

---

# INFERENCIA ESTADÍSTICA

---

*Basado en las clases de María del Carmen Morcillo Aixelá*

Autor:  
Jorge Rodríguez Domínguez



# Índice general



# Capítulo 0

## Distribuciones importantes

### 0.1. Distribuciones discretas

#### 0.1.1. Distribución Uniforme

Diremos que  $X \sim U\{x_1, \dots, x_n\}$ ,  $x_i \in \mathbb{R}$ , si  $P(X = x_i) = \frac{1}{n}$ ,  $i = 1, \dots, n$ . Puede verse también de la forma

$$P(X = x_i) = \frac{1}{n} I_{\{x_1, \dots, x_n\}}(x)$$

Además,  $E(X) = \frac{x_1 + \dots + x_n}{n}$ .

#### 0.1.2. Distribución de Bernoulli

Diremos que  $X \sim Ber(p)$ ,  $0 < p < 1$  si

$$P(X = 1) = p \quad \text{y} \quad P(X = 0) = 1 - p \equiv q$$

Visto de otra forma:

$$P(X = x) = p^x (1 - p)^{1-x}, \quad x = 0, 1$$

Además,  $E(X) = p$  y  $V(X) = pq$ .

#### 0.1.3. Distribución Binomial

Realizamos  $n$  pruebas independientes de Bernoulli todas ellas con probabilidad de éxito igual a  $p$ .

Definimos  $X = \text{número de éxitos en las } n \text{ pruebas de Bernoulli}$ . Entonces  $X \sim Bi(n, p)$ .

- Su distribución de probabilidad es

$$P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i \in D_X = \{0, 1, \dots, n\}$$

- $E(X) = np$  y  $V(X) = npq$ .
- *Reproductividad*: Si  $X_1 \sim Bi(n_1, p)$  y  $X_2 \sim Bi(n_2, p)$  entonces  $X_1 + X_2 \sim Bi(n_1 + n_2, p)$ .

#### 0.1.4. Distribución Geométrica

Realizamos pruebas independientes de Bernoulli, todas con probabilidad de éxito igual a  $p$ . Definimos  $X = \text{número de fracasos hasta conseguir el primer éxito}$ . Entonces  $X \sim Ge(p)$ .

- Su distribución de probabilidad es

$$P(X = i) = (1 - p)^i p, \quad i \in \{0, 1, 2, \dots\}$$

- $E(X) = \frac{q}{p}$  y  $V(X) = \frac{q}{p^2}$ .

### 0.1.5. Distribución Binomial Negativa

Realizamos pruebas independientes de Bernoulli, todas con probabilidad de éxito igual a  $p$ . Definimos  $X = \text{número de fracasos hasta conseguir el } m\text{-ésimo éxito}$ . Entonces  $X \sim BN(m, p)$ .

- Su distribución de probabilidad es

$$P(X = i) = \binom{m+i-1}{i} p^m (1-p)^i, \quad i \in \{0, 1, 2, \dots\}$$

- $E(X) = \frac{mq}{p}$  y  $V(X) = \frac{q}{p^2}$ .
- *Reproductividad*: Si  $X_1 \sim BN(m_1, p)$  y  $X_2 \sim BN(m_2, p)$  son independientes, entonces

$$X_1 + X_2 \sim BN(m_1 + m_2, p)$$

### 0.1.6. Distribución de Poisson

Diremos que  $X \sim Po(\lambda)$ ,  $\lambda > 0$  si

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad n \in \{0, 1, 2, \dots\}$$

- $E(X) = \lambda$  y  $V(X) = \lambda$ .
- *Reproductividad*: Si  $X_1 \sim Po(\lambda_1)$  y  $X_2 \sim Po(\lambda_2)$  son independientes, entonces

$$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$$

## 0.2. Distribuciones continuas

### 0.2.1. Distribución uniforme

Diremos que  $X \sim U(a, b)$  si su función de densidad es

$$f(x) = \frac{1}{b-a} \cdot \mathbf{I}_{(a,b)}(x)$$

Además,  $E(X) = \frac{b+a}{2}$  y  $V(X) = \frac{(b-a)^2}{12}$ .

### 0.2.2. Distribución Exponencial

Diremos que  $X \sim Exp(\lambda)$ ,  $\lambda > 0$  si su función de densidad es

$$f(x) = \lambda e^{-\lambda x} \cdot \mathbf{I}_{(0,+\infty)}(x)$$

Además,  $E(X) = \frac{1}{\lambda}$  y  $V(X) = \frac{1}{\lambda^2}$ .

### 0.2.3. Distribución Gamma

Diremos que  $X \sim Ga(\alpha, \beta)$  si su función de densidad es

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} \cdot \mathbf{I}_{(0,+\infty)}(x)$$

- $E(X) = \frac{\alpha}{\beta}$  y  $V(X) = \frac{\alpha}{\beta^2}$ .
- *Reproductividad*: Si  $X_1, \dots, X_n$  son variables aleatorias independientes con  $X_i \sim Exp(\lambda) = Ga(\alpha = 1, \beta = \lambda)$  entonces  $\sum_{i=1}^n X_i \sim Ga(\alpha = n, \beta = \lambda)$ .

Si  $X_1, \dots, X_n$  son variables aleatorias independientes con  $X_i \sim Ga(\alpha_i, \beta)$  entonces  $\sum_{i=1}^n X_i \sim Ga(\alpha = \sum_{i=1}^n \alpha_i, \beta)$ .

### 0.2.4. Distribución $\chi_{(n)}^2$

Diremos que  $X \sim \chi_{(n)}^2$  si  $X \sim Ga\left(\alpha = \frac{n}{2}, \beta = \frac{1}{2}\right)$ .

- $E(X) = n$  y  $V(X) = 2n$ .
- *Relación con otras distribuciones:*
  1.  $X \sim U(0, 1)$ , entonces  $-2\log(X) \sim \chi_{(2)}^2$ .
  2.  $X \sim N(\mu, \sigma)$ , entonces  $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_{(1)}^2$ .
  3.  $X_1, \dots, X_m$  son variables aleatorias independientes con  $X_i \sim \chi_{(n_i)}^2$ , entonces

$$\sum_{i=1}^m X_i \sim \chi_{(n)}^2, \quad n = \sum_{i=1}^m n_i$$

### 0.2.5. Distribución Beta

Diremos que  $X \sim Be(\alpha, \beta)$ ,  $\alpha, \beta > 0$ , si su función de densidad es

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \cdot \mathbf{I}_{(0,1)}(x)$$

**Observación 0.2.1.**  $X \sim Be(\alpha = 1, \beta = 1) \equiv U(0, 1)$ .

### 0.2.6. Distribución Normal

Diremos que  $X \sim N(\mu, \sigma)$ ,  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , si su función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \mathbb{R}$$

- $E(X) = \mu$  y  $V(X) = \sigma^2$ .
- La distribución es simétrica respecto de  $\mu$ .
- $E(X) = Me = Mo = \mu$ .
- Distribución  $N(0, 1)$ . Se suele denotar como  $Z \sim N(0, 1)$  y su función de densidad es

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R}$$

Si  $X \sim N(\mu, \sigma)$ , entonces  $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ .

- *Propiedades:*
  1.  $X \sim N(\mu, \sigma)$ , entonces  $Y = a + bX \sim N(a + b\mu, |b|\sigma)$ .
  2.  $X_1, \dots, X_n$  son variables aleatorias independientes con  $X_i \sim N(\mu_i, \sigma_i)$ , entonces

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

3.  $X_1, \dots, X_n$  son variables aleatorias independientes con  $X_i \sim N(\mu, \sigma)$ , entonces

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

4. Si  $X \sim N(\mu, \sigma)$ , entonces  $Y = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi_{(1)}^2$ .

### 0.3. Cambio de variables

**Teorema 0.3.1** (Teorema de cambio de variables). Sea  $X$  una variable aleatoria absolutamente continua con función de densidad  $f_X$  concentrada en un intervalo  $I \subseteq \mathbb{R}$ . Sea  $g : I \rightarrow \mathbb{R}$  una función medible, derivable, con derivada continua y estrictamente monótona. Entonces  $Y = g(x)$  es variable aleatoria absolutamente continua con función de densidad

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |(g^{-1}(y))'|.$$

Si no se verifican las condiciones del teorema de cambio de variables, entonces calculamos la función de distribución de  $Y$ .

### 0.4. Vectores aleatorios

**Definición 0.4.1.** Sea  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vector de funciones definido en  $(\Omega, \mathcal{A}, P)$  a  $\mathbb{R}^n$ , donde para cada  $\omega \in \Omega$

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$$

Diremos que  $\mathbf{X}$  es un vector aleatorio o una variable aleatoria  $n$ -dimensional si, para cada vector de números reales,  $\mathbf{a} = (a_1, a_2, \dots, a_n)$ , la imagen inversa del intervalo  $n$ -dimensional  $I = \{(x_1, \dots, x_n) : x_i \leq a_i, i = 1, 2, \dots, n\}$  pertenece a la  $\sigma$ -álgebra  $\mathcal{A}$ , es decir,

$$\mathbf{X}^{-1}(I) = \{\omega \in \Omega : X_1(\omega) \leq a_1, \dots, X_n(\omega) \leq a_n\} \in \mathcal{A}.$$

#### 0.4.1. Función de distribución conjunta

Dado un vector aleatorio  $\mathbf{X}$ , se define la función de distribución conjunta como la función  $F_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$  definida de la forma

$$F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

**Teorema 0.4.2.** La función de distribución de probabilidad de un vector aleatorio satisface las siguientes propiedades:

$$(P1) \quad F(+\infty, +\infty, \dots, +\infty) = 1.$$

$$(P2) \quad F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0.$$

(P3) La función de distribución conjunta es creciente para cada componente.

(P4) La función de distribución conjunta es continua por la derecha para cada componente.

#### 0.4.2. Distribución del máximo y del mínimo

Sea  $\mathbf{X} = (X_1, \dots, X_n)$  un vector aleatorio con función de distribución  $F_{\mathbf{X}}$  y sean las variables aleatorias  $M = \max(X_1, \dots, X_n)$  y  $N = \min(X_1, \dots, X_n)$  definidas de la forma

$$M(\omega) = \max(X_1(\omega), \dots, X_n(\omega))$$

$$N(\omega) = \min(X_1(\omega), \dots, X_n(\omega))$$

Tanto  $M$  como  $N$  son variables aleatorias y

$$F_M(x) = P[M \leq x] = P[X_1 \leq x, \dots, X_n \leq x] = F_{\mathbf{X}}(x, \dots, x)$$

$$F_N(x) = P[N \leq x] = 1 - P[N > x] = 1 - P[X_1 > x, \dots, X_n > x].$$



# Capítulo 1

## Introducción a la Inferencia

### 1.1. Tipos de inferencia

La **Estadística** es la rama de las Matemáticas que se ocupa de recoger datos, analizarlos y organizarlos, y de realizar predicciones que sobre esos datos puedan deducirse. Tiene dos vertientes básicas:

- a) **Estadística Descriptiva:** Se ocupa de, a partir de ciertos datos, analizarlos y organizarlos. Es aquí donde tiene sentido calcular la media, mediana, moda, desviación media, desviación típica, etc.
- b) **Estadística Inferencial:** Se ocupa de predecir, sacar conclusiones, para una población tomando como base una muestra de dicha población. Como todas las predicciones, siempre han de hacerse bajo un cierto grado de fiabilidad o confianza.

