligned}$$

□

**Proposición.** Si  $X$  e  $Y$  son dos variables aleatorias independientes, entonces

$$E(XY) = E(X) E(Y)$$

#### 4.1.3. Covarianza y correlación

**Definición.** Sean  $X$  e  $Y$  dos variables aleatorias. La **covarianza** entre  $X$  e  $Y$  se define como

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = \begin{cases} \sum_x \sum_y (x - E(X))(y - E(Y)) p_{XY}(x, y) \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))(y - E(Y)) f_{XY} dx dy \end{cases}$$

según sean  $X$  e  $Y$  discretas o continuas, respectivamente.

*Observación.*  $\text{Cov}(X, X) = V(X)$ .

**Proposición.**  $\text{Cov}(X, Y) = E(XY) - E(X) E(Y)$ .

*Demostración.*

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] = \sum_x \sum_y (x - E(X))(y - E(Y)) p_{XY}(x, y) = \\ &= \sum_x \sum_y (xy - x E(X) - y E(X) - E(X) E(Y)) p_{XY}(x, y) = \\ &= \sum_x \sum_y x y p_{XY}(x, y) - E(X) \sum_x \sum_y x p_{XY}(x, y) \\ &\quad - E(X) \sum_x \sum_y y p_{XY}(x, y) - E(X) E(Y) \sum_x \sum_y p_{XY}(x, y) = \\ &= E(XY) - E(Y) \sum_x x \sum_y p_{XY}(x, y) - E(X) \sum_y y \sum_x p_{XY}(x, y) + E(X) E(Y) = \\ &= E(XY) - E(Y) \sum_x x p_X(x) - E(X) \sum_y y p_Y(y) + E(X) E(Y) = \\ &= E(XY) - E(X) E(Y) - E(X) E(Y) + E(X) E(Y) = E(XY) - E(X) E(Y) \end{aligned}$$

□

**Propiedad.** Si  $X$  e  $Y$  son variables aleatorias independientes,  $\text{Cov}(X, Y) = 0$ . La recíproca no es cierta en general.

*Demostración.* Es inmediato a partir del hecho de que si  $X$  e  $Y$  son independientes,  $E(XY) = E(X)E(Y)$ .  $\square$

**Proposición.** Sean  $X$  e  $Y$  dos variables aleatorias discretas o continuas, y sean  $a$  y  $b$  números reales. Entonces

$$V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X, Y)$$

*Demostración.*

$$\begin{aligned} V(aX + bY) &= E\left(\left[(aX + bY) - E(aX + bY)\right]^2\right) = \\ &= E\left(\left[(aX + bY) - (aE(X) + bE(Y))\right]^2\right) = \\ &= E\left(\left[(aX - aE(X)) + (bY - bE(Y))\right]^2\right) = \\ &= E\left(\left[a(X - E(X))\right]^2\right) + E\left(\left[b(Y - E(Y))\right]^2\right) + 2E[ab(X - E(X))(Y - E(Y))] = \\ &= a^2 V(X) + b^2 V(Y) + 2ab \text{Cov}(X, Y) \end{aligned}$$

$\square$

**Definición.** Sean  $X$  e  $Y$  dos v.a. con varianza positiva. El **coeficiente de correlación** entre  $X$  e  $Y$  se define como

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

siendo  $\sigma_X$  y  $\sigma_Y$  los desvíos estándar de  $X$  e  $Y$ , respectivamente.

**Proposición.**

1. Sean  $a$ ,  $b$ ,  $c$  y  $d$  números reales,  $a \neq 0$ ,  $c \neq 0$ , y  $X$  e  $Y$  dos v.a. cualesquiera con varianza positiva. Entonces

$$\rho(aX + b, cY + d) = \text{sg}(ac)\rho(X, Y)$$

donde  $\text{sg}$  denota la función signo.

2.  $-1 \leq \rho(X, Y) \leq 1$ .

3.  $|\rho(X, Y)| = 1 \Leftrightarrow Y = aX + b$  con probabilidad 1, para ciertos valores reales  $a \neq 0$  y  $b$ . Observemos que el coeficiente de correlación mide relación lineal entre las variables aleatorias.

*Demostración.*

1. Por un lado,

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= E[(aX + b)(cY + d)] - E(aX + b)E(cY + d) = \\ &= E[acXY + adX + bcY + bd] - (aE(X) + b)(cE(Y) + d) = \\ &= acE(XY) + adE(X) + bcE(Y) \\ &\quad + bd - [acE(X)E(Y) + adE(X) + bcE(Y) + bd] = \\ &= ac[E(XY) - E(X)E(Y)] = ac \text{Cov}(X, Y) \end{aligned}$$

Por otra parte,  $\sigma_{aX+b} = |a|\sigma_X$  y  $\sigma_{cY+d} = |c|\sigma_Y$ , y por lo tanto

$$\rho(aX + b, cY + d) = \frac{\text{Cov}(aX + b, cY + d)}{\sigma_{aX+b}\sigma_{cY+d}} = \frac{ac \text{Cov}(X, Y)}{|a||c|\sigma_X \sigma_Y} = \text{sg}(ac)\rho(X, Y) \quad \square$$

2. Consideremos la función real

$$q(t) = E[(Y - E(Y)) - t(X - E(X))]^2 = E[V - tW]^2$$

siendo  $V = Y - E(Y)$ ,  $W = X - E(X)$ . Como

$$q(t) = E[V - tW]^2 = E(V^2) - 2tE(VW) + t^2E(W^2)$$

es una función cuadrática en  $t$  que toma valores mayores o iguales que 0, o su gráfico no corta al eje  $t$  o lo hace en un solo punto. Es decir que la ecuación  $q(t) = 0$  tiene a lo sumo una raíz, y por lo tanto su discriminante es menor o igual que 0. En nuestro caso, el discriminante es

$$4[E(VW)]^2 - 4E(V^2)E(W^2)$$

y, por lo tanto,

$$4[E(VW)]^2 - 4E(V^2)E(W^2) \leq 0 \Leftrightarrow \frac{[E(VW)]^2}{E(V^2)E(W^2)} \leq 1 \Leftrightarrow$$

$$\frac{[E((X - E(X))(Y - E(Y)))]^2}{E[(Y - E(Y))^2]E[(X - E(X))^2]} \leq 1 \Leftrightarrow [\rho(X, Y)]^2 \leq 1 \Leftrightarrow -1 \leq \rho(X, Y) \leq 1$$

3. ( $\Rightarrow$ ) Si  $\rho^2(X, Y) = 1$ , volviendo a la demostración de la propiedad anterior, existe  $q_0$  tal que  $q(t_0) = 0$ , es decir, que  $E[V - t_0W]^2 = 0$ . Pero además,  $E(V - t_0W) = 0$ , pues  $V$  y  $W$  tienen esperanza igual a 0. Entonces, la v.a.  $V - t_0W$  tiene varianza cero y por lo tanto es constante con probabilidad 1, es decir

$$P(V - t_0W = E(V - t_0W)) = P(V - t_0W = 0) = 1$$

o sea

$$P((Y - E(Y)) - t_0(X - E(X)) = 0) = 1 \Leftrightarrow P(Y = t_0X + E(Y) - t_0E(X)) = 1$$

Entonces,  $Y = aX + b$  con probabilidad 1, siendo  $a = t_0$  y  $b = E(Y) - t_0E(X)$ . Falta verificar que  $a = t_0 \neq 0$ . En efecto, si  $t_0$  fuese igual a 0, esto implicaría que  $E(V^2) = V(Y) = 0$ .

( $\Leftarrow$ ) Sea  $Y = aX + b$  para ciertos valores  $a \neq 0$  y  $b$ . Entonces

$$\begin{aligned} \rho(X, Y) &= \rho(X, aX + b) = \frac{\text{Cov}(X, aX + b)}{\sigma_X \sigma_{aX + b}} = \frac{E(X(aX + b)) - E(X)E(aX + b)}{\sigma_X |a| \sigma_X} = \\ &= \frac{aE(X^2) + bE(X) - a[E(X)]^2 - bE(X)}{|a| \sigma_X^2} = \\ &= \frac{a(E(X^2) - [E(X)]^2)}{|a| \sigma_X^2} = \frac{a \sigma_X^2}{|a| \sigma_X^2} = \pm 1 \end{aligned}$$

## 4.2. Extensión a más de dos dimensiones

### 4.2.1. Vectores aleatorios discretos

**Definición.** Sean  $X_1, \dots, X_k$  variables aleatorias discretas. La **función de probabilidad conjunta** del vector aleatorio  $(X_1, \dots, X_k)$  se define como

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k)$$

y, dado cualquier conjunto  $A \subseteq \mathbb{R}^k$ ,

$$P((X_1, \dots, X_k) \in A) = \sum_{(x_1, \dots, x_k) \in A} \dots \sum p_{X_1, \dots, X_k}(x_1, \dots, x_k)$$

Esta función satisface las siguientes propiedades

- $p_{X_1, \dots, X_k}(x_1, \dots, x_k) \geq 0 \quad \forall (x_1, \dots, x_k)$ .
- $\sum_{x_1} \cdots \sum_{x_k} p_{X_1, \dots, X_k}(x_1, \dots, x_k) = 1$ .