En un sentido amplio, se entiende por **Inferencia** a la parte de la Estadística que estudia grandes colectivos (población) a partir de una pequeña parte de ésta (muestra).

Un factor clave de diferenciación entre los diversos problemas de Inferencia Estadística. que pueden plantearse, es el tipo de conclusiones que se quieren establecer sobre la situación aleatoria en cuestión. Distinguimos entonces:

- **Estimación puntual:** Situaciones en las que el objetivo es utilizar la información suministrada por las observaciones para obtener un pronóstico numérico único acerca de un determinado parámetro de la distribución de la característica o población en estudio.
- **Estimación por intervalos:** Situaciones en las que el objetivo es proporcionar un margen de variación para un determinado parámetro de la distribución desconocida; es decir: precisar un intervalo numérico en el que pueda afirmarse. razonablemente, que varía. el parámetro en cuestión.
- **Contrastes de Hipótesis:** Tratar de corroborar o invalidar una determinada afirmación acerca de la distribución de probabilidad del fenómeno estudiado.

La población en estudio viene representada por una variable aleatoria,  $X$ , con una determinada distribución de probabilidad  $X \sim F_X$ . Dependiendo del grado de conocimiento de esta distribución, se distinguen dos métodos para realizar procesos de inferencia.

- a) **Inferencia Paramétrica:** Es aquella en la que se admite que la distribución de la población pertenece a una cierta familia paramétrica de distribuciones, siendo necesario únicamente precisar el valor de los parámetros para determinar la distribución poblacional.
  - 1) **Enfoque clásico:** Los parámetros de la distribución de la población se consideran constantes.

- 2) **Enfoque bayesiano:** Considera los parámetros de la distribución como variables aleatorias permitiendo introducir información sobre ellos a través de la distribución a priori.
- b) **Inferencia no Paramétrica:** No supone ninguna distribución de probabilidad de la población, exigiendo solo hipótesis muy generales como puede ser la de simetría de la distribución. Los procedimientos no paramétricos se pueden clasificar en:
  - 1) **Procedimientos de localización**, que estudian los parámetros de localización de la distribución.
  - 2) **Procedimientos de estructura**, que estudian las condiciones que se dan en la distribución de la variable.
  - 3) **Procedimientos sobre las condiciones de la muestra**, que comprueban si se verifican las hipótesis exigibles a los valores muestrales como pueden ser la independencia, ausencia de valores atípicos, etc.

## 1.2. Elementos de la Inferencia Estadística

Todo problema de inferencia estadística está motivado por un cierto grado de desconocimiento de la ley de probabilidad que rige cierto fenómeno aleatorio. El caso más simple es cuando el interés se centra en una determinada variable aleatoria  $X$  cuya distribución  $F$  es más o menos desconocida. La distribución  $F$  involucrada en el problema de inferencia estadística se denomina *distribución teórica* o *distribución de la población*.

Los elementos de la Inferencia Estadística son:

- **Población:** Conjunto de elementos sobre los que se observa una característica común. Esta característica viene representada por una variable aleatoria cuya distribución  $F$  es desconocida o no se conoce su totalidad.
- **Muestra:** Conjunto de unidades de una población. Cuanto más significativa sea, mejor será la muestra.
- **Parámetros poblacionales:** Son los índices centrales y de dispersión que definen a una población.

En los modelos de Inferencia Estadística el mayor o menor grado de desconocimiento acerca de la distribución teórica  $F$  se refleja mediante una familia  $\mathcal{F}$  de distribuciones. La situación más simple es aquella en la que la familia  $\mathcal{F}$  está compuesta por distribuciones que tienen una forma funcional fija y conocida y que dependen de un parámetro  $\theta$ , de una o más dimensiones, que varía en un subconjunto  $\Theta \subset \mathbb{R}^k$ , llamado **espacio paramétrico**

$$\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^k\}.$$

## 1.3. Muestreo aleatorio

Para que la muestra sea representativa debe reflejar las similitudes y diferencias encontradas en la población. Los errores más comunes que se pueden cometer son:

- 1) **Error de muestreo:** Hacer conclusiones muy generales a partir de la observación de sólo una parte de la población.
- 2) **Error de inferencia:** Hacer conclusiones hacia una población de mayor tamaño que de la que se tomó la muestra.

Los métodos para muestreo se clasifican en:

- (i) Muestreos probabilísticos.
- (ii) Muestreos no probabilísticos.

### 1.3.1. Muestreo probabilístico

Los muestreos probabilísticos se basan en el principio de equiprobabilidad: todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra.

- **Muestreo aleatoria simple con reemplazamiento:** Todas las unidades muestrales tienen la misma probabilidad de ser elegidas, pudiendo medirse varias veces el mismo individuo.
- **Muestreo aleatoria simple sin reemplazamiento:** Todas las unidades muestrales tienen la misma probabilidad de ser elegidas, pero los individuos no tienen posibilidad de aparecer en la muestra más de una vez.
- **Muestreo estratificado:** La población está dividida en subpoblaciones o estratos que contienen elementos parecidos entre sí. La composición de la muestra se distribuye entre los distintos estratos mediante un procedimiento que recibe el nombre de *afijación*.
  - a) **Afijación uniforme:** En la muestra habrá el mismo número de representantes por cada estrato. Si hay  $k$  estratos y el tamaño de la muestra es  $n$ , se extraerán aproximadamente  $n/k$  elementos de cada estrato.
  - b) **Afijación proporcional:** En la muestra habrá un número de representantes de cada estrato proporcional a su tamaño. Si el estrato  $i$  tiene  $N_i$  elementos y la población total está formada por  $N$  elementos, al estrato  $i$  le corresponden  $(N_i/N) \cdot n$  elementos muestrales, donde  $n$  es el tamaño de muestra que se quiere tomar.
  - c) La asignación de unidades muestrales se hace teniendo en cuenta tanto el **tamaño de la muestra** como su **variabilidad**. De esta forma en un estrato más heterogéneo se necesitan más unidades muestrales mientras que en uno más homogéneo se necesitan menos unidades muestrales.

Si  $\sigma_i$  representa a la desviación típica del  $i$ -ésimo estrato, la asignación de las unidades muestrales para dicho estrato será

$$n_i = n \cdot \frac{N_i \sigma_i}{\sum_{j=1}^k \sigma_j N_j}.$$

- **Muestreo por conglomerados:** Dada una población, se establecen grupos de elementos físicamente próximos entre ellos frecuentemente constituidos por una partición geográfica de la población. Los conglomerados deben ser tan heterogéneos como la población a estudiar para que la muestra la represente bien.

Obsérvese que en el muestreo estratificado un estrato es homogéneo (sus elementos tienen las mismas características), mientras que un conglomerado es heterogéneo porque debe representar a la población.

- **Muestreo sistemático:** Para obtener una muestra de tamaño  $n$  se elige un individuo al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra.

### 1.3.2. Muestreo no probabilístico

Cuando el muestreo probabilístico resulta costoso o no se tiene asegurado que cualquier elemento de la población pueda pertenecer a una muestra, se consideran muestreos no probabilísticos. Estos métodos no sirven para realizar generalizaciones pues no se tiene la certeza de que la muestra sea representativa.

En este tipo de muestreo se da preferencia en la muestra a aquellos individuos que al cumplir con cierta cualidad o característica benefician a la investigación.

- **Muestreo por cuotas:** Se supone que existe un buen conocimiento de los estratos de la población y/o los individuos más representativos o adecuados para la investigación. Tiene semejanzas con el muestreo estratificado pero no tiene el carácter de aleatoriedad.

- **Muestreo intencional o de conveniencia:** Este muestreo se caracteriza por intentar obtener muestras representativas mediante la inclusión en la muestra de grupos típicos. Es utilizado en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.
- **Bola de nieve:** Se localiza a algunos individuos que a su vez conducen a otros, y estos a otros, y así sucesivamente hasta conseguir una muestra suficiente. Se usa cuando se hacen estudios en poblaciones marginales, delincuentes, sectas, determinados tipos de enfermos, etc.

## 1.4. Muestra aleatoria simple

**Definición 1.4.1.** Una **muestra aleatoria simple de tamaño  $n$**  de una variable aleatoria  $X$  con distribución teórica  $F_X$ , es un conjunto de  $n$  variables aleatorias independientes  $(X_1, \dots, X_n)$  e igualmente distribuidas con distribución común  $F_X$ .

Cada  $X_i$  es una variable aleatoria que representa la característica bajo estudio del elemento  $i$ -ésimo de la muestra. Cuando las mediciones se hayan llevado a cabo, es decir, una vez realizado el muestreo, los resultados obtenidos,  $(x_1, \dots, x_n)$ , se denominarán **realización de la muestra**.

Como las variables aleatorias han de ser independientes, la función de distribución conjunta de la muestra aleatoria simple será

$$F(x_1, \dots, x_n) = F_X(x_1) \cdots F_X(x_n).$$

No habrá ambigüedad en representar mediante el mismo símbolo,  $F$ , la función de distribución conjunta y las marginales puesto que el número de variables las distingue.

## 1.5. Concepto de estadístico y su distribución en el muestreo

**Definición 1.5.1.** Llamamos **estadístico** a cualquier función medible de las variables aleatorias observables de la muestra aleatoria simple que no depende de ningún parámetro desconocido en el estudio.

En general un estadístico es una función medible  $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$  aplicada a la muestra aleatoria simple,  $T(X_1, \dots, X_n)$ .

Un estadístico  $T$  es una nueva variable aleatoria (o vector aleatorio) que tendrá una distribución asociada llamada *distribución muestral* con una función de distribución llamada *función de distribución empírica*.

### 1.5.1. Estadísticos más utilizados

Los estadísticos más utilizados son:

- **Media muestral**

$$T(X_1, X_2, \dots, X_n) = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

- **Varianza muestral**

$$T(X_1, X_2, \dots, X_n) = \text{Var}_n(\mathbf{X}) = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

■ **Cuasivarianza muestral**

$$T(X_1, X_2, \dots, X_n) = S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

■ **Máximo y Mínimo**

$$\begin{aligned} T(X_1, X_2, \dots, X_N) &= X_{(n)} = \max(X_1, X_2, \dots, X_n) \\ T(X_1, X_2, \dots, X_N) &= X_{(1)} = \min(X_1, X_2, \dots, X_n) \end{aligned}$$

**Observación 1.5.2.**

$$S^2 = \frac{n}{n-1} \text{Var}_n(\mathbf{X})$$

### 1.5.2. Función de distribución muestral

Sea  $X \sim F_\theta$ , donde  $\{F_\theta : \theta \in \Theta\}$ . Identifiquemos  $F_\theta \equiv F$ . ¿Cómo podemos estimar  $F$ ?

Tomamos  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple ( $X_i \sim F$ ,  $i = 1, \dots, n$  e independientes). Definimos

$$F_{(X_1, X_2, \dots, X_n)}^*(x) = \frac{\text{número de variables tales que } X_i \leq x}{n}, \quad x \in \mathbb{R},$$

que es una nueva variable aleatoria. Observamos que podemos escribir dicha variable aleatoria de la siguiente forma

$$F_{(X_1, X_2, \dots, X_n)}^*(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i).$$

Si llamamos

$$Y_i = I_{(-\infty, x]}(X_i) = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases}$$

es una variable aleatoria e  $Y_i \sim \text{Ber}(p)$ . Encontremos dicho  $p$ . Observamos que

$$P(Y_i = 1) = P(X_i \leq x) = F_{X_i}(x) = F(x).$$

Entonces

$$F_{(X_1, X_2, \dots, X_n)}^*(x) = \frac{1}{n} \sum_{i=1}^n Y_i,$$

como  $Y_i$  son variables aleatorias independientes, entonces

$$nF_{(X_1, X_2, \dots, X_n)}^*(x) = \sum_{i=1}^n Y_i \sim \text{Bi}(n, F(x)).$$

De aquí deducimos que

$$\begin{aligned} E(F_{(X_1, X_2, \dots, X_n)}^*(x)) &= \frac{1}{n} nF(x) = F(x) \\ V(F_{(X_1, X_2, \dots, X_n)}^*(x)) &= \frac{1}{n^2} nF(x)(1 - F(x)) = \frac{1}{n} F(x)(1 - F(x)) \end{aligned}$$

Además, por el Teorema Central del Límite

$$F_{(X_1, X_2, \dots, X_n)}^*(x) \sim N\left(F(x), \sqrt{\frac{F(x)(1 - F(x))}{n}}\right)$$

### 1.5.3. Momentos centrales y no centrales

Sea  $X \sim F_X$  y sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria simple ( $X_i \sim F$ ,  $i = 1, \dots, n$  e independientes). Entonces

- **Momento ordinario de orden k respecto del origen**

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

- **Momento central de orden k**

$$M_k = \frac{\sum_{i=1}^n (X_i - \bar{X})^k}{n}$$

## Capítulo 2

# Distribución Normal. Distribuciones asociadas al muestreo de poblaciones normales

**Definición 2.0.1.** Sea  $\mathbf{X} = (X_1, X_2, \dots, X_n)^t$  un vector aleatorio y sea  $\boldsymbol{\mu} = (E(X_1), E(X_2), \dots, E(X_n))^t$  el vector de medias. Se define la matriz de varianzas y covarianzas  $\Sigma$ , como

$$\Sigma = E((\mathbf{X} - \boldsymbol{\mu}) \cdot (\mathbf{X} - \boldsymbol{\mu})^t)$$

### 2.1. Distribución Normal Bivalente

**Definición 2.1.1.** Se dice que  $(X_1, X_2)^t$  tiene una distribución **Normal Bivalente** con vector de medias  $\boldsymbol{\mu} = (\mu_1, \mu_2)^t$  y matriz de covarianzas

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

siendo  $\rho$  el coeficiente de correlación entre  $X_1$  y  $X_2$  y lo notaremos  $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ , si

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \sigma_2^2 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

donde  $Z_1$  y  $Z_2$  son variables aleatorias independientes distribuidas según una  $N(0, 1)$ .

Si  $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$ , su función de densidad conjunta viene dada de la forma

$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^t} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1, x_2)}, \quad \mathbf{x} \in \mathbb{R}^2$$

donde  $Q(x_1, x_2)$  es la forma cuadrática

$$Q(x_1, x_2) = \frac{1}{1-\rho^2} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} \right)$$

**Observación 2.1.2.** Recordemos que el coeficiente de correlación entre  $X_1$  y  $X_2$  es

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)}\sqrt{V(X_2)}}$$

Si  $X_1$  y  $X_2$  son independientes entonces  $\rho = 0$ . Sin embargo, el recíproco no es cierto en general, aunque si  $(X_1, X_2) \sim N_2(\boldsymbol{\mu}, \Sigma)$  entonces  $X_1$  y  $X_2$  son independientes si y solo si  $\rho = 0$ .

## 2.2. Distribución $\chi_{(n)}^2$

**Definición 2.2.1.** Una variable aleatoria  $X$  decimos que se distribuye según una  $\chi_{(n)}^2$  cuando  $X \sim Ga(\alpha = \frac{n}{2}, \beta = \frac{1}{2})$ , con  $n \in \mathbb{N}$ . La función de densidad viene dada de la forma

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1}, \quad x > 0.$$

### Génesis de la distribución $\chi_{(n)}^2$

Dadas  $X_1, X_2, \dots, X_n$  variables aleatorias independientes con distribución  $N(0, 1)$ , el vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)^t \sim N_n(\mathbf{0}, \mathbf{I}_n)$ . Si consideramos la variable aleatoria  $Y = \sum_{i=1}^n X_i^2$  podemos decir que  $Y \sim \chi_{(n)}^2$ .

Por lo tanto podemos que una  $\chi_{(n)}^2$  viene dada como la suma de los cuadrados de  $n$  variables aleatorias independientes todas  $N(0, 1)$ .

Esta distribución depende de un sólo parámetro,  $n$ , cuya denominación de *grados de libertad* hace referencia al número de sumandos que aportan variabilidad a la suma.

### Parámetros de la distribución $\chi_{(n)}^2$

Si  $Y \sim \chi_{(n)}^2$ , la esperanza es  $E(Y) = n$  y la varianza es  $V(Y) = 2n$ .

### Aproximación por el Teorema Central del Límite

Si  $Y \sim \chi_{(n)}^2$ , mediante el Teorema Central del Límite se tiene que

$$\begin{aligned} \frac{Y - n}{\sqrt{2n}} &\rightarrow N(0, 1) \\ \sqrt{2Y} - \sqrt{2n - 1} &\rightarrow N(0, 1) \end{aligned}$$

### Reproductividad

Si  $T \sim \chi_{(n)}^2$  y  $W \sim \chi_{(m)}^2$  son variables aleatorias independientes, entonces  $T + W \sim \chi_{(n+m)}^2$

### Distribución de la forma cuadrática de la distribución Normal Multivariante

Si  $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ . Entonces la formma cuadrática

$$(\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})^t \sim \chi_{(n)}^2.$$

### Teorema de Fisher

**Teorema 2.2.2** (Teorema de Fisher). Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población  $N(0, 1)$ , entonces  $(n - 1)S^2 (= \sum_{i=1}^n (X_i - \bar{X})^2)$  y  $\bar{X}$  son variables aleatorias independientes y la distribución del muestreo es

$$\begin{aligned} (n - 1)S^2 &\sim \chi_{(n-1)}^2 \\ \bar{X} &\sim N\left(0, \frac{1}{\sqrt{n}}\right) \end{aligned}$$



**Teorema 2.2.3** (Teorema de Fisher generalizado). Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población  $N(\mu, \sigma)$ , entonces  $S^2$  y  $\bar{X}$  son variables aleatorias independientes y la distribución del muestreo es

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

## 2.3. Distribución t-Student

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de una población  $N(\mu, \sigma)$ , entonces la distribución en el muestreo de  $\bar{X}$  es  $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$  o equivalentemente

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ tiene como distribución } N(0, 1).$$

No será de gran utilidad esta distribución si  $\sigma$  es desconocido. De hecho no podremos utilizarla para dar predicciones acerca de la diferencia  $\bar{X} - \mu$ . Podríamos sustituir  $\sigma$  por algún estadístico que sirva para estimar valores de  $\sigma$ . Parece razonable sustituir  $\sigma$  por  $S$ , donde  $S^2$  es la cuasivarianza muestral. Esta idea llevó a Student (Gosset) a considerar el estadístico

$$t = \sqrt{n-1} \frac{\bar{X} - \mu}{S}$$

La cuestión ahora es determinar la distribución de este nuevo estadístico y confiar en que no dependa de  $\sigma$ . En caso contrario no lo podremos usar como estadístico.

**Definición 2.3.1.** Si  $X, X_1, X_2, \dots, X_n$ ,  $n+1$  variables aleatorias independientes con distribución  $N(0, \sigma)$ . Sea  $Y = \sum_{i=1}^n X_i^2$  y  $U = \sqrt{Y/n}$ . Entonces

$$t = \frac{X}{U} = \frac{X}{\sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}}$$

La distribución de  $t$  se denomina **distribución t-Student con  $n$  grados de libertad** y lo notaremos  $t \sim t_n$ .

La distribución  $t_n$  es la distribución  $N(0, 1)$  dividida por la raíz cuadrada de una  $\chi_{(n)}^2$  por sus grados de libertad, ambas independientes.

### Parámetros y soporte de la distribución t-Student

La distribución  $t$  de Student depende de un único parámetro  $n$  (el número de sumandos que intervienen en la suma del denominador). La distribución tiene como soporte  $\mathbb{R}$ . Su densidad es simétrica respecto al origen con gráfica parecida a la distribución.

### Función de distribución y características numéricas

- Para  $n = 1$ , la distribución  $t_1$  es la distribución de Cauchy con densidad

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

Esta distribución no posee momentos de primer orden y por consiguiente carece de varianza.

- Si  $n > 1$  la media es finita y vale cero. Para  $n > 2$  la varianza existe y es  $\frac{n}{n-2}$ , de forma que la dispersión decrece rápidamente hacia 1 cuando  $n$  crece.

La función de distribución asociada a la densidad de  $t_n$  tampoco puede expresarse explícitamente y sus valores numéricos aparecen tabulados.

### Aproximación a la $N(0, 1)$

Cuando  $n \rightarrow +\infty$ , la distribución  $t_n$  se puede aproximar por una  $N(0, 1)$ .

## 2.4. Distribución F-Snedecor

Consideremos dos poblaciones normales independientes  $X \sim N(\mu_1, \sigma_1)$  e  $Y \sim N(\mu_2, \sigma_2)$  que queremos comparar. Supuesto que las varianzas son desconocidas, la comparación entre las poblaciones exige también la comparación de las varianzas. El método para obtener información acerca de la relación estará basado en las cuasivarianzas respectivas  $S_1^2$  y  $S_2^2$ . En vez de considerar  $S_1^2 - S_2^2$ , de la que no es posible encontrar la distribución de forma fácil, utilizaremos el estadístico  $\frac{S_1^2}{S_2^2}$  cuya distribución en el muestreo se puede calcular de forma explícita.

**Definición 2.4.1.** Si  $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$  son variables aleatorias independientes con distribución  $N(0, 1)$ , entonces la distribución del estadístico

$$\frac{\sum_{i=1}^n \frac{X_i^2}{n}}{\sum_{j=1}^m \frac{Y_j^2}{m}}$$

es una distribución **F-Snedecor con  $n$  y  $m$  grados de libertad** y lo notaremos  $F_{n,m}$ .

Una distribución  $F_{n,m}$  es la distribución del cociente de dos distribuciones  $\chi^2$  independientes, de  $n$  y  $m$  grados de libertad respectivamente, divididas cada una de ellas por sus grados de libertad.

### Características numéricas

La media de esta distribución es  $\frac{m}{m-2}$  si  $m > 2$  y su varianza existe si  $m > 4$ .

## Capítulo 3

# Estimación puntual paramétrica

### 3.1. Estadísticos y estimadores. Métodos para la construcción de estimadores

**Definición 3.1.1.** Sea  $X \sim \{F_\theta : \theta \in \Theta\}$  y sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple. Un **estadístico** es una función medible,  $T : (X_1, \dots, X_n) \rightarrow \mathbb{R}^k$ , que no depende de ningún parámetro desconocido de la distribución.