En forma similar al caso bidimensional, pueden definirse las funciones de probabilidad marginal. Por ejemplo, la función de probabilidad marginal de  $X_1$  está dada por

$$p_{X_1}(x_1) = \sum_{x_2} \cdots \sum_{x_k} p_{X_1, \dots, X_k}(x_1, \dots, x_k) = 1$$

y la función de probabilidad marginal de  $(X_1, X_2)$  está dada por

$$p_{X_1, X_2}(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_k} p_{X_1, \dots, X_k}(x_1, \dots, x_k) = 1$$

### Distribución multinomial.

**Definición.** Un *experimento multinomial* es aquel en el que se repite  $n$  veces una experiencia, de forma tal que en cada repetición hay  $k \geq 2$  resultados posibles, cada uno de los cuales ocurre con probabilidad  $p_i$  ( $1 \leq i \leq k$ ) constante a lo largo de todas las repeticiones. Si definimos las variables aleatorias

$$X_i : \text{número de veces que ocurre el resultado } i$$

para  $1 \leq i \leq k$ , decimos que el vector aleatorio  $(X_1, \dots, X_k)$  tiene **distribución conjunta multinomial** con parámetros  $n, p_1, \dots, p_k$ , y lo notamos  $(X_1, \dots, X_n) \sim M(n, p_1, \dots, p_k)$ .

Su función de probabilidad conjunta está dada por

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = \begin{cases} \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i} & \text{si } 0 \leq x_i \leq n \quad \forall i, \quad \sum_{i=1}^k x_i = n \\ 0 & \text{en otro caso} \end{cases}$$

*Observación.* La distribución marginal de  $X_i$  es binomial de parámetros  $n$  y  $p_i$  para todo  $1 \leq i \leq k$ . En general, las marginales de una distribución multinomial son binomiales o multinomiales.

#### 4.2.2. Vectores aleatorios continuos

**Definición.** El vector aleatorio  $(X_1, \dots, X_k)$  es continuo si existe una función  $f_{X_1, \dots, X_k} : \mathbb{R}^k \rightarrow \mathbb{R}_{\geq 0}$ , denominada **función de densidad conjunta**, tal que

$$P((X_1, \dots, X_k) \in A) = \int \cdots \int_A f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_1 \cdots dx_k \quad \forall A \subseteq \mathbb{R}^k$$

Esta función satisface las siguientes propiedades

- $f_{X_1, \dots, X_k}(x_1, \dots, x_k) \geq 0 \quad \forall (x_1, \dots, x_k)$ .
- $\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_1 \cdots dx_k = 1$ .

En forma similar al caso bidimensional, pueden definirse las funciones de densidad marginal. Por ejemplo, la función de densidad marginal de  $X_1$  está dada por

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_2 \cdots dx_k$$

y la función de probabilidad marginal de  $(X_1, X_2)$  está dada por

$$f_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_3 \cdots dx_k$$

### 4.2.3. Independencia de variables aleatorias

**Definición.**  $X_1, \dots, X_k$  son variables aleatorias **independientes** si y solo si

- $p_{X_1, \dots, X_k}(x_1, \dots, x_k) = \prod_{i=1}^k p_{X_i}(x_i) \forall (x_1, \dots, x_k)$  en el caso discreto.
- $f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \prod_{i=1}^k f_{X_i}(x_i) \forall (x_1, \dots, x_k)$ , salvo a lo sumo en un conjunto de probabilidad cero, en el caso continuo.

### 4.2.4. Distribución de la suma de variables aleatorias independientes

**Proposición.** Sean  $X$  e  $Y$  dos v.a. independientes. Entonces, la función generadora de momentos de la suma  $X + Y$  es el producto de las funciones generadoras de momentos de  $X$  e  $Y$ , es decir

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

*Demostración.* En el caso continuo, tenemos que, si  $X$  e  $Y$  son v.a. independientes

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E} \left( e^{t(X+Y)} \right) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{t(x+y)} f_{XY}(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{tx} e^{ty} f_X(x) f_Y(y) dx dy = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx \int_{-\infty}^{+\infty} e^{ty} f_Y(y) dy = \\ &= \mathbb{E} \left( e^{tX} \right) \mathbb{E} \left( e^{tY} \right) = M_X(t)M_Y(t) \end{aligned}$$

La demostración en el caso discreto es similar. □

*Observación.* Es inmediato verificar que, si  $X_1, \dots, X_n$  son v.a. independientes,

$$M_{X_1+\dots+X_n}(t) = \prod_{i=1}^n M_{X_i}(t)$$

A partir del resultado anterior, puede conocerse la distribución de la suma o de combinaciones lineales de ciertas variables aleatorias independientes, dado que la función generadora de momentos determina la distribución de la v.a.

1. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim \text{Bi}(n_i, p)$ , entonces

$$\sum_{i=1}^n X_i \sim \text{Bi} \left( \sum_{i=1}^n n_i, p \right)$$

En particular, si  $X_i \sim \text{Bi}(1, p) \forall i$ , entonces  $\sum_{i=1}^n X_i \sim \text{Bi}(n, p)$ .

2. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim \mathcal{P}(\lambda_i)$ , entonces

$$\sum_{i=1}^n X_i \sim \mathcal{P} \left( \sum_{i=1}^n \lambda_i \right)$$

3. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim \text{G}(p)$ , entonces

$$\sum_{i=1}^n X_i \sim \text{BN}(n, p)$$

4. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim \mathcal{E}(\lambda)$ , entonces

$$\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

5. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim \Gamma(n_i, \lambda)$ , entonces

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n n_i, \lambda\right)$$

6. Si  $X_1, \dots, X_n$  son v.a. independientes tales que  $X_i \sim N(\mu_i, \sigma_i^2)$ , y  $a_1, \dots, a_n$  son números reales, entonces

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

En particular, si  $X_1, \dots, X_n$  son v.a. independientes e idénticamente distribuidas tales que  $X_i \sim N(\mu, \sigma^2) \forall i$ , entonces

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{y} \quad T = \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$

#### 4.2.5. Sumas y promedios de variables aleatorias

**Proposición.** Sean  $X_1, \dots, X_n$  variables aleatorias cualesquiera, y  $a_1, \dots, a_n$  números reales. Entonces

$$\begin{aligned} \mathbb{E}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i \mathbb{E}(X_i) \\ \mathbb{V}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i^2 \mathbb{V}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j) \end{aligned}$$

*Demostración.* La expresión para la esperanza puede ser probada mediante inducción en  $n$ . El caso  $n = 2$  ya fue demostrado anteriormente. Supongamos ahora que la expresión es cierta para  $n = k$  y probémosla para  $n = k + 1$ .

$$\mathbb{E}\left(\sum_{i=1}^{k+1} a_i X_i\right) = \mathbb{E}\left(\sum_{i=1}^k a_i X_i + a_{k+1} X_{k+1}\right) = \mathbb{E}(Y + a_{k+1} X_{k+1})$$

siendo  $Y = \sum_{i=1}^k a_i X_i$ . Como para  $n = 2$  la igualdad se cumple, se obtiene

$$\mathbb{E}\left(\sum_{i=1}^{k+1} a_i X_i\right) = \mathbb{E}(Y + a_{k+1} X_{k+1}) = \mathbb{E}(Y) + a_{k+1} \mathbb{E}(X_{k+1}) = \mathbb{E}\left(\sum_{i=1}^k a_i X_i\right) + a_{k+1} \mathbb{E}(X_{k+1})$$

y, utilizando la hipótesis inductiva,

$$\mathbb{E}\left(\sum_{i=1}^{k+1} a_i X_i\right) = \sum_{i=1}^k a_i \mathbb{E}(X_i) + a_{k+1} \mathbb{E}(X_{k+1}) = \sum_{i=1}^{k+1} a_i \mathbb{E}(X_i)$$

como queríamos demostrar.

Probemos ahora la expresión correspondiente a la varianza.

$$\begin{aligned}
V\left(\sum_{i=1}^n a_i X_i\right) &= \text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{i=1}^n a_i X_i\right) = \\
&= E\left(\sum_{i=1}^n a_i X_i \cdot \sum_{j=1}^n a_j X_j\right) - E\left(\sum_{i=1}^n a_i X_i\right) E\left(\sum_{j=1}^n a_j X_j\right) = \\
&= E\left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j X_i X_j\right) - \left(\sum_{i=1}^n a_i E(X_i)\right) \left(\sum_{j=1}^n a_j E(X_j)\right) = \\
&= \left(\sum_{i=1}^n \sum_{j=1}^n a_i a_j E(X_i X_j)\right) - \sum_{i=1}^n \sum_{j=1}^n a_i a_j E(X_i) E(X_j) = \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j (E(X_i X_j) - E(X_i) E(X_j)) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)
\end{aligned}$$

Teniendo en cuenta que si  $i = j$ ,  $\text{Cov}(X_i, X_i) = V(X_i)$  y que  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ , obtenemos el resultado que queríamos demostrar.  $\square$

**Corolario.** Sean  $X_1, \dots, X_n$  variables aleatorias independientes, y  $a_1, \dots, a_n$  números reales. Entonces

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(x_i) \quad y \quad V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i)^2$$

**Corolario.** Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas, con  $E(X_i) = \mu$  y  $V(X_i) = \sigma^2 \forall i = 1, \dots, n$ , y  $a_1, \dots, a_n$  números reales. Entonces

$$E\left(\sum_{i=1}^n a_i X_i\right) = \mu \sum_{i=1}^n a_i \quad y \quad V\left(\sum_{i=1}^n a_i X_i\right) = \sigma^2 \sum_{i=1}^n a_i^2$$

**Propiedad.** Sean  $X_1, \dots, X_n$  variables aleatorias independientes e idénticamente distribuidas, con  $E(X_i) = \mu$  y  $V(X_i) = \sigma^2 \forall i = 1, \dots, n$ . Entonces

$$\begin{aligned}
E\left(\sum_{i=1}^n X_i\right) &= n\mu \quad y \quad V\left(\sum_{i=1}^n X_i\right) = n\sigma^2 \\
E(\bar{X}) &= E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \mu \quad y \quad V(\bar{X}) = V\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sigma^2}{n}
\end{aligned}$$

### 4.3. Desigualdad de Chebyshev

**Proposición** (Desigualdad de Márkov). Sea  $X$  una v.a. cualquiera que solo toma valores no negativos. Entonces,

$$\forall \varepsilon > 0, \quad P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$$

*Demostración.* Dado  $\varepsilon > 0$ , sea la v.a.

$$I = \begin{cases} 1 & \text{si } X \geq \varepsilon \\ 0 & \text{en otro caso} \end{cases}$$

Notemos que, dado que  $X \geq 0$ ,

$$I \leq \frac{X}{\varepsilon}$$

Tomando esperanzas en la desigualdad anterior, resulta

$$E(I) \leq \frac{E(X)}{\varepsilon}$$

Teniendo en cuenta que  $E(I) = P(X \geq a)$ , se obtiene el resultado que queríamos probar.  $\square$

**Proposición** (Desigualdad de Chebyshev). *Sea  $X$  una v.a. con  $E(X) = \mu$  y  $V(X) = \sigma^2 < \infty$ . Entonces*

$$\forall \varepsilon > 0, \quad P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

*Demostración.* Dado que  $(X - \mu)^2$  es una v.a. no negativa, podemos aplicar la desigualdad de Márkov, obteniendo

$$P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{E((X - \mu)^2)}{\varepsilon^2}$$

lo cual, teniendo en cuenta que  $(X - \mu)^2 \geq \varepsilon^2 \Leftrightarrow |X - \mu| \geq \varepsilon$ , equivale a

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \quad \square$$

*Demostración alternativa.* En el caso continuo,

$$\begin{aligned} \sigma^2 &= E((X - \mu)^2) = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x) dx = \\ &= \int_{\{|X - \mu| \geq \varepsilon\}} (x - \mu)^2 f_X(x) dx + \int_{\{|X - \mu| < \varepsilon\}} (x - \mu)^2 f_X(x) dx \geq \\ &\geq \int_{\{|X - \mu| \geq \varepsilon\}} (x - \mu)^2 f_X(x) dx \geq \int_{\{|X - \mu| \geq \varepsilon\}} \varepsilon^2 f_X(x) dx = \varepsilon^2 P(|X - \mu| \geq \varepsilon) \end{aligned}$$

Entonces  $\frac{\sigma^2}{\varepsilon^2} \geq P(|X - \mu| \geq \varepsilon)$ , como queríamos demostrar.  $\square$

*Observación.* La desigualdad de Chebyshev provee una cota para la probabilidad de un evento descrito en términos de una variable aleatoria que no depende de la distribución, sino solo de la esperanza y la varianza de la v.a.

Sin embargo, esta cota puede ser grosera, o incluso no informativa si, por ejemplo,  $\varepsilon^2 \leq \sigma^2$ .

#### Formas equivalentes de la desigualdad de Chebyshev.

- $\forall \varepsilon > 0, \quad P(|X - \mu| \leq \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$
- $\forall k > 1, \quad P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}.$
- $\forall k > 1, \quad P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}.$

Las últimas dos formas muestran cómo el desvío estándar mide el grado de concentración de la distribución alrededor de  $\mu = E(X)$ .

#### 4.4. Ley de los Grandes Números

**Definición.** Sea  $(X_n)$ ,  $n \geq 1$ , una sucesión de variables aleatorias. Diremos que  $X_n$  **converge en probabilidad** a la variable aleatoria  $X$ , y lo notaremos  $X_n \xrightarrow{p} X$ , si

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

**Ley de los Grandes Números.** Sean  $X_1, X_2, \dots$  variables aleatorias independientes e idénticamente distribuidas (muestra aleatoria), con  $E(X) = \mu$  y  $V(X) = \sigma^2 < \infty$ . Entonces

$$\bar{X}_n \xrightarrow{p} \mu$$

siendo  $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$  el promedio muestral.

*Demostración.* Sabemos que  $E(\bar{X}_n) = \mu$  y  $V(\bar{X}_n) = \frac{\sigma^2}{n}$ . Aplicando la desigualdad de Chebyshev

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \quad \forall \varepsilon > 0$$

y, por lo tanto

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n\varepsilon^2} = 0 \quad \forall \varepsilon > 0$$

Luego,  $\bar{X}_n \xrightarrow{p} \mu$ , como queríamos demostrar.  $\square$

**Proposición** (Versión Bernoulli de la Ley de los Grandes Números). Consideremos  $n$  repeticiones independientes de un experimento aleatorio, y un suceso  $A$  con probabilidad  $p$  constante a lo largo de las  $n$  repeticiones. Si llamamos  $n_A$  al total de veces que ocurre  $A$  en las  $n$  repeticiones, y  $f_A = \frac{n_A}{n}$  a la frecuencia relativa de  $A$ , tenemos que

$$f_A \xrightarrow{p} p$$

es decir

$$\lim_{n \rightarrow \infty} P(|f_A - p| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

*Demostración.* Como  $n_A \sim \text{Bi}(n, p)$  con  $p = P(A)$ , entonces  $E(n_A) = np$  y  $V(n_A) = np(1-p)$ . Luego

$$E(f_A) = E\left(\frac{n_A}{n}\right) = p \quad \text{y} \quad V(f_A) = V\left(\frac{n_A}{n}\right) = \frac{p(1-p)}{n}$$

Aplicando la desigualdad de Chebyshev

$$P(|f_A - p| \geq \varepsilon) \leq \frac{V(f_A)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \quad \forall \varepsilon > 0$$

Luego

$$\lim_{n \rightarrow \infty} P(|f_A - p| \geq \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{p(1-p)}{n\varepsilon^2} = 0 \quad \forall \varepsilon > 0$$

como queríamos demostrar.  $\square$

#### 4.5. Teorema Central del Límite

**Teorema** (Central del Límite). Sean  $X_1, X_2, \dots$  v.a. independientes e idénticamente distribuidas, con  $E(X_i) = \mu$  y  $V(X_i) = \sigma^2 < \infty \forall i$ . Entonces, si  $n$  es suficientemente grande

$$\frac{T - n\mu}{\sqrt{n}\sigma} \stackrel{(a)}{\approx} N(0, 1) \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \stackrel{(a)}{\approx} N(0, 1)$$

o, dicho de otro modo

$$\frac{T - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z \sim N(0, 1) \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1)$$

donde la convergencia en distribución ( $\xrightarrow{d}$ ) se interpreta en el siguiente sentido

$$P\left(\frac{T - n\mu}{\sqrt{n}\sigma} \leq a\right) \cong \Phi(a) \quad P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq a\right) \cong \Phi(a)$$

es decir, que las funciones de distribución convergen a la función de distribución normal estándar.

*Demostración.* Lo demostraremos bajo la hipótesis de que la función generadora de momentos de  $X_i$ ,  $M_{X_i}(t)$ , existe y es finita.