**Definición 3.1.2.** Un **estimador** es un estadístico,  $T$ , definido de la forma

$$T : (X_1, \dots, X_n) \rightarrow \Theta$$

**Ejemplo 3.1.3.** Sea  $X \sim \text{Ber}(p)$  y sea  $(X_1, X_2, X_3)$  una muestra aleatoria simple.

- Tenemos que

$$T(X_1, X_2, X_3) = \sum_{i=1}^3 X_i$$

es un estadístico para  $p$ . Además es un estimador para el número de veces que se alcanza el éxito en una muestra de tamaño 3.

- Tenemos que

$$T(X_1, X_2, X_3) = \frac{\sum_{i=1}^3 X_i}{3}$$

es un estimador para  $p$ .

#### 3.1.1. Método de los momentos

**Definición 3.1.4.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X$  que se distribuye con  $F(x; \vec{\theta})$ , siendo  $\vec{\theta} = (\theta_1, \dots, \theta_k)$  el vector de parámetros que caracteriza la distribución. Si existen los momentos ordinarios hasta orden  $k$  de  $F(x; \vec{\theta})$ , podemos encontrar un estimador para  $\vec{\theta}$  igualando los momentos poblacionales y muestrales.

**Observación 3.1.5.** Si  $\vec{\theta} = (\theta_1, \dots, \theta_k)$  es constante no podemos aplicar el método de los momentos para estimar  $\vec{\theta}$ .

**Ejemplo 3.1.6.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria de una variable aleatoria  $X \sim Po(\lambda)$ . Entonces  $\theta = \lambda > 0$  y  $\Theta = (0, +\infty)$ .

Sabemos que  $E_\lambda(X) = \lambda$  y

$$m_1 = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Usando este estimador para calcular  $\lambda$  tenemos que

$$E_\lambda(T) = \frac{1}{n} E_\lambda \left( \sum_{i=1}^n X_i \right) = \frac{1}{n} \cdot n\lambda = \lambda$$

### 3.1.2. Método de la máxima verosimilitud

**Definición 3.1.7.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X$  cuya distribución pertenece a  $\{F_\theta : \theta \in \Theta\}$ . Llamamos **función de verosimilitud** a la función de densidad conjunta o distribución de probabilidad conjunta considerada como función de  $\theta$  para valores fijos de la muestra.

- $L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$  para variables aleatorias absolutamente continuas.
- $L(\theta; x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n; \theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$  para variables aleatorias discretas.

**Ejemplo 3.1.8.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X \sim U(0, \theta)$ . Entonces

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^n I_{[\max\{x_i\}, \infty)}(\theta)$$

**Definición 3.1.9.** Sea  $X \sim \{F_\theta : \theta \in \Theta\}$ . Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple. Decimos que  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = T = T(X_1, \dots, X_n)$  es un **estimador de máxima verosimilitud** para  $\theta$  si

$$\hat{\theta} = \max_{\theta \in \Theta} L(\theta; x_1, \dots, x_n)$$

para toda realización de muestra  $(x_1, \dots, x_n)$  y siempre que dicha exista.

**Ejemplo 3.1.10.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X \sim Ber(\theta)$ , donde  $\theta \in \Theta = [0, 1]$ . Sabemos que

$$P_\theta(X_i = x_i) = \theta^{x_i} (1 - \theta)^{1-x_i}, \quad x_i = 0, 1$$

Consideramos una relación de la muestra  $(x_1, \dots, x_n)$  ( $x_i = 0, 1$ ). La función de verosimilitud es

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n P_\theta(X_i = x_i) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$$

Consideramos ahora la media muestral, es decir,

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad \text{que es un valor en } [0, 1]$$

Entonces podemos escribir la función de verosimilitud como

$$L(\theta) = \theta^{\bar{x}} (1 - \theta)^{n - n\bar{x}}, \quad \theta \in [0, 1]$$

Tomando logaritmos

$$\log(L(\theta)) = n\bar{x} \log(\theta) + (n - n\bar{x}) \log(1 - \theta)$$

$$l(\theta; x_1, \dots, x_n) = \log(L(\theta; x_1, \dots, x_n)) \text{ se llama log-verosimilitud.}$$

Entonces

$$\begin{aligned} l(\theta) &= n\bar{x} \log(\theta) + (n - n\bar{x}) \log(1 - \theta) \\ l'(\theta) &= \frac{n\bar{x}}{\theta} - \frac{n - n\bar{x}}{1 - \theta} \end{aligned}$$

Observamos que

$$\begin{aligned} l'(\theta) = 0 &\iff \frac{n\bar{x}}{\theta} - \frac{n - n\bar{x}}{1 - \theta} = 0 \\ &\iff n\bar{x} - n\bar{x}\theta - n\theta + n\bar{x}\theta = 0 \\ &\iff n\bar{x} = n\theta; \quad \bar{x} = \hat{\theta} \end{aligned}$$

Luego en  $\hat{\theta} = \bar{x}$  hay un punto crítico de  $l$  (faltaría comprobar que  $l''(\hat{\theta}) < 0$ ).

Conclusión: Para la realización de una muestra,  $(x_1, \dots, x_n)$ , el valor de  $\theta$  donde se alcanza el máximo de la función de verosimilitud es

$$\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

En general, sea cual sea la realización de la muestra

$$\hat{\theta} = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

**Ejemplo 3.1.11.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X \sim U(0, \theta)$ . Sea  $(x_1, \dots, x_n) \in \mathfrak{X}$  una realización de la muestra aleatoria simple.

$$\mathfrak{X} = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \in (0, \theta), i = 1, \dots, n\}$$

La función de verosimilitud es por tanto

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{1}{\theta} I_{(0, \theta)}(x_i) = \left(\frac{1}{\theta}\right)^n I_{(\max\{x_i\}, \infty)}(\theta)$$

Observamos que  $\hat{\theta} = \max\{x_i\} > 0$  y el estimador de máxima verosimilitud para  $\theta$  es  $\hat{\theta} = \max\{x_i\}$ .

**Ejemplo 3.1.12.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu, \sigma^2)$ . Definimos  $\theta = (\mu, \sigma^2)$  y tenemos que  $\theta \in \Theta = \mathbb{R} \times (0, +\infty)$ . Sea  $(x_1, \dots, x_n)$  una realización de la muestra. La función de verosimilitud es

$$\begin{aligned} L(\mu, \sigma^2; x_1, \dots, x_n) &= f(x_1, \dots, x_n; \mu, \sigma^2) = \prod_{i=1}^n f(x_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \end{aligned}$$

Con esto, tenemos que

$$l(\mu, \sigma^2) = -n \log(\sqrt{2\pi}) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Para calcular el estimador de máxima verosimilitud, veamos cuando ocurre que

$$(1) \frac{\partial l(\mu, \sigma^2)}{\partial \mu} = 0 \quad \text{y} \quad (2) \frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} = 0$$

Recordamos que  $s = \sigma^2$  ( $\sqrt{s} = \sigma$ )

$$(1) \frac{\partial l(\mu, \sigma^2)}{\partial \mu} = 0 \iff \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \iff \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

$$(2) \frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} = 0 \iff -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \iff \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

Por tanto, los estimadores de máxima verosimilitud para  $\mu$  y  $\sigma^2$  son

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{y} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

**Observación 3.1.13.** En general, dada  $(X_1, \dots, X_n)$ , el estimador de máxima verosimilitud de  $\mu$  y  $\sigma^2$  es

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{y} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \text{Var}_n(\vec{X})$$

El estimador de máxima verosimilitud para  $\mu$  es insesgado, es decir

$$E(\hat{\mu}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \mu$$

El estimador de máxima verosimilitud para  $\sigma^2$  no es insesgado, es decir

$$E(\hat{\sigma}^2) = E\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right) \neq \sigma^2$$

**Teorema 3.1.14** (Teorema de invarianza). Sea  $X \sim f(x, \theta)$  con  $\theta$  desconocido y sea  $\hat{\theta}$  el estimador de máxima verosimilitud de  $\theta$ . Sea  $\tau : \Theta \rightarrow \Lambda$ . Si  $\tau$  es biyectiva, entonces el estimador de máxima verosimilitud para  $\tau(\theta)$  es  $\tau(\hat{\theta})$ .

*Demostración.* Esta demostración es cortesía de Miguel Ángel Quesada Martínez.

Definimos  $\rho = \tau(\theta)$ , entonces  $\theta = \tau^{-1}(\rho)$ . Reparametrizemos la verosimilitud usando el nuevo parámetro  $\rho$  en lugar de  $\theta$ .

$$L^*(\rho; x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \tau^{-1}(\rho)) = L(\tau^{-1}(\rho); x_1, \dots, x_n)$$

Con esto, tenemos que

$$\begin{aligned} L^*(\hat{\rho}; x_1, \dots, x_n) &= \sup_{\rho} L^*(\rho; x_1, \dots, x_n) = \sup_{\rho} L(\tau^{-1}(\rho); x_1, \dots, x_n) \\ &= \sup_{\theta} L(\theta; x_1, \dots, x_n) = L(\hat{\theta}; x_1, \dots, x_n) = L^*(\tau(\hat{\theta}); x_1, \dots, x_n) \end{aligned}$$

Por tanto, tenemos que el máximo de  $L^*(\rho; x_1, \dots, x_n)$  se alcanza cuando  $\hat{\rho} = \tau(\hat{\theta})$ , luego el estimador de máxima verosimilitud para  $\rho = \tau(\theta)$  es  $\tau(\hat{\theta})$ .  $\square$

**Ejemplo 3.1.15.** Sea  $X \sim \text{Exp}(\theta)$ ,  $\theta > 0$ . Calculemos el estimador de máxima verosimilitud para  $P(X \leq t^*)$ ,  $t^*$  fijo. Para ellos vamos a aplicar el Teorema de la invarianza. Obtendremos primero el estimador de máxima verosimilitud para  $\theta$  y luego para  $F_X(t^*)$  ya que

$$F_X(t^*) = P(X \leq t^*) = 1 - e^{-\theta t^*}$$

Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de  $X$  y  $(x_1, \dots, x_n)$  una realización de la muestra. La función de verosimilitud es

$$L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

Entonces

$$\begin{aligned} l(\theta) &= n \log(\theta) - \theta \sum_{i=1}^n x_i \\ l'(\theta) &= \frac{n}{\theta} - \sum_{i=1}^n x_i \\ l'(\theta) = 0 &\iff \frac{n}{\theta} - \sum_{i=1}^n x_i = 0 \iff \hat{\theta} = \frac{n}{\sum_{i=1}^n x_i} \end{aligned}$$

Observamos que

$$l''(\theta) = -\frac{n}{\theta^2} < 0 \iff l''(\hat{\theta}) < 0$$

El estimador de máxima verosimilitud para  $\theta$  es entonces

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n x_i}$$

Por lo tanto, el estimador de máxima verosimilitud para  $F_X(t^*)$  será  $T = 1 - e^{\hat{\theta} t^*}$

## 3.2. Propiedades de los estimadores

**Definición 3.2.1.** Sean  $T_1$  y  $T_2$  dos estimadores de  $g(\theta)$ . Diremos que  $T_1$  es más concentrado que  $T_2$  si y solo si

$$P_{\theta}(g(\theta) - \lambda \leq T_1 \leq g(\theta) + \lambda) \geq P_{\theta}(g(\theta) - \lambda \leq T_2 \leq g(\theta) + \lambda)$$

para todo  $\lambda > 0$  y para todo  $\theta \in \Theta$ .

Un estimador  $T$  de  $g(\theta)$  diremos que es el más concentrado si y solo si  $T$  es más concentrado que  $T'$ , para todo  $T'$  estimador de  $g(\theta)$ .