Supongamos inicialmente que  $\mu = 0$  y  $\sigma^2 = 1$ . En este caso, la función generadora de momentos de  $\frac{T}{\sqrt{n}}$  está dada por

$$M_{\frac{T}{\sqrt{n}}}(t) = M_T\left(\frac{t}{\sqrt{n}}\right) = M_{\sum_{i=1}^n X_i}\left(\frac{t}{\sqrt{n}}\right) = \prod_{i=1}^n M_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(M_{X_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

por ser las  $X_i$  independientes. Sea  $L(u) = \ln(M_{X_i}(u))$ , entonces

$$\begin{aligned} L(0) &= \ln(M_{X_i}(0)) = \ln(1) = 0 \\ L'(0) &= \left. \frac{\partial \ln(M_{X_i}(u))}{\partial u} \right|_{u=0} = \left. \frac{M'_{X_i}(u)}{M_{X_i}(u)} \right|_{u=0} = \frac{M'_{X_i}(0)}{M_{X_i}(0)} = \frac{\mu}{1} = \mu = 0 \\ L''(0) &= \left. \frac{\partial^2 \ln(M_{X_i}(u))}{\partial u^2} \right|_{u=0} = \left. \frac{M''_{X_i}(u)M_{X_i}(u) - [M'_{X_i}(u)]^2}{[M_{X_i}(u)]^2} \right|_{u=0} = \\ &= \frac{M''_{X_i}(0)M_{X_i}(0) - [M'_{X_i}(0)]^2}{[M_{X_i}(0)]^2} = E(X_i^2) = 1 \end{aligned}$$

Ahora, para poder probar el teorema, demostraremos que  $M_{\frac{T}{\sqrt{n}}} \rightarrow e^{\frac{t^2}{2}}$ , o, equivalentemente, que  $nL\left(\frac{t}{\sqrt{n}}\right) \rightarrow \frac{t^2}{2}$ . Aplicando la regla de L'Hospital dos veces

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{\frac{1}{n}} &= \lim_{n \rightarrow \infty} \frac{-L'\left(\frac{t}{\sqrt{n}}\right)tn^{-\frac{3}{2}}}{-2n^{-2}} = \lim_{n \rightarrow \infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right)t}{2n^{-\frac{1}{2}}} = \\ &= \lim_{n \rightarrow \infty} \frac{-L''\left(\frac{t}{\sqrt{n}}\right)t^2n^{-\frac{3}{2}}}{-2n^{-\frac{3}{2}}} = \lim_{n \rightarrow \infty} \frac{L''\left(\frac{t}{\sqrt{n}}\right)t^2}{2} = \frac{t^2}{2} \end{aligned}$$

por lo tanto, hemos probado el Teorema Central del Límite para  $\mu = 0$  y  $\sigma^2 = 1$ . El caso general resulta considerando las v.a. estandarizadas,  $\frac{X_i - \mu}{\sigma} = X_i^*$ .  $\square$

## Aplicaciones del Teorema Central del Límite

**Aproximación de la distribución binomial por la normal.** Sea  $X \sim \text{Bi}(n, p)$ , entonces  $X$  es el número de éxitos en  $n$  repeticiones de un experimento binomial con probabilidad de éxito igual a  $p$ , y  $\frac{X}{n}$  es la proporción muestral de éxitos. Si definimos las variables aleatorias

$$X_i = \begin{cases} 1 & \text{si se obtuvo éxito en la } i\text{-ésima repetición} \\ 0 & \text{si se obtuvo fracaso en la } i\text{-ésima repetición} \end{cases}$$

para  $i = 1, \dots, n$ , estas v.a. son independientes, con  $X_i \sim \text{Bi}(1, p) \forall i$  y  $\sum_{i=1}^n X_i = X$ .

Aplicando el Teorema Central del Límite, si  $n$  es suficientemente grande,

$$X \stackrel{(a)}{\approx} N(np, np(1-p)) \quad \frac{X}{n} \stackrel{(a)}{\approx} N\left(p, \frac{p(1-p)}{n}\right)$$

Se considera que la aproximación es buena si  $np \geq 5$  y  $n(1-p) \geq 5$ .

**Corrección por continuidad.** Cuando se aproxima una distribución discreta por una continua, como es el caso de la aproximación de la distribución binomial por la normal, es necesario efectuar una corrección. En general, cuando la v.a. es discreta y  $x_i - x_{i-1} = 1$ , la corrección se realiza en la forma:

$$P(X \leq a) = P(X \leq a + 0,5)$$

$$P(X \geq a) = P(X \geq a - 0,5)$$

**Aproximación de la distribución de Poisson por la normal.** Sean  $X_1, \dots, X_n$  v.a. independientes e idénticamente distribuidas, con distribución de Poisson de parámetro  $(\lambda)$ . Entonces

$$\sum_{i=1}^n X_i \sim \mathcal{P}(n\lambda)$$

Por lo tanto, cualquier v.a. con distribución de Poisson con parámetro suficientemente grande puede ser aproximada por la distribución normal.

**Aproximación de la distribución gamma por la normal.** Sean  $X_1, \dots, X_n$  v.a. independientes e idénticamente distribuidas, tales que  $X_i \sim \Gamma(n_i, \lambda)$ . Entonces

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n n_i, \lambda\right)$$

Por lo tanto, cualquier v.a. con distribución  $\Gamma(\alpha, \lambda)$  con parámetro  $\alpha$  suficientemente grande puede ser aproximada por la distribución normal.

## 5. Inferencia estadística

### 5.1. Conceptos básicos de una investigación estadística

**Población:** Conjunto total de los sujetos o unidades de análisis de interés en el estudio

**Muestra:** Cualquier subconjunto de sujetos o unidades de análisis de la población en estudio.

**Unidad de análisis o de observación:** Objeto bajo estudio. Puede ser una persona, una familia, un país, una institución o, en general, cualquier objeto.

**Variable:** Cualquier característica de la unidad de observación que interese registrar y que en el momento de ser registrada puede ser transformada en un número.

**Valor de una variable, dato, observación o medición:** Número que describe a la característica de interés en una unidad de observación particular.

**Caso o registro:** Conjunto de mediciones realizadas sobre una unidad de observación.

#### 5.1.1. Medidas de resumen

Son medidas de fácil interpretación que reflejan las características más relevantes de los datos provenientes de variables numéricas.

**Medidas de posición o centrado.** ¿Cuál es el valor central o que mejor representa a los datos?

Supongamos que disponemos de un conjunto de  $n$  datos, que representaremos por  $x_1, \dots, x_n$ .

**Promedio o media muestral.**  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ .

**Mediana muestral.** Sean los estadísticos de orden muestrales  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ . Definimos como mediana  $\tilde{x} = \begin{cases} x_{(k+1)} & \text{si } n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2} & \text{si } n = 2k \end{cases}$ .

**Media  $\alpha$ -podada.**  $\bar{x}_\alpha = \frac{x_{(\lfloor n\alpha \rfloor + 1)} + \dots + x_{(n - \lfloor n\alpha \rfloor)}}{n - 2\lfloor n\alpha \rfloor}$ .

**Medidas de dispersión o variabilidad.** ¿Cuán dispersos están los datos? ¿Cuán cercanos son los datos al valor típico?

**Rango muestral.**  $R(x) = \max(x_i) - \min(x_i)$ .

**Varianza muestral.**  $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$ .

**Desvío estándar muestral.**  $S = \sqrt{S^2}$ .

**Distancia intercuartil.**  $d_I =$  cuartil superior  $-$  cuartil inferior  $= x_{0,75} - x_{0,25}$ , donde  $x_\alpha$ ,  $0 \leq \alpha \leq 1$ , es el  $100\alpha$ -ésimo percentil de la muestra, es decir, el valor por debajo del cual se encuentra el  $100\alpha\%$  de la muestra ordenada.

**Distancia entre cuartos.**  $d_C =$  cuarto superior  $-$  cuarto inferior, donde los cuartos inferior y superior son la mediana de la primera y la segunda mitad de la muestra ordenada, respectivamente (en caso de que el tamaño de la muestra sea impar, se extienden ambas mitades con la mediana de la muestra completa).

**Desvío absoluto mediano.**  $MAD = \text{mediana}(|x_i - \tilde{x}|)$ .

## 5.2. Estimación puntual

La mayoría de las distribuciones de probabilidad dependen de cierto número de parámetros. El objetivo de la *estimación puntual* es usar una muestra para obtener números que representen lo mejor posible a los verdaderos valores de los parámetros de interés.

Supongamos que se selecciona una muestra de tamaño  $n$  de una población. Antes de obtener la muestra, la primera observación puede ser considerada una v.a.  $X_1$ , la segunda una v.a.  $X_2$ , etc. Por lo tanto, antes de obtener la muestra denotaremos  $X_1, X_2, \dots, X_n$  a las observaciones y, una vez obtenida la muestra, denotaremos  $x_1, x_2, \dots, x_n$  a los valores observados.

**Definición.** Dado un parámetro  $\theta$ , un **estimador puntual** del parámetro es un valor que se obtiene a partir de alguna función de la muestra y que puede considerarse representativo del parámetro. Lo denotaremos  $\hat{\theta}$ .