**Definición 3.2.2.** Sea  $T$  un estimador de  $g(\theta)$ . Llamaremos **error cuadrático medio del estimador  $T$** , a

$$ECM_T(\theta) = MSE_T(\theta) = E_{\theta}((T - g(\theta))^2)$$

Dados  $T_1$  y  $T_2$  estimadores de  $g(\theta)$ ,  $T_1$  es preferible a  $T_2$  si

$$ECM_{T_1}(\theta) \leq ECM_{T_2}(\theta)$$

**Observación 3.2.3.** El estimador  $S^2$  es preferible al estimador  $\text{Var}_n(X)$ .

**Proposición 3.2.4.** El error cuadrático medio se puede escribir de la forma

$$ECM_T(\theta) = Var_\theta(T) + (E_\theta(T) - g(\theta))^2$$

*Demostración.*

$$\begin{aligned} ECM_T(\theta) &= E_\theta((T - g(\theta))^2) = E_\theta((T - E_\theta(T) + E_\theta(T) - g(\theta))^2) \\ &= E_\theta((T - E_\theta(T))^2) + E_\theta((E_\theta(T) - g(\theta))^2) + 2E_\theta((T - E_\theta(T))(E_\theta(T) - g(\theta))) \\ &= Var_\theta(T) + (E_\theta(T) - g(\theta))^2 \end{aligned}$$

Observamos que en la última igualdad usamos que  $2E_\theta((T - E_\theta(T))(E_\theta(T) - g(\theta))) = 0$  puesto que  $(E_\theta(T) - g(\theta))$  es una constante con respecto a la esperanza, luego

$$\begin{aligned} 2E_\theta((T - E_\theta(T))(E_\theta(T) - g(\theta))) &= 2(E_\theta(T) - g(\theta))E_\theta((T - E_\theta(T))) \\ &= 2(E_\theta(T) - g(\theta))(E_\theta(T) - E_\theta(T)) = 0 \end{aligned}$$

□

**Definición 3.2.5.** ■ **Precisión del estimador T** :  $Var_\theta(T) = E_\theta((T - E_\theta(T))^2)$ .

■ **Sesgo del estimador T**:  $b_\theta(T) = E_\theta(T) - g(\theta)$ .

### 3.2.1. Estimadores insesgados o centrados

**Definición 3.2.6.** Diremos un estimador  $T$  de  $g(\theta)$  es insesgado si y solo si  $E_\theta(T) = g(\theta)$  para todo  $\theta \in \Theta$ .

**Observación 3.2.7.** Si  $T$  es estimador insesgado de  $g(\theta)$ , entonces  $ECM_T(\theta) = Var_\theta(T)$  para todo  $\theta \in \Theta$ .

**Ejemplo 3.2.8.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria

$$X \sim \{F_\theta : \theta \in \Theta\}$$

¿Es posible encontrar un estimador insesgado para  $\alpha_k = E(X^k)$ ? Sea  $\theta = E(X^k)$ . Consideremos el estimador

$$T(X_1, \dots, X_n) = m_k = \frac{\sum_{i=1}^n X_i^k}{n}$$

Observamos que

$$E_\theta(T) = E_\theta\left(\frac{\sum_{i=1}^n X_i^k}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^n X_i^k\right) = \frac{1}{n} \cdot n \cdot E(X^k) = E(X^k)$$

**Ejemplo 3.2.9.** Sea  $X_1$  una muestra aleatoria simple de una variable aleatoria  $X \sim Ber(\theta)$ ,  $\theta \in (0, 1)$ . ¿Es posible encontrar un estimador insesgado para  $\theta$  a partir de la muestra de tamaño 1?

Sea  $T : \mathfrak{X} \rightarrow (0, 1)$ ,  $T = X_1$  Entonces

- $P(T = 0) = P(X_1 = 0) = P(X = 0) = 1 - \theta$ .
- $P(T = 1) = P(X_1 = 1) = P(X = 1) = \theta$ .
- $E_\theta(T) = 0 \cdot P(T = 0) + 1 \cdot P(T = 1) = 0 \cdot (1 - \theta) + \theta = \theta$ .



**Ejemplo 3.2.10.** Sea  $X_1$  una muestra aleatoria simple de una variable aleatoria  $X \sim \text{Ber}(\theta)$ ,  $\theta \in (0, 1)$ . ¿Es posible encontrar un estimador insesgado para  $\theta^2 = g(\theta)$  a partir de la muestra de tamaño 1?

Sea  $T : \mathfrak{X} \rightarrow (0, 1)$ ,  $T = X_1^2$ . Entonces

- $P(T = 0) = P(X_1^2 = 0) = P(X_1 = 0) = 1 - \theta$ .
- $P(T = 1) = P(X_1^2 = 1) = P(X_1 = 1) = \theta$ .
- $E_\theta(T) = 0 \cdot P(T = 0) + 1 \cdot P(T = 1) = 0 \cdot (1 - \theta) + \theta = \theta \neq \theta^2$ .

¿Cómo definimos  $T$  para que sea un estimador insesgado para  $\theta^2$ ? Para que ocurra que

$$E_\theta(T) = T(0)(1 - \theta) + T(1)\theta = \theta^2$$

Algo razonable es que  $T(1) = \theta$ , pero esto no puede ser porque  $T$  es estimador, y en su expresión no puede algo que dependa de los que queremos estimar.

Con esto hemos demostrado que

- Si  $T$  es insesgado para  $\theta$ , entonces  $g(T)$  no tiene por qué ser insesgado para  $g(\theta)$ .
- No siempre es posible encontrar un estimador para  $g(\theta)$ .

**Ejemplo 3.2.11.** Sea  $X_1$  una muestra aleatoria simple de una variable aleatoria  $X \sim \text{Po}(\theta)$ ,  $\theta > 0$ . ¿Es posible encontrar un estimador insesgado para  $e^{-3\theta} = g(\theta)$  a partir de la muestra de tamaño 1? Observamos que

- $E_\theta(X) = \theta$  y  $V_\theta(X) = \theta$ .
- $\Theta = (0, +\infty)$  y  $\mathfrak{X} = \{0, 1, 2, \dots\}$ .
- La distribución de probabilidad de  $X$  es

$$P_\theta(X = i) = e^{-\theta} \frac{\theta^i}{i!}, \quad i \in \{0, 1, 2, \dots\}$$

Queremos  $T$  estimador insesgado para  $g(\theta) = e^{-3\theta}$ , es decir, que  $E_\theta(T) = e^{-3\theta}$ .

$$E_\theta(T) = \sum_{i=0}^{\infty} T(i)P(X_1 = i) = \sum_{i=1}^{\infty} T(i)e^{-\theta} \frac{\theta^i}{i!}$$

Observamos que

$$\sum_{i=1}^{\infty} T(i)e^{-\theta} \frac{\theta^i}{i!} = e^{-3\theta} \iff \sum_{i=1}^{\infty} T(i) \frac{\theta^i}{i!} = e^{-2\theta} = \sum_{i=1}^{\infty} \frac{(-2\theta)^i}{i!}$$

Luego, podemos tomar  $T(i) = (-2)^i$ . Pero,  $T(1) = -2 < 0$ , por tanto, no podemos encontrar un estimador insesgado.

### 3.2.2. Estimadores consistentes

**Definición 3.2.12.** Sea  $\{T_n = T(X_1, \dots, X_n)\}$  una sucesión de estimadores. Se dice que es **consistente débil** para estimar una función  $g(\theta)$  del parámetro poblacional si

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| > \varepsilon) = 0$$

para todo  $\varepsilon > 0$  y para todo  $\theta \in \Theta$ .

En esta definición estamos utilizando la convergencia en probabilidad,  $P_\theta$ , de la sucesión de estimadores  $\{T_n\}$  hacia la función del parámetro  $g(\theta)$ .

**Definición 3.2.13.** Sea  $\{T_n = T(X_1, \dots, X_n)\}$  una sucesión de estimadores de  $g(\theta)$ . Diremos que  $\{T_n\}$  es **consistente en media cuadrática** si y solo si

$$\lim_{n \rightarrow \infty} ECM_{T_n}(\theta) = \lim_{n \rightarrow \infty} E_\theta((T_n - g(\theta))^2) = 0$$

para todo  $\theta \in \Theta$ .

**Observación 3.2.14.** La propiedad de consistencia en media cuadrática implica que

$$\lim_{n \rightarrow \infty} V_\theta(T_n) = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} b_\theta(T_n) = 0$$

**Lema 3.2.15.** Si  $\{T_n\}$ , con  $T_n = T(X_1, \dots, X_n)$ , es una sucesión de estimadores consistente en media cuadrática de  $g(\theta)$ , entonces  $\{T_n\}$  es un estimador consistente débil de  $g(\theta)$ .

*Demostración.* Tenemos que ver que

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| > \varepsilon) = 0 \iff \lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| \leq \varepsilon) = 1$$

Observamos que, por la desigualdad de Chebychev

$$P_\theta(|T_n - g(\theta)| \leq \varepsilon) = P_\theta(-\varepsilon + g(\theta) \leq T_n \leq \varepsilon + g(\theta)) \geq 1 - \frac{E((T_n - g(\theta))^2)}{\varepsilon^2}$$

Como  $\{T_n\}$  es consistente en media cuadrática de  $g(\theta)$ , se tiene que

$$\lim_{n \rightarrow \infty} ECM_{T_n}(\theta) = \lim_{n \rightarrow \infty} E_\theta((T_n - g(\theta))^2) = 0$$

Por tanto,

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| \leq \varepsilon) \geq \lim_{n \rightarrow \infty} \left(1 - \frac{E((T_n - g(\theta))^2)}{\varepsilon^2}\right) = 1$$

□

### 3.3. Estadísticos suficientes

A lo largo del curso estudiaremos dos principios de reducción de datos:

- El **principio de verosimilitud**, que describe una función del parámetro  $\theta$  que contiene toda la información sobre ese parámetro a partir de unos valores de muestra fijos y que proporcionan estimadores de máxima verosimilitud.
- El **principio de suficiencia**, que es un método de reducción de datos pero sin abandonar la información que se tenga sobre  $\theta$  y que proporciona estadísticos suficientes.

#### 3.3.1. El principio de suficiencia

**Definición 3.3.1** (Suficiente I). Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim \{F_\theta : \theta \in \Theta\}$  ( $\theta$  unidimensional). Diremos que un estadístico  $T = T(X_1, \dots, X_n)$  es **suficiente** para  $\theta$  si y solo si la distribución de  $(X_1, \dots, X_n)$  dado que  $T = t$  no depende de  $\theta$  para cualquier valor  $t$  de  $T$ .

**Definición 3.3.2** (Suficiente II). Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim \{F_\theta : \theta \in \Theta\}$  ( $\theta$  unidimensional). Diremos que un estadístico  $T = T(X_1, \dots, X_n)$  es **suficiente** si y solo si la distribución condicionada de  $S|T = t$  no depende de  $\theta$ , para todo estadístico  $S$ .

**Proposición 3.3.3.** *Suficiente I  $\iff$  Suficiente II.*

**Ejemplo 3.3.4.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una variable aleatoria  $X \sim \text{Ber}(\theta)$ . Consideremos el estadístico (que veremos que es suficiente)  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$ . Calculemos la distribución de  $(X_1, \dots, X_n)$  dado que  $T = t$ .