### 5.2.1. Método de estimación por momentos

**Definición.** Dada una muestra aleatoria  $X_1, \dots, X_n$ , se denomina **momento muestral** de orden  $k$  a

$$\frac{\sum_{i=1}^n X_i^k}{n}$$

**Definición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución cuya función de probabilidad puntual o función de densidad depende de  $m$  parámetros  $\theta_1, \dots, \theta_m$ . Los **estimadores de momentos** de estos parámetros son los valores  $\hat{\theta}_1, \dots, \hat{\theta}_m$  que se obtienen igualando  $m$  momentos poblacionales (los correspondientes la distribución cuyos parámetros se busca estimar) con los correspondientes momentos muestrales. Se obtienen, en general, resolviendo el sistema de ecuaciones

$$\frac{\sum_{i=1}^n X_i^k}{n} = E(X^k) \quad k = 1, \dots, m$$

### 5.2.2. Método de estimación de máxima verosimilitud

**Definición.** Sean  $X_1, \dots, X_n$  v.a. con función de probabilidad conjunta  $p_X(x_1, \dots, x_n)$  o función de densidad conjunta  $f_X(x_1, \dots, x_n)$  que depende de  $m$  parámetros  $\theta_1, \dots, \theta_m$ . Si  $(x_1, \dots, x_n)$  son los valores observados, la **función de verosimilitud** de la muestra es la función  $L: \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$  que se obtiene al considerar a la función de probabilidad o de densidad conjunta como función de los parámetros  $\theta_1, \dots, \theta_m$ , y se denota  $L(\theta_1, \dots, \theta_m)$ .

**Definición.** Los **estimadores de máxima verosimilitud** (EMV) de  $\theta_1, \dots, \theta_m$  son los valores  $\hat{\theta}_1, \dots, \hat{\theta}_m$  que maximizan la función de verosimilitud, es decir

$$L(\hat{\theta}_1, \dots, \hat{\theta}_m) \geq L(\tilde{\theta}_1, \dots, \tilde{\theta}_m) \quad \forall \tilde{\theta}_1, \dots, \tilde{\theta}_m$$

La forma general de los EMV se obtiene reemplazando los valores observados  $x_i$  por las variables aleatorias  $X_i$ .

**Propiedad** (Invarianza de los EMV). Si  $\hat{\theta}$  es el EMV de un parámetro  $\theta$  y  $h$  es una función inyectiva con dominio en el rango de valores posibles de  $\theta$ , entonces el EMV de  $h(\theta)$  es  $h(\hat{\theta})$ .

### 5.2.3. Propiedades de los estimadores

**Definición.** Un estimador puntual  $\hat{\theta}$  del parámetro  $\theta$ , es **insesgado** si

$$E_{\theta}(\hat{\theta}) = \theta \quad \forall \theta$$

Si  $\hat{\theta}$  no es insesgado, se denomina **sesgo** de  $\hat{\theta}$  a  $b(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta$ .

**Definición.** Un estimador puntual  $\hat{\theta}$  del parámetro  $\theta$ , basado en una muestra  $X_1, \dots, X_n$  es **asintóticamente insesgado** si

$$\lim_{n \rightarrow \infty} E_{\theta}(\hat{\theta}) = \theta \quad \forall \theta$$

**Principio de estimación insesgada de mínima varianza.** Entre todos los estimadores insesgados de  $\theta$ , conviene elegir el de menor varianza. Este estimador se denomina IMVU (insesgado de mínima varianza uniformemente).

**Teorema.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu, \sigma^2)$ . Entonces  $\bar{X}$  es estimador IMVU de  $\mu$ .

**Definición.** El **error estándar** de un estimador  $\hat{\theta}$  es su desvío estándar, es decir

$$\sigma_{\hat{\theta}} = \sqrt{V_{\theta}(\hat{\theta})}$$

En caso de depender de parámetros desconocidos, estos se reemplazan por un estimador y se obtiene el error estándar estimado.

**Definición.** Sea  $\hat{\theta}$  un estimador de  $\theta$ . Su error cuadrático medio es

$$ECM_{\theta}(\hat{\theta}) = E_{\theta} [(\hat{\theta} - \theta)^2]$$

Si el estimador  $\hat{\theta}$  es insesgado, el error cuadrático medio es igual a la varianza del estimador.

**Proposición.**

$$ECM_{\theta}(\hat{\theta}) = V_{\theta}(\hat{\theta}) + [b(\hat{\theta})^2]$$

siendo  $b(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta$  el sesgo del estimador.

*Demostración.*

$$\begin{aligned} \text{ECM}_\theta(\hat{\theta}) &= \text{E}_\theta \left[ (\hat{\theta} - \theta)^2 \right] = \text{E}_\theta \left[ (\hat{\theta} - \text{E}_\theta(\hat{\theta}) + \text{E}_\theta(\hat{\theta}) - \theta)^2 \right] = \\ &= \text{E}_\theta \left[ (\hat{\theta} - \text{E}_\theta(\hat{\theta}))^2 + (\text{E}_\theta(\hat{\theta}) - \theta)^2 + 2(\hat{\theta} - \text{E}_\theta(\hat{\theta}))(\text{E}_\theta(\hat{\theta}) - \theta) \right] = \\ &= \text{E}_\theta \left[ (\hat{\theta} - \text{E}_\theta(\hat{\theta}))^2 \right] + \text{E}_\theta \left[ (\text{E}_\theta(\hat{\theta}) - \theta)^2 \right] + 2 \text{E}_\theta \left[ (\hat{\theta} - \text{E}_\theta(\hat{\theta}))(\text{E}_\theta(\hat{\theta}) - \theta) \right] \end{aligned}$$

Usando que la esperanza de una v.a. es una constante y la esperanza de una constante es igual a esta, se obtiene

$$\text{ECM}_\theta(\hat{\theta}) = \underbrace{\text{E}_\theta \left[ (\hat{\theta} - \text{E}_\theta(\hat{\theta}))^2 \right]}_{V_\theta(\hat{\theta})} + \underbrace{(\text{E}_\theta(\hat{\theta}) - \theta)^2}_{(b(\hat{\theta}))^2} + 2(\text{E}_\theta(\hat{\theta}) - \theta) \underbrace{(\text{E}_\theta(\hat{\theta}) - \text{E}_\theta(\hat{\theta}))}_0$$

y, por lo tanto,  $\text{ECM}_\theta(\hat{\theta}) = V_\theta(\hat{\theta}) + [b(\hat{\theta})]^2$ , como queríamos demostrar.  $\square$

**Principio de estimación de menor error cuadrático medio.** Dados dos o más estimadores del parámetro  $\theta$ , conviene elegir el de menor error cuadrático medio.

**Definición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución que depende de un parámetro  $\theta$ , y sea  $\hat{\theta}_n$  un estimador puntual de  $\theta$  basado en esa muestra. Diremos que  $\{\hat{\theta}_n\}$  es una sucesión **consistente**, o más brevemente, que  $\hat{\theta}_n$  es un estimador consistente de  $\theta$ , si

$$\hat{\theta}_n \xrightarrow{p} \theta$$

es decir, si  $\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0$ .

**Proposición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución que depende de un parámetro  $\theta$ , y sea  $\hat{\theta}_n$  un estimador de  $\theta$  basado en esta muestra. Si

1.  $\lim_{n \rightarrow \infty} \text{E}_\theta(\hat{\theta}_n) = \theta$  (es decir, si el estimador es asintóticamente insesgado), y
2.  $\lim_{n \rightarrow \infty} V_\theta(\hat{\theta}_n) = 0$

entonces  $\hat{\theta}_n$  es un estimador consistente de  $\theta$ .

*Demostración.* Si el estimador es insesgado, la demostración es inmediata a partir de la desigualdad de Chebyshev. No daremos la demostración en el caso general.  $\square$

### 5.3. Intervalos de confianza

**Definición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución que depende de un parámetro  $\theta$ . Dadas dos funciones de la muestra  $a(X_1, \dots, X_n)$  y  $b(X_1, \dots, X_n)$ , tales que

$$P(a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n)) = 1 - \alpha$$

con  $\alpha$  pequeño, el intervalo  $[a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$  se denomina **intervalo de confianza** de nivel  $1 - \alpha$  para el parámetro  $\theta$ .

#### 5.3.1. Intervalos de confianza para los parámetros de una distribución normal

**Definición.** Sean dos v.a.  $Z \sim N(0, 1)$  y  $U \sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$  independientes, entonces

$$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n$$

tiene **distribución t de Student** con  $n$  grados de libertad.

Esta distribución está tabulada para diferentes valores de  $n$ . Su distribución es simétrica con respecto al 0 y tiene forma de campana, pero con colas más pesadas que la distribución normal estándar. Cuando  $n$  tiende a infinito, la distribución t de Student tiende a la distribución normal estándar.