- Si  $x_1 + \dots + x_n \neq t$ , entonces

$$P(X_1 = x_1, \dots, X_n = x_n | T = t) = \frac{P(X_1 = x_1, \dots, X_n = x_n; T = t)}{P(T = t)} = 0$$

- Si  $x_1 + \dots + x_n = t$ , entonces

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n | T = t) &= \frac{P(X_1 = x_1, \dots, X_n = x_n; T = t)}{P(T = t)} \\ &= \frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} \end{aligned}$$

- $P(X_1 = x_1, \dots, X_n = x_n) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} = \theta^t (1 - \theta)^{n-t}$ .
- $P(T = t) = \binom{n}{t} \theta^t (1 - \theta)^{n-t}$ , pues  $T = \sum_{i=1}^n X_i \sim \text{Bi}(n, \theta)$ .

Luego,

$$\frac{P(X_1 = x_1, \dots, X_n = x_n)}{P(T = t)} = \frac{\theta^t (1 - \theta)^{n-t}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

que no depende de  $\theta$ , por tanto,  $T$  es un estimador suficiente.

**Teorema 3.3.5** (Teorema de Factorización de Neyman-Fisher). Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población cuya distribución pertenece a la familia de distribuciones  $\{F_\theta : \theta \in \Theta\}$ . Entonces, un estadístico  $T(X_1, \dots, X_n)$  es suficiente para  $\theta$  si y solo si

$$f(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n)$$

donde  $g$  y  $h$  son funciones medibles no negativas tales que  $g$  depende de  $\theta$  y  $h$  no depende de  $\theta$ .

*Demostración.* Esta demostración es cortesía de Miguel Ángel Quesada Martínez. Haremos la demostración solamente para el caso discreto.

$\implies$  Supongamos que  $T$  es un estadístico suficiente para  $\theta$ , por definición, esto quiere decir que  $P(X = x | T = t)$  es independiente de  $\theta$  para cualquier  $t$  y podemos escribir:

$$\begin{aligned} P_\theta(X = x) &= P_\theta[X_1 = x_1, \dots, X_n = x_n, T(X_1, \dots, X_n) = t] \\ &= P_\theta[X_1 = x_1, \dots, X_n = x_n | T(X_1, \dots, X_n) = t] \cdot P_\theta[T(X_1, \dots, X_n) = t] \end{aligned}$$

Si definimos

$$g(T(x_1, \dots, x_n); \theta) = P_\theta[T(x_1, \dots, x_n) = t]$$

$$h(x_1, \dots, x_n) = \begin{cases} 0 & \text{si } P(X = x) = 0 \\ P_\theta[X_1 = x_1, \dots, X_n = x_n | T(X_1, \dots, X_n)] & \text{si } P(X = x) > 0 \end{cases}$$

Es claro que  $P_\theta(X_1 = x_1, \dots, X_n = x_n) = g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n)$ .

Supongamos que  $P_\theta(X_1 = x_1, \dots, X_n = x_n) = g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n)$ . Si fijamos  $t_0$  tenemos que

$$P_\theta(T = t_0) = \sum_{\{x: T(x)=t_0\}} P_\theta(X = x) = \sum_{\{x: T(x)=t_0\}} g(t_0, \theta) h(x) = g_\theta(t_0) \cdot \sum_{\{x: T(x)=t_0\}} h(x)$$

Supongamos que  $P_\theta(T = t_0) > 0$ , entonces

$$P_\theta(X = x | T = t_0) = \frac{P(X = x, T = t_0)}{P(T = t_0)} = \begin{cases} 0 & \text{si } T(x) \neq t_0 \\ \frac{P_\theta(X=x)}{P_\theta(T=t_0)} & \text{si } T(x) = t_0 \end{cases}$$

Así vemos que si  $T(x) = t_0$

$$P_\theta(X = x | T = t_0) = \frac{P(X = x, T = t_0)}{P(T = t_0)} = \frac{g_\theta(t_0) h(x)}{g_\theta(t_0) \sum_{\{x: T(x)=t_0\}} h(x)} = \frac{h(x)}{\sum_{\{x: T(x)=t_0\}} h(x)}$$

la cual no depende de  $\theta$ . □

Propiedades de los estadísticos suficientes: Sea  $T$  estadístico suficiente para  $\{F_\theta : \theta \in \Theta\}$  :

1.  $T$  es suficiente para  $\{F_\theta : \theta \in \Theta' \subset \Theta\}$ .
2. Si  $T = m(U)$ ,  $m$  medible y  $U$  estadístico para  $\theta$ , entonces  $U$  es suficiente para  $\theta$ .
3. Si  $m$  es biyectiva, entonces  $m(T)$  es suficiente.

**Definición 3.3.6.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X$  con distribución  $f(x, \vec{\theta})$ , con  $\vec{\theta} = (\theta_1, \dots, \theta_r)$ ,  $\vec{\theta} \in \Theta$ , vector de parámetros desconocidos. Diremos que un estadístico  $(T_1, \dots, T_k)$  es un **estadístico conjuntamente suficiente** si y solo si la distribución de  $(X_1, \dots, X_n)$  dado que  $(T_1 = t_1, \dots, T_k = t_k)$  no depende de  $\theta$ .

**Teorema 3.3.7** (Teorema de factorización generalización). Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X$  con distribución  $f(x, \vec{\theta})$ . Un conjunto de estadísticos  $(T_1, \dots, T_k)$  es un estadístico conjuntamente suficiente si y solo si la densidad conjunta de  $(X_1, \dots, X_n)$  se puede factorizar de la forma

$$f(x_1, \dots, x_n; \vec{\theta}) = g(t_1, \dots, t_k; \vec{\theta}) h(x_1, \dots, x_n)$$

donde  $g$  y  $h$  son funciones medibles no negativas.

**Ejemplo 3.3.8.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim U(\theta_1, \theta_2)$ , con  $0 < \theta_1 < \theta_2$ . Entonces la distribución de  $(X_1, \dots, X_n)$  viene dada por

$$\begin{aligned} f(x_1, \dots, x_n; \theta_1, \theta_2) &= \prod_{i=1}^n f(x_i; \theta_1, \theta_2) = \left( \frac{1}{\theta_2 - \theta_1} \right)^n \prod_{i=1}^n I_{(\theta_1, \theta_2)}(x_i) \\ &= \left( \frac{1}{\theta_2 - \theta_1} \right)^n \cdot I_{(0, \min\{x_i\})}(\theta_1) \cdot I_{(\max\{x_i\}, \infty)}(\theta_2) \\ &= g(t_1, t_2; \theta_1, \theta_2) h(x_1, \dots, x_n) \end{aligned}$$

donde

$$g(t_1, t_2; \theta_1, \theta_2) = \left( \frac{1}{\theta_2 - \theta_1} \right)^n \cdot I_{(0, \min\{x_i\})}(\theta_1) \cdot I_{(\max\{x_i\}, \infty)}(\theta_2)$$

$$h(x_1, \dots, x_n) = 1$$

Entonces  $(T_1, T_2) = (\min_{i=1, \dots, n} \{x_i\}, \max_{i=1, \dots, n} \{x_i\})$  es un estadístico conjuntamente suficiente para  $(\theta_1, \theta_2)$ .

**Ejemplo 3.3.9.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim N(\mu, \sigma)$ . Veamos que

$$(T_1, T_2) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

es un estadístico conjunto suficiente para  $(\mu, \sigma^2)$ .

$$\begin{aligned} f(x_1, \dots, x_n; \mu, \sigma^2) &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right] \\ &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i^2 + \mu^2 - 2\mu x_i) \right] \\ &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{\sigma^2} \left( \sum_{i=1}^n x_i^2 + n\mu^2 - 2\mu \sum_{i=1}^n x_i \right) \right] \end{aligned}$$

Si aplicamos el teorema de factorización, se tiene que

$$f(x_1, \dots, x_n; \mu, \sigma^2) = g(t_1, t_2; \mu, \sigma^2) h(x_1, \dots, x_n)$$

siendo

$$g(t_1, t_2; \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{\sigma^2} \left( \sum_{i=1}^n x_i^2 + n\mu^2 - 2\mu \sum_{i=1}^n x_i \right) \right]$$

$$h(x_1, \dots, x_n) = 1$$

Por lo tanto,  $(T_1, T_2) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$  es un estadístico conjuntamente suficiente para  $(\mu, \sigma^2)$ .

**Observación 3.3.10.** Los estadísticos suficientes no son únicos. Es fácil demostrar que  $T = \bar{X}$  es un estadístico suficiente y que  $(T_1, T_2) = (\bar{X}, S^2)$  es un estadístico conjuntamente suficiente para  $(\mu, \sigma^2)$ .

### 3.3.2. Familia exponencial uniparamétrica

**Definición 3.3.11.** Diremos que  $f(x; \theta)$  pertenece a la familia exponencial uniparamétrica si

$$f(x; \theta) = a(\theta)b(x) \exp[c(\theta)d(x)] \text{ para cada } \theta \in \Theta,$$

siendo  $a(\theta) > 0$  y  $b(x) > 0$ .

**Observación 3.3.12.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población con distribución  $f(x; \theta)$  que pertenece a la familia exponencial uniparamétrica. Entonces

$$\begin{aligned} f(x_1, \dots, x_n; \theta) &= \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n a(\theta)b(x_i) \exp[c(\theta)d(x_i)] \\ &= (a(\theta))^n \prod_{i=1}^n b(x_i) \exp \left[ c(\theta) \sum_{i=1}^n d(x_i) \right] \\ &= g(t; \theta) h(x_1, \dots, x_n) \end{aligned}$$

siendo

$$g(t; \theta) = (a(\theta))^n \exp \left[ c(\theta) \sum_{i=1}^n d(x_i) \right]$$

$$h(x_1, \dots, x_n) = \prod_{i=1}^n b(x_i)$$

Por tanto, el estadístico  $T = \sum_{i=1}^n d(X_i)$  es un estadístico suficiente para  $\theta$ .

**Ejemplo 3.3.13.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim Po(\theta)$ ,  $\theta > 0$ . ¿Pertenece a la familia exponencial uniparamétrica?

$$P(X = x; \theta) = e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{x!} e^{-\theta} e^{x \log \theta} = \exp(-\theta) \frac{1}{x!} \exp(x \log \theta)$$

$$= a(\theta) b(x) \exp[c(\theta) d(x)]$$

siendo

$$a(\theta) = \exp(-\theta) \quad b(x) = \frac{1}{x!}$$

$$c(\theta) = \log \theta \quad d(x) = x$$

Usando la observación anterior tenemos que  $T = \sum_{i=1}^n X_i$  es un estadístico suficiente para  $\theta$ .

### 3.3.3. Familia exponencial $k$ -paramétrica

**Definición 3.3.14.** Diremos que  $f(x; \vec{\theta})$ ,  $\vec{\theta} = (\theta_1, \dots, \theta_k)$ , pertenece a la familia exponencial  $k$ -paramétrica si

$$f(x; \vec{\theta}) = a(\vec{\theta}) b(x) \exp \left[ \sum_{j=1}^k c_j(\vec{\theta}) d_j(x) \right]$$

**Ejemplo 3.3.15.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X \sim Ga(\alpha, \beta)$ . ¿Pertenece a la familia exponencial biparamétrica?