**Proposición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu, \sigma^2)$ . Entonces

1.  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1)$ .
2.  $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ , donde  $S^2$  es la varianza muestral.
3.  $\bar{X}$  y  $S^2$  son independientes.
4.  $\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1}$ .

*Demostración.*

1. Ya hemos visto que cualquier combinación lineal de v.a. normales independientes es normal, y el promedio es una combinación lineal particular.
- 2 y 3. Están fuera del alcance de este curso.
4. Resulta de 1, 2 y 3, pues

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1) \quad \text{y} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

son v.a. independientes. Entonces, por la definición de la distribución t de Student

$$\frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1} \quad \square$$

A partir de los resultados anteriores, podemos construir intervalos de confianza para los parámetros de una distribución normal.

**Media de una distribución normal con varianza conocida.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu, \sigma_0^2)$ , con varianza  $\sigma_0^2$  conocida. Entonces

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \sim N(0, 1) \Rightarrow P\left(-z_{\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma_0} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

de donde se deduce el siguiente intervalo de confianza de nivel  $1 - \alpha$  para  $\mu$

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}\right]$$

**Media de una distribución normal con varianza desconocida.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu, \sigma^2)$ . Entonces

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \sim t_{n-1} \Rightarrow P\left(-t_{n-1, \frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X} - \mu}{S} \leq t_{n-1, \frac{\alpha}{2}}\right) = 1 - \alpha$$

de donde se deduce el siguiente intervalo de confianza de nivel  $1 - \alpha$  para  $\mu$

$$\left[\bar{X} - t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

**Varianza de una distribución normal con media conocida.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu_0, \sigma^2)$ , con media  $\mu_0$  conocida. Entonces

$$\frac{X_i - \mu_0}{\sigma} \sim N(0, 1) \quad \forall 1 \leq i \leq n \Rightarrow \left( \frac{X_i - \mu_0}{\sigma} \right)^2 \sim \chi_1^2 = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \quad \forall 1 \leq i \leq n$$

Como además las variables aleatorias son independientes,

$$\sum_{i=1}^n \left( \frac{X_i - \mu_0}{\sigma} \right)^2 \sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \Rightarrow P\left(\chi_{n, 1-\frac{\alpha}{2}}^2 \leq \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sigma^2} \leq \chi_{n, \frac{\alpha}{2}}^2\right) = 1 - \alpha$$

de donde se obtiene el siguiente intervalo de confianza de nivel  $1 - \alpha$  para  $\sigma^2$

$$\left[ \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, \frac{\alpha}{2}}^2}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi_{n, 1-\frac{\alpha}{2}}^2} \right]$$

**Varianza de una distribución normal con media desconocida.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $N(\mu, \sigma^2)$ . Entonces

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \Rightarrow P\left(\chi_{n-1, 1-\frac{\alpha}{2}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{n-1, \frac{\alpha}{2}}\right) = 1 - \alpha$$

de donde se deduce el siguiente intervalo de confianza de nivel  $1 - \alpha$  para  $\sigma^2$

$$\left[ \frac{(n-1)S^2}{\chi_{n-1, \frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi_{n-1, 1-\frac{\alpha}{2}}} \right]$$

### 5.3.2. Método general para obtener intervalos de confianza

**Definición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución que depende de un parámetro  $\theta$ . Una función **pivote** una función  $T(X_1, \dots, X_n, \theta)$ , cuya distribución no depende de  $\theta$  ni de ningún otro parámetro desconocido. Si existe un pivote para  $\theta$ , entonces existen dos valores  $a$  y  $b$  tales que

$$P(a \leq T(X_1, \dots, X_n, \theta) \leq b) = 1 - \alpha$$

de donde es posible obtener un intervalo de confianza de nivel  $1 - \alpha$  para  $\theta$ .

### 5.3.3. Intervalos de confianza de nivel asintótico

En muchos problemas no es posible encontrar intervalos de confianza de nivel exacto  $1 - \alpha$ , o son de muy difícil construcción, ya sea porque no es posible encontrar una función pivote que no dependa del parámetro, o porque no se conoce la distribución exacta de la función pivote. En estos casos, cuando la muestra es grande, es posible obtener intervalos de confianza de nivel aproximado.

**Definición.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución que depende de un parámetro  $\theta$ . Dadas dos sucesiones  $\{a_n(X_1, \dots, X_n)\}$  y  $\{b_n(X_1, \dots, X_n)\}$ , tales que

$$\lim_{n \rightarrow \infty} P(a_n(X_1, \dots, X_n) \leq \theta \leq b_n(X_1, \dots, X_n)) = 1 - \alpha$$

la sucesión de intervalos  $[a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]$  es una sucesión de **intervalos de confianza de nivel asintótico**  $1 - \alpha$  para el parámetro  $\theta$ . También se dice que, si  $n$  es suficientemente grande, el intervalo  $[a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]$  tiene nivel aproximado  $1 - \alpha$ .

**Intervalo de confianza de nivel aproximado para la media de una distribución cualquiera.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con media  $\mu$  y varianza  $\sigma^2 < \infty$ .

Sabemos que  $\bar{X}$  es un estimador insesgado y consistente de  $\mu$ . Si bien no conocemos su distribución exacta, sabemos que

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{d} N(0, 1)$$

Si  $\sigma^2$  es conocido, puede usarse esta función como pivote para obtener un intervalo de nivel aproximado. Si  $\sigma^2$  es desconocido, podemos aplicar la siguiente propiedad

**Propiedad.**

$$\left. \begin{array}{l} Y_n \xrightarrow{d} Y \\ U_n \xrightarrow{p} a \end{array} \right\} \implies U_n Y_n \xrightarrow{d} aY$$

Como  $S$  es un estimador consistente de  $\sigma$ ,  $S \xrightarrow{p} \sigma$ , y por lo tanto  $\frac{\sigma}{S} \xrightarrow{p} 1$ . Luego,

$$\left. \begin{array}{l} \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{d} N(0, 1) \\ \frac{\sigma}{S} \xrightarrow{p} 1 \end{array} \right\} \implies \sqrt{n} \frac{\bar{X} - \mu}{S} \xrightarrow{d} N(0, 1)$$

si  $n$  es lo suficientemente grande

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \stackrel{(a)}{\approx} N(0, 1)$$

A partir de este resultado, para  $n$  suficientemente grande

$$P \left( -z_{\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X} - \mu}{S} \leq z_{\frac{\alpha}{2}} \right) \cong 1 - \alpha$$

de donde se deduce el siguiente intervalo de confianza de nivel aproximado  $1 - \alpha$  para  $\mu$

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

**Intervalo de confianza de nivel asintótico para una proporción.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{Bi}(1, p)$ . Entonces  $X = \sum_{i=1}^n X_i \sim \text{Bi}(n, p)$ . Queremos construir un intervalo de confianza de nivel asintótico  $1 - \alpha$  para  $p$ .

Por el Teorema Central del Límite, tenemos que

$$\hat{p} = \frac{X}{n} = \frac{\sum_{i=1}^n X_i}{n} \stackrel{(a)}{\approx} N \left( p, \frac{p(1-p)}{n} \right)$$

y, por lo tanto

$$P \left( -z_{\frac{\alpha}{2}} \leq \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\frac{\alpha}{2}} \right) \cong 1 - \alpha$$

Hay dos formas de obtener un intervalo de confianza para  $p$  a partir de esta expresión.

1. Como, por la Ley de los Grandes Números,  $\frac{X}{n} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{p} p$ , podemos aplicar la propiedad enunciada antes, reemplazando  $p$  por su estimador  $\frac{X}{n}$  en el denominador del pivote. Entonces

$$P \left( \frac{X}{n} - z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} \leq p \leq \frac{X}{n} + z_{\frac{\alpha}{2}} \sqrt{\frac{\frac{X}{n} \left(1 - \frac{X}{n}\right)}{n}} \right) \cong 1 - \alpha$$

obteniendo un intervalo para  $p$  de nivel aproximado  $1 - \alpha$ .

2. Podemos reescribir la expresión de la forma

$$P \left( \left| \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \right| \leq z_{\frac{\alpha}{2}} \right) \cong 1 - \alpha \quad \Leftrightarrow \quad P \left( \frac{\left(\frac{X}{n} - p\right)^2}{\frac{p(1-p)}{n}} \leq z_{\frac{\alpha}{2}}^2 \right) \cong 1 - \alpha$$

Observemos que

$$\begin{aligned} \frac{\left(\frac{X}{n} - p\right)^2}{\frac{p(1-p)}{n}} \leq z_{\frac{\alpha}{2}}^2 &\quad \Leftrightarrow \quad \left(\frac{X}{n} - p\right)^2 \leq z_{\frac{\alpha}{2}}^2 \frac{p(1-p)}{n} \\ \Leftrightarrow \quad \left(\frac{X}{n}\right)^2 - 2p\frac{X}{n} + p^2 - z_{\frac{\alpha}{2}}^2 \frac{p(1-p)}{n} &\leq 0 \\ \Leftrightarrow \quad p^2 \left(1 + \frac{z_{\frac{\alpha}{2}}^2}{n}\right) - p \left(\frac{2X}{n} + \frac{z_{\frac{\alpha}{2}}^2}{n}\right) + \left(\frac{X}{n}\right)^2 &\leq 0 \end{aligned}$$

Buscando las raíces de esta ecuación de segundo grado, que llamaremos  $\hat{p}_1$  y  $\hat{p}_2$ , obtenemos un intervalo para  $p$  de nivel aproximado  $1 - \alpha$  dado por

$$[\hat{p}_1, \hat{p}_2]$$

#### 5.4. Tests de hipótesis

**Definición.** Un **test de hipótesis** es una regla que permite decidir, basándose en una muestra, entre dos hipótesis referidas a un parámetro de una distribución. A la hipótesis que implica que no hay cambio con respecto a la situación inicial se la denomina *hipótesis nula* y se designa  $H_0$ . La segunda hipótesis se denomina *hipótesis alternativa* y se designa  $H_1$ . El test se basa en un *estadístico*, que es una función o función de la muestra, y en una *zona de rechazo*, que es un conjunto de valores del estadístico para los cuales se rechaza la hipótesis  $H_0$  en función de  $H_1$ .

**Definición.** En un test que contrasta las hipótesis  $H_0$  y  $H_1$ , llamamos **nivel de significación** del test, y denotamos  $\alpha$ , a la probabilidad de cometer un error de tipo I, es decir, de rechazar  $H_0$  cuando es verdadera.

Por otra parte, denotamos  $\beta$  a la probabilidad de cometer un error de tipo II, es decir, de no rechazar  $H_0$  cuando es falsa.

**Definición.** Dados un test de hipótesis y una muestra aleatoria, llamamos **p-valor** de la muestra a la probabilidad, bajo el supuesto de que  $H_0$  es verdadera, de obtener un valor del estadístico al menos tan extremo como el que fue observado.

**Definición.** La **función de potencia** de un test,  $\pi(\mu)$ , es la probabilidad de rechazar la hipótesis nula cuando el verdadero valor del parámetro es  $\mu$ . Permite obtener una expresión general para los dos tipos de errores, pues

$$\pi(\mu) = \begin{cases} \alpha(\mu) & \text{si } \mu \in H_0 \\ 1 - \beta(\mu) & \text{si } \mu \in H_1 \end{cases}$$

donde  $\alpha(\mu)$  y  $\beta(\mu)$  denotan las probabilidades de error de tipo I y II, respectivamente, cuando el verdadero valor del parámetro es  $\mu$ .

**Tipos de hipótesis a testear.** La forma de la región de rechazo dependerá de la hipótesis alternativa a testear.

**Hipótesis unilaterales.**

$$H_0 : \theta = \theta_0 \text{ (o } \theta \leq \theta_0) \quad \text{vs.} \quad H_1 : \theta > \theta_0$$

$$H_0 : \theta = \theta_0 \text{ (o } \theta \geq \theta_0) \quad \text{vs.} \quad H_1 : \theta < \theta_0$$

**Hipótesis bilaterales.**

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

#### 5.4.1. Test de hipótesis para los parámetros de una distribución normal

**Media de una distribución normal con varianza conocida.** Supongamos que  $\sigma^2 = \sigma_0^2$  es conocida, y consideremos la siguientes hipótesis

1.  $H_0 : \mu = \mu_0 \text{ (o } \mu \leq \mu_0) \quad \text{vs.} \quad H_1 : \mu > \mu_0.$
2.  $H_0 : \mu = \mu_0 \text{ (o } \mu \geq \mu_0) \quad \text{vs.} \quad H_1 : \mu < \mu_0.$
3.  $H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$

**Estadístico del test.**

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0}$$

Bajo  $H_0 : \mu = \mu_0$ ,  $T \sim N(0, 1)$ .

**Región de rechazo.** Según la hipótesis alternativa, está dada en cada caso por

1.  $T \geq z_\alpha.$
2.  $T \leq -z_\alpha.$
3.  $|T| \geq z_{\frac{\alpha}{2}}.$

**Función de potencia.** Está dada, en cada caso, por

1.  $\pi(\mu) = 1 - \Phi\left(z_\alpha + \frac{\mu_0 - \mu}{\frac{\sigma_0}{\sqrt{n}}}\right).$
2.  $\pi(\mu) = \Phi\left(-z_\alpha + \frac{\mu_0 - \mu}{\frac{\sigma_0}{\sqrt{n}}}\right).$
3.  $\pi(\mu) = 1 - \Phi\left(z_{\frac{\alpha}{2}} + \frac{\mu_0 - \mu}{\frac{\sigma_0}{\sqrt{n}}}\right) + \Phi\left(-z_{\frac{\alpha}{2}} + \frac{\mu_0 - \mu}{\frac{\sigma_0}{\sqrt{n}}}\right).$

**Media de una distribución normal con varianza desconocida.** Supongamos ahora que la varianza es desconocida, y consideremos las mismas hipótesis sobre  $\mu$

1.  $H_0 : \mu = \mu_0 \text{ (o } \mu \leq \mu_0) \quad \text{vs.} \quad H_1 : \mu > \mu_0.$
2.  $H_0 : \mu = \mu_0 \text{ (o } \mu \geq \mu_0) \quad \text{vs.} \quad H_1 : \mu < \mu_0.$
3.  $H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$

**Estadístico del test.**

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

Bajo  $H_0 : \mu = \mu_0$ ,  $T \sim t_{n-1}$ .

**Región de rechazo.** Según la hipótesis alternativa, está dada en cada caso por

1.  $T \geq t_{n-1, \alpha}$ .
2.  $T \leq -t_{n-1, \alpha}$ .
3.  $|T| \geq t_{n-1, \frac{\alpha}{2}}$ .

**Varianza de una distribución normal con media desconocida.** Consideremos las siguientes hipótesis

1.  $H_0 : \sigma^2 = \sigma_0^2$  (o  $\sigma^2 \leq \sigma_0^2$ ) vs.  $H_1 : \sigma^2 > \sigma_0^2$ .
2.  $H_0 : \sigma^2 = \sigma_0^2$  (o  $\sigma^2 \geq \sigma_0^2$ ) vs.  $H_1 : \sigma^2 < \sigma_0^2$ .
3.  $H_0 : \sigma^2 = \sigma_0^2$  vs.  $H_1 : \sigma^2 \neq \sigma_0^2$ .

**Estadístico del test.**

$$U = \frac{(n-1)S^2}{\sigma_0^2}$$

Bajo  $H_0 : \sigma^2 = \sigma_0^2$ ,  $U \sim \chi_{n-1}^2$ .

**Región de rechazo.** Según la hipótesis alternativa, está dada en cada caso por

1.  $U \geq \chi_{n-1, \alpha}^2$ .
2.  $U \leq \chi_{n-1, 1-\alpha}^2$ .
3.  $U \geq \chi_{n-1, \frac{\alpha}{2}}^2$  o  $U \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2$ .

### 5.4.2. Tests de hipótesis de nivel aproximado

**Test de nivel aproximado para la media de una distribución cualquiera.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con media  $\mu$  y varianza  $\sigma^2 < \infty$ . Como ya enunciamos anteriormente,  $\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow{d} N(0, 1)$ .

Supongamos que se desea testear a nivel aproximado  $\alpha$  alguna de las siguientes hipótesis

1.  $H_0 : \mu = \mu_0$  (o  $\mu \leq \mu_0$ ) vs.  $H_1 : \mu > \mu_0$ .
2.  $H_0 : \mu = \mu_0$  (o  $\mu \geq \mu_0$ ) vs.  $H_1 : \mu < \mu_0$ .
3.  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu \neq \mu_0$ .

y que  $n$  es suficientemente grande. Utilizando como estadístico

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

las siguientes regiones de rechazo proveen tests de nivel aproximado  $\alpha$  para cada una de las hipótesis

1.  $T \geq z_\alpha$ .
2.  $T \leq -z_\alpha$ .
3.  $|T| \geq z_{\frac{\alpha}{2}}$ .

**Test de hipótesis de nivel aproximado para una proporción.** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución  $\text{Bi}(1, p)$ . Entonces  $X = \sum_{i=1}^n X_i \sim \text{Bi}(n, p)$ . Por el Teorema Central del Límite, tenemos que, si  $n$  es suficientemente grande

$$\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} Z \sim N(0, 1)$$

siendo  $\bar{X}$  la proporción muestral o frecuencia relativa de éxitos.