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1} = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} e^{\alpha \log x} \frac{1}{x}$$

$$= a(\alpha, \beta) b(x) \exp \left[ \sum_{j=1}^2 c_j(\alpha, \beta) d_j(x) \right]$$

siendo

$$a(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \quad c_1(\alpha, \beta) = \alpha \quad d_1(x) = \log x$$

$$b(x) = \frac{1}{x} \quad c_2(\alpha, \beta) = -\beta \quad d_2(x) = x$$

Por tanto

$$(T_1, T_2) = \left( \sum_{i=1}^n d_1(X_i), \sum_{i=1}^n d_2(X_i) \right) = \left( \sum_{i=1}^n \log(X_i), \sum_{i=1}^n X_i \right)$$

es un estadístico conjuntamente suficiente para  $(\alpha, \beta)$ .

### 3.4. Estimadores insesgados y de mínima varianza

#### 3.4.1. Teorema de Cramer-Rao. Información de Fisher

**Definición 3.4.1.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de la distribución  $f(x; \theta)$ ,  $\theta \in \Theta$ . Un estimador  $T^*(X_1, \dots, X_n)$  de  $g(\theta)$  diremos que es un **estimador insesgado uniformemente de mínima varianza** de  $g(\theta)$  si y solo si

- (i)  $E_\theta(T^*) = g(\theta)$ .
- (ii)  $\text{Var}_\theta(T^*) \leq \text{Var}_\theta(T)$  para cualquier otro estimador de  $g(\theta)$  que verifique que  $E_\theta(T) = g(\theta)$  para todo  $\theta \in \Theta$ .

**Teorema 3.4.2 (Cramer-Rao).** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de la distribución  $f(x; \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}$ . Sea  $T^*$  un estimador insesgado para  $g(\theta)$ . Si se verifican las siguientes condiciones de regularidad

- (i)  $\{(x_1, \dots, x_n) \in \mathfrak{X} : f(x_1, \dots, x_n; \theta) > 0\}$  no depende de  $\theta$ .
- (ii) Existe  $\frac{\partial}{\partial \theta} \log(f(x; \theta))$  para todo  $x$  y todo para todo  $\theta \in \Theta$ .
- (iii)

$$\frac{\partial}{\partial \theta} \int \cdots \int \prod_{i=1}^n f(x_i; \theta) dx_1 \dots dx_n = \int \cdots \int \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i; \theta) dx_1 \dots dx_n$$

- (iv)

$$\begin{aligned} \frac{\partial}{\partial \theta} \int \cdots \int T^*(x_1, \dots, x_n) \prod_{i=1}^n f(x_i; \theta) dx_1 \dots dx_n = \\ \int \cdots \int T^*(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i; \theta) dx_1 \dots dx_n \end{aligned}$$

- (v) Para todo  $\theta \in \Theta$

$$0 < E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log(f(X; \theta)) \right)^2 \right] < \infty$$

entonces

$$\text{Var}_\theta(T^*) \geq \frac{(g'(\theta))^2}{n E_\theta \left[ \left( \frac{\partial}{\partial \theta} \log(f(X; \theta)) \right)^2 \right]}$$

**Observación 3.4.3.** Si existe un estimador  $T$  para el que se alcanza la igualdad, entonces  $T$  es un estimador insesgado uniformemente de mínima varianza y a ese estimador lo llamaremos **estimador eficiente**.

**Ejemplo 3.4.4.** Sea  $(X_1, \dots, X_n)$  una muestra aleatoria simple de una población  $X$  con función de densidad

$$f(x; \theta) = \theta e^{-\theta x}$$

para todo  $x > 0$  y siendo  $\theta > 0$ . Sea  $g(\theta) = \frac{1}{\theta}$ . Calculemos una cota de Cramer-Rao para un estimador insesgado  $T$  de  $g(\theta)$ .

**Observación 3.4.5.** Bajo las condiciones del Teorema de Cramer-Rao y si existe la segunda derivada de  $\log f(X; \theta)$  respecto de  $\theta$ , se tiene que

$$E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \right] = -E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right]$$

Observamos que  $\log f(X; \theta) = \log \theta - \theta X$ . Derivando

$$\begin{aligned} \frac{\partial}{\partial \theta} \log f(X; \theta) &= \frac{1}{\theta} - X \\ \frac{\partial^2}{\partial \theta^2} \log f(X; \theta) &= -\frac{1}{\theta^2} \implies E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right] = E_{\theta} \left[ -\frac{1}{\theta^2} \right] = -\frac{1}{\theta^2} \end{aligned}$$

Aplicando el teorema de Cramer-Rao

$$\text{Cota} = \frac{(g'(\theta))^2}{n E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(X; \theta) \right)^2 \right]} = \frac{(g'(\theta))^2}{n - E_{\theta} \left[ \frac{\partial^2}{\partial \theta^2} \log f(X; \theta) \right]} = \frac{\left( \frac{1}{\theta} \right)^2}{\frac{n}{\theta^2}} = \frac{1}{n\theta^2}$$

Calculemos ahora un estimador insesgado por el método de los momentos para  $g(\theta) = \frac{1}{\theta}$ . Observamos que si  $X \sim \text{Exp}(\theta)$ , entonces  $E(X) = \frac{1}{\theta}$ . Mediante el método de los momentos, el estimador para la media,  $\frac{1}{\theta}$ , sería  $\bar{X}$ .

Sea  $T = \bar{X}$ .  $T$  es estimador insesgado para  $g(\theta)$  y

$$\text{Var}_{\theta}(T) = \text{Var}_{\theta} \bar{X} = \frac{\frac{1}{\theta^2}}{n} = \frac{1}{n\theta^2}$$

Por tanto,  $T$  alcanza la cota de Cramer-Rao, lo que nos dice que  $T = \bar{X}$  es estimador insesgado uniformemente de mínima varianza para  $g(\theta) = \frac{1}{\theta}$ .

Aplicando el teorema de invarianza, vimos que el estimador de máxima verosimilitud para  $\theta$  es  $\frac{n}{\sum_{i=1}^n X_i}$ , y por tanto, el estimador de máxima verosimilitud para  $\frac{1}{\theta}$  es  $\frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ .

### 3.4.2. Teorema de Rao-Blackwell. Teorema de Lehmann-Scheffé

**Teorema 3.4.6** (Rao-Blackwell). Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de  $X$  con distribución  $f(x; \theta)$  y sea  $\{S_1, \dots, S_k\}$  un conjunto de estadísticos conjuntamente suficientes para  $\theta$ . Sea  $T$  un estimador insesgado de  $g(\theta)$ . Sea  $T^*$  dado como  $T^* = E_{\theta}[T|S_1, \dots, S_k]$ . Entonces:

- (i)  $T^*$  es un estadístico y es función de los estadísticos suficientes  $S_1, \dots, S_k$ .
- (ii)  $E_{\theta}(T^*) = g(\theta)$  ( $T^*$  es un estimador insesgado).
- (iii)  $\text{Var}_{\theta}(T^*) \leq \text{Var}_{\theta}(T)$ , para todo  $\theta \in \Theta$ .

Además existe  $\theta_0 \in \Theta$  tal que

$$\text{Var}_{\theta_0}(T^*) < \text{Var}_{\theta_0}(T)$$

salvo que  $T = T^*$  con probabilidad 1.

**Ejemplo 3.4.7.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una distribución  $\text{Ber}(\theta)$ . Consideremos  $T = X_1$ , que es un estimador insesgado para  $\theta$  y  $S = \sum_{i=1}^n X_i$ , que es un estadístico suficiente para  $\theta$  y sea  $g(\theta) = \theta$ . Veamos que se verifica el teorema de Rao-Blackwell.

Sea  $T^* = E(T|S) = E(X_1 | \sum_{i=1}^n X_i)$ . El teorema afirma que  $T^*$  es un estimador insesgado para  $\theta$ .



Calculamos ahora los valores de la variable aleatoria  $E(T|S)$ .

$$\begin{aligned}
 E(T|S=s) &= E\left(T \left| \sum_{i=1}^n X_i = s \right.\right) = 0 \cdot P\left(X_1 = 0 \left| \sum_{i=1}^n X_i = s \right.\right) + 1 \cdot P\left(X_1 = 1 \left| \sum_{i=1}^n X_i = s \right.\right) \\
 &= P\left(X_1 = 1 \left| \sum_{i=1}^n X_i = s \right.\right) = \frac{P(\sum_{i=1}^n X_i = s | X_1 = 1) \cdot P(X_1 = 1)}{P(\sum_{i=1}^n X_i = s)} \\
 &= \frac{P(\sum_{i=2}^n X_i = s-1 | X_1 = 1) \cdot P(X_1 = 1)}{P(\sum_{i=1}^n X_i = s)} \\
 &= \frac{\theta \binom{n-1}{s-1} \theta^{s-1} (1-\theta)^{n-1-(s-1)}}{\binom{n}{s} \theta^s (1-\theta)^{n-s}} = \dots = \frac{s}{n}
 \end{aligned}$$

Consideramos el estimador

$$T^* = \frac{S}{n} = E(T|S) = \frac{\sum_{i=1}^n X_i}{n}$$

Comprobamos que el teorema se verifica:

- (i)  $T^*$  es un estadístico (en este caso estimador) que es función del estadístico  $S$ .
- (ii)  $T^*$  es un estimador insesgado para  $\theta$ .
- (iii)

$$Var_{\theta}(T^*) = \frac{\theta(1-\theta)}{n} \leq \theta(1-\theta) = Var_{\theta}(T)$$

**Definición 3.4.8.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una población  $X$  con distribución  $f(x; \theta)$ ,  $\theta \in \Theta$ . Un estadístico  $T = T(X_1, \dots, X_n)$  se dice completo si y solo si el hecho de que  $E_{\theta}[g(T)] = 0$  implica que  $P_{\theta}[g(T) = 0] = 1$  para todo  $\theta \in \Theta$ , donde  $g(T)$  es un estadístico.

Al estadístico  $T = T(X_1, \dots, X_n)$  se le llama **estadístico completo**.

**Teorema 3.4.9.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con distribución  $f(x; \theta)$ ,  $\theta \in \Theta$ . Si  $f(x; \theta)$  pertenece a la familia exponencial uniparamétrica entonces  $T = \sum_{i=1}^n d(X_i)$  es un estadístico minimal, suficiente y completo para  $\theta$ .

**Definición 3.4.10.** Un estadístico  $T$  suficiente diremos que es **minimal** si dado otro estadístico suficiente  $T'$ , existe una función medible  $\varphi$  tal que  $T = \varphi(T')$ .

**Teorema 3.4.11** (Lehmann-Scheffé I). Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con distribución  $f(x; \theta)$ ,  $\theta \in \Theta$ . Sea  $S$  un estadístico suficiente y completo para  $h(\theta)$ . Si  $T^*(S)$  es un estimador insesgado para  $h(\theta)$ , entonces  $T^*(S)$  es un estimador insesgado uniformemente de mínima varianza para  $h(\theta)$ .