Un test de nivel aproximado  $\alpha$  para las hipótesis

1.  $H_0 : p = p_0$  vs.  $H_1 : p > p_0$
2.  $H_0 : p = p_0$  vs.  $H_1 : p < p_0$
3.  $H_0 : p = p_0$  vs.  $H_1 : p \neq p_0$

se basa en el estadístico

$$\frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

que, bajo  $H_0$ , tiene distribución aproximada  $N(0, 1)$ .

Las regiones de rechazo están dadas por

1.  $\frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \geq z_\alpha.$
2.  $\frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq -z_\alpha.$
3.  $\left| \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right| \geq z_{\frac{\alpha}{2}}.$

### 5.4.3. Relación entre tests de hipótesis bilaterales e intervalos de confianza

**Proposición.** Sea  $IC(X_1, \dots, X_n)$  un intervalo de confianza de nivel  $1 - \alpha$  para un parámetro  $\theta$ , obtenido a partir de una muestra aleatoria  $X_1, \dots, X_n$ . Consideremos el problema de testear las hipótesis

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta \neq \theta_0$$

El test que rechaza  $H_0$  cuando  $\theta_0 \notin IC(X_1, \dots, X_n)$  tiene nivel  $\alpha$ .

## 6. Cadenas de Márkov<sup>1</sup>

**Definición.** Consideremos una secuencia de variables aleatorias  $X_0, X_1, \dots$ , que toman valores en un conjunto  $\mathcal{S}$  finito o infinito numerable. Si interpretamos  $X_n$  como el estado de un sistema en el momento  $n$ , asumiendo que el sistema se encuentra en el estado  $i$  en el momento  $n$  si  $X_n = i$ , dicha sucesión es un *proceso estocástico*. Decimos que la secuencia de variables aleatorias es una **cadena de Márkov** si, cada vez que el sistema se encuentra en el estado  $i$ ,

<sup>1</sup>Esta sección es un breve resumen del capítulo sobre cadenas de Márkov del libro *A First Course in Probability*, de Sheldon Ross, que aparece en la bibliografía sugerida por la cátedra. Lo incluí dado que en el tema se enseñó en la cursada del segundo cuatrimestre de 2014, aunque no puedo asegurar que los contenidos se correspondan exactamente con los del programa de la materia.

existe una probabilidad fija  $p_{ij}$  de que el siguiente estado del sistema sea  $j$ . Es decir, para todo  $i_0, \dots, i_{n-1}, i, j$

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i) = p_{ij}$$

Los valores  $p_{ij}$ ,  $i, j \in \mathcal{S}$  se denominan *probabilidades de transición* de la cadena de Márkov. Las probabilidades de transición satisfacen

$$p_{ij} \geq 0 \quad \forall i, j \in \mathcal{S} \quad \text{y} \quad \sum_{j \in \mathcal{S}} p_{ij} = 1 \quad \forall i \in \mathcal{S}$$

y conforman una *matriz de transición*,  $P = \begin{pmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \\ \vdots & & \ddots \end{pmatrix}$ .

**Proposición** (Ecuaciones de Chapman-Kolmogorov). *En general, definimos las probabilidades de transición en  $n$  pasos de la cadena de Márkov como*

$$p_{ij}^{(n)} = P(X_{n+m} = j \mid X_m = i)$$

*Las ecuaciones de Chapman-Kolmogorov muestran cómo estas probabilidades pueden ser calculadas*

$$p_{ij}^{(n)} = \sum_{k \in \mathcal{S}} p_{ik}^{(r)} p_{kj}^{(n-r)} \quad \forall 0 < r < n$$

*Demostración.*

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = j \mid X_0 = i) \\ &= \sum_k P(X_n = j, X_r = k \mid X_0 = i) \\ &= \sum_k P(X_n = j \mid X_r = k, X_0 = i) P(X_r = k \mid X_0 = i) \\ &= \sum_k P_{kj}^{(n-r)} p_{ik}^{(r)} \end{aligned} \quad \square$$

Aunque las  $p_{ij}^{(n)}$  denotan probabilidades condicionales, podemos usarlas para derivar expresiones para probabilidades incondicionales, condicionando el estado inicial. Por ejemplo

$$P(X_n = j) = \sum_i P(X_n = j \mid X_0 = i) P(X_0 = i) = \sum_i p_{ij}^{(n)} P(X_0 = i)$$

**Proposición.** *Las probabilidades de transición en  $n$  pasos de una cadena de Márkov conforman una matriz de transición en  $n$  pasos,  $P^{(n)}$ , que es igual a la  $n$ -ésima potencia de la matriz de transición. Es decir,  $P^{(n)} = P^n$ .*

*Demostración.* Lo probaremos por inducción en  $n$ . Si  $n = 1$ , no hay nada que probar, puesto que como  $p_{ij} = p_{ij}^{(1)} \forall i, j$ ,  $P = P^{(1)}$ . Supongamos ahora que la expresión es cierta para  $n = m$ , y probemos que también vale para  $n = m + 1$ . Los elementos de  $P^{(m+1)}$  son las probabilidades  $p_{ij}^{(m+1)}$ . Por las ecuaciones de Chapman-Kolmogorov, con  $r = 1$

$$\{P^{(m+1)}\}_{ij} = p_{ij}^{(m+1)} = \sum_{k \in \mathcal{S}} p_{ik} p_{kj}^{(m)}$$

Utilizando la hipótesis inductiva

$$\{P^{(m+1)}\}_{ij} = p_{ij}^{(m+1)} = \sum_{k \in \mathcal{S}} \{P\}_{ik} \{P^m\}_{kj} = \{P \cdot P^m\}_{ij} = \{P^{m+1}\}_{ij}$$

Luego, los elementos de  $P^{(m+1)}$  y  $P^{m+1}$  son los mismos, es decir,  $P^{(m+1)} = P^{m+1}$ , lo cual completa la demostración.  $\square$

Para un gran número de cadenas de Márkov, si  $n \rightarrow \infty$ , la probabilidad  $P_{ij}^{(n)}$  converge a un valor  $\pi_j$  que depende solo de  $j$ . Es decir, para valores grandes de  $n$ , la probabilidad de que el sistema se encuentre en el estado  $j$  luego de  $n$  transiciones es aproximadamente igual a  $\pi_j$ , independientemente del estado inicial.

Puede demostrarse que es condición suficiente para que esto suceda que, para algún  $n > 0$ ,  $p_{ij}^{(n)} > 0 \forall i, j \in \mathcal{S}$ . Las cadenas de Márkov que satisfacen esta condición se denominan *ergódicas*. Dado que

$$p_{ij}^{(n+1)} = \sum_{k \in \mathcal{S}} P_{ij}^{(n)} P_{kj}$$

resulta que, para una cadena ergódica, si  $n \rightarrow \infty$

$$\pi_j = \sum_{k \in \mathcal{S}} \pi_k p_{kj} \quad (1)$$

y, como  $1 = \sum_{j \in \mathcal{S}} P_{ij}^{(n)}$ , obtenemos que si  $n \rightarrow \infty$

$$\sum_{j \in \mathcal{S}} \pi_j = 1 \quad (2)$$

De hecho, puede probarse que las únicas soluciones no negativas de las ecuaciones (1) y (2) son los  $\pi_j$ ,  $j \in \mathcal{S}$ . Al vector  $\pi = (\pi_1, \pi_2, \dots)$  se lo denomina *medida invariante* de la cadena de Márkov.

La cantidad  $\pi_j$  también es igual a la proporción de tiempo que pasa el sistema, a largo plazo, en el estado  $j$ . En efecto, sea  $P_j$  esta proporción. Entonces, como la proporción de tiempo que el sistema pasa en el estado  $k$  es  $P_k$ , y la cadena pasa del estado  $k$  al estado  $j$  con probabilidad  $p_{kj}$ , tenemos que la proporción de tiempo que la cadena de Márkov pasa entrando al estado  $j$  desde el estado  $k$  es igual a  $P_k p_{kj}$ . Sumando sobre  $k$ , tenemos que  $P_j = \sum_k P_k p_{kj}$ . Como es claro que

$$\sum_j P_j = 1$$

y dado que los  $\pi_j$  son solución única de la ecuación anterior, deducimos que  $P_j = \pi_j \forall j \in \mathcal{S}$ .

Esta interpretación de los valores  $\pi_j$  es válida en general, incluso para cadenas no ergódicas.

De la ecuación (1), podemos deducir que la medida invariante de la cadena es un autovector a la izquierda de la matriz de transición con autovalor 1, es decir,  $\pi P = \pi$ . Esto quiere decir que

$$\sum_i \pi_i P(X_1 = j | X_0 = i) = \pi_j$$

y en general, para todo  $n \in \mathbb{N}$ ,  $\pi P^n = \pi$ :

$$\sum_i \pi_i P(X_n = j | X_0 = i) = \pi_j$$

Es decir, si la distribución de  $X_0$  es  $\pi$ , la distribución de  $X_n$  es  $\pi \forall n \geq 0$ .