**Teorema 3.4.12** (Lehmann-Scheffé II). *Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con distribución  $f(x; \theta)$ ,  $\theta \in \Theta$ . Si  $T_2$  es un estadístico suficiente y completo para  $h(\theta)$  y  $T_1$  es un estadístico insesgado para  $h(\theta)$ , entonces  $E_\theta(T_1|T_2)$  es el estimador insesgado uniformemente de mínima varianza para  $h(\theta)$ .*

## Capítulo 4

# Intervalos de confianza

### 4.1. Intervalos de confianza

#### 4.1.1. Definición de intervalo de confianza

**Definición 4.1.1.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con función de densidad  $f(x; \theta)$ ,  $\theta \in \Theta$ . Sean  $T_1 = t_1(X_1, \dots, X_n)$  y  $T_2 = t_2(X_1, \dots, X_n)$  dos estadísticos que satisfacen  $T_1 \leq T_2$  tales que  $P_\theta[T_1 < \tau(\theta) < T_2] \equiv \gamma$ , donde  $\gamma$  no depende de  $\theta$ . Entonces, el intervalo aleatorio  $(T_1, T_2)$  se llama **intervalo de  $\tau(\theta)$  al  $100\gamma\%$  de confianza**. A  $\gamma$  se le llama **coeficiente de confianza**.

Un valor  $(t_1, t_2)$  del intervalo aleatorio  $(T_1, T_2)$  también se llama intervalo de  $\tau(\theta)$  al  $100\gamma\%$  de confianza.

**Ejemplo 4.1.2.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu = \theta, \sigma^2 = 9)$ . Sea  $T_1 = \bar{X} - \frac{6}{\sqrt{n}}$  y  $T_2 = \bar{X} + \frac{6}{\sqrt{n}}$ . Entonces  $(T_1, T_2)$  constituye un intervalo aleatorio y es un intervalo de confianza para  $\tau(\theta) = \theta$  cuyo coeficiente de confianza es

$$\gamma = P_\theta \left[ \bar{X} - \frac{6}{\sqrt{n}} < \theta < \bar{X} + \frac{6}{\sqrt{n}} \right] = P_\theta \left[ -2 < \frac{\bar{X} - \theta}{3/\sqrt{n}} < 2 \right]$$

Como  $\bar{X} \sim N\left(\theta, \frac{\theta}{\sqrt{n}}\right)$ , tenemos que  $\gamma = \Phi(2) - \Phi(-2) = 0.9544$ . Además, si en una muestra aleatoria de 25 observaciones se tiene una media muestral de, digamos, 17.5, entonces el intervalo  $\left(17.5 - \frac{6}{\sqrt{25}}, 17.5 + \frac{6}{\sqrt{25}}\right)$  es un intervalo de confianza para  $\theta$ .

#### 4.1.2. Método del pivote

**Definición 4.1.3.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con función de densidad  $f(x; \theta)$ ,  $\theta \in \Theta$ . Sea  $Q = q(X_1, \dots, X_n; \theta)$  una función de  $X_1, \dots, X_n$  y  $\theta$ . Si  $Q$  tiene una distribución que no depende de  $\theta$ , entonces decimos que  $Q$  es un **pivote**.

**Ejemplo 4.1.4.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu = \theta, \sigma^2 = 9)$ . Tenemos que  $\bar{X} - \theta$  es un pivote ya que  $\bar{X} - \theta$  tiene una distribución  $N(\mu = 0, \sigma^2 = \frac{9}{n})$ , que no depende de  $\theta$ .

**Observación 4.1.5.** Si tenemos una muestra aleatoria simple  $X_1, \dots, X_n$  de una variable aleatoria  $X$  con función de distribución  $F(x; \theta)$  continua, entonces

- $F(X_i; \theta) \sim U([0, 1])$  para cada  $i = 1, \dots, n$ .
- $-\log F(X_i; \theta) \sim \text{Exp}(\lambda = 1)$  para cada  $i = 1, \dots, n$ .

Por tanto, tenemos que los siguientes pivotes:

- $Q = \prod_{i=1}^n F(X_i; \theta) \sim U([0, 1])$ .
- $Q = \sum_{i=1}^n -\log F(X_i; \theta) \sim Ga(\alpha = n, \beta = 1)$ .

**Ejemplo 4.1.6.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu, \sigma)$ ,  $\mu$  y  $\sigma$  desconocidos.

¿Intervalo de confianza para  $\mu$ ? Por el Teorema de Fisher sabemos que  $\bar{X}$  y  $S^2$  son independientes y:

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

Consideramos el pivote:

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

que es independiente de  $\mu$  y  $\sigma$ . Sean  $q_1 < q_2$  y sea  $x_1, \dots, x_n$  una realización de la muestra, entonces

$$\left\{ q_1 < \frac{\bar{x} - \mu}{s/\sqrt{n}} < q_2 \right\} \iff \left\{ \bar{x} - q_2 \frac{s}{\sqrt{n}} < \mu < \bar{x} - q_1 \frac{s}{\sqrt{n}} \right\}$$

Definimos:

$$P\left[ q_1 < \frac{\bar{X} - \mu}{S/\sqrt{n}} < q_2 \right] = \int_{q_1}^{q_2} f_T(t) dt = \gamma$$

con lo que tenemos que  $\left( \bar{X} - q_2 \frac{S}{\sqrt{n}}, \bar{X} - q_1 \frac{S}{\sqrt{n}} \right)$  es un intervalo para  $\mu$  al  $100\gamma\%$  de confianza. La longitud de dicho intervalo es  $(q_2 - q_1)(S/\sqrt{n})$ . Intentemos minimizar la longitud del intervalo sujeto a que la confianza es  $\gamma$ , es decir, hemos de resolver el siguiente problema de optimización:

$$\begin{aligned} \text{Min: } L &= \frac{S}{\sqrt{n}}(q_2 - q_1) \\ \text{s.a } &\int_{q_1}^{q_2} f_T(t) dt = \gamma \end{aligned}$$

Vamos a considerar que  $q_2$  es función de  $q_1$ , así que para minimizar  $L$ , tenemos que hacer:

$$\frac{dL}{dq_1} = 0 \iff \frac{S}{\sqrt{n}} \left( \frac{dq_2}{dq_1} - 1 \right) = 0$$

Además:

$$\int_{q_1}^{q_2} f_T(t) dt = \gamma \implies F_T(q_2) - F_T(q_1) = \gamma \xrightarrow{\text{derivando}} f_T(q_2) \frac{dq_2}{dq_1} - f_T(q_1) = 0$$

Con esto tenemos que

$$0 = \frac{S}{\sqrt{n}} \left( \frac{dq_2}{dq_1} - 1 \right) = \frac{S}{\sqrt{n}} \left( \frac{f_T(q_1)}{f_T(q_2)} - 1 \right) = 0$$

y esto puede ocurrir si  $f_T(q_1) = f_T(q_2)$ , lo que implica que  $q_1 = -q_2$  (no puede ser  $q_1 = q_2$  porque estamos suponiendo que  $q_1 < q_2$ ). Por tanto el intervalo para  $\mu$  al  $100\gamma\%$  de confianza con longitud mínima es  $\left( \bar{X} - q_2 \frac{S}{\sqrt{n}}, \bar{X} + q_2 \frac{S}{\sqrt{n}} \right)$ .

¿Intervalo de confianza para  $\sigma$ ? Por el Teorema de Fisher:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

Consideramos el pivote:

$$Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2$$

que es independiente de  $\mu$  y  $\sigma$ . Sean  $a_1 < q_2$  y sea  $x_1, \dots, x_n$  una realización de la muestra, entonces

$$\left\{ q_1 < \frac{(n-1)S^2}{\sigma^2} < q_2 \right\} \iff \left\{ \frac{(n-1)s^2}{q_2} < \sigma^2 < \frac{(n-1)s^2}{q_1} \right\}$$

Definimos:

$$P \left[ \frac{(n-1)s^2}{q_2} < \sigma^2 < \frac{(n-1)s^2}{q_1} \right] = \gamma$$

con lo que tenemos que  $\left( \frac{(n-1)S^2}{q_2}, \frac{(n-1)S^2}{q_1} \right)$  es un intervalo para  $\sigma^2$  al  $100\gamma\%$  de confianza.

**Ejemplo 4.1.7.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu, \sigma)$ ,  $\mu$  y  $\sigma$  desconocidos. ¿Región de confianza para  $(\mu, \sigma^2)$ ?

$$Q_1 = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}, \quad Q_2 = \frac{(n-1)S^2}{\sigma^2}$$

son pivotes para  $\mu$  y  $\sigma^2$  respectivamente. Definimos

$$P \left[ -q_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < q_1 \right] = \gamma_1$$

$$P \left[ q'_2 < \frac{(n-1)S^2}{\sigma^2} < q''_2 \right] = \gamma_2$$

Si hacemos que se alcancen las igualdades, podemos representar una región de confianza para  $(\mu, \sigma^2)$

$$-q_1 = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \implies \sigma^2 = \frac{\sqrt{n}}{q_1}(\bar{X} - \mu)$$

$$q_1 = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \implies \sigma^2 = \frac{\sqrt{n}}{q_1}(\bar{X} - \mu)$$

$$q'_2 = \frac{(n-1)S^2}{\sigma^2} \implies \sigma^2 = \frac{(n-1)S^2}{q'_2}$$

$$q''_2 = \frac{(n-1)S^2}{\sigma^2} \implies \sigma^2 = \frac{(n-1)S^2}{q''_2}$$

Como  $Q_1$  y  $Q_2$  son independientes:

$$P[-q_1 < Q_1 < q_1, q'_2 < Q_2 < q''_2] = P[-q_1 < Q_1 < q_1] \cdot P[q'_2 < Q_2 < q''_2] = \gamma_1 \cdot \gamma_2$$

Luego

$$\left\{ (\mu, \sigma^2) : q_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < q_1, q'_2 < \frac{(n-1)S^2}{\sigma^2} < q''_2 \right\}$$

es una región de confianza para  $(\mu, \sigma^2)$  a un  $100\gamma_1\gamma_2\%$ .

**Ejemplo 4.1.8.** Intervalo de confianza para la diferencia de medias de distribuciones normales. Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\mu_1, \sigma)$  y sea  $Y_1, \dots, Y_k$  una muestra aleatoria simple de una variable aleatoria  $Y \sim N(\mu_2, \sigma)$ . Sabemos entonces que  $\bar{X} \sim N\left(\mu_1, \frac{\sigma}{\sqrt{n}}\right)$  y  $\bar{Y} \sim N\left(\mu_2, \frac{\sigma}{\sqrt{k}}\right)$ . Por tanto  $\bar{Y} - \bar{X} \sim N\left(\mu_2 - \mu_1, \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{k}}\right)$ . Por el Teorema de Fisher:

de donde deducimos que

$$\frac{1}{\sigma^2} \left( \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^k (Y_j - \bar{Y})^2 \right) \sim \chi_{(n+k-2)}^2$$

Definimos

$$Q_1 = \frac{\frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{\sigma^2/n + \sigma^2/k}}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{j=1}^k (y_j - \bar{y})^2}{\sigma^2(n+k-2)}}} = \frac{\frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sigma \sqrt{1/n + 1/k}}}{S_p / \sigma} = \frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{S_p \sqrt{1/n + 1/k}} \sim t_{n+k-2}$$

siendo

$$S_p^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^k (Y_j - \bar{Y})^2$$

## 4.2. Determinación del tamaño muestral

Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con función de densidad  $f(x; \theta)$ . Sea  $Q$  un pivote. Supongamos que

$$P(q_1 < Q < q_2) = \gamma = P(t_1(x_1, \dots, x_n) < \theta < t_2(x_1, \dots, x_n))$$

para ciertos  $t_1, t_2$  estadísticos y  $x_1, \dots, x_n$  una realización de la muestra.

Si suponemos que  $X \sim N(\mu, \sigma)$ . Fijamos  $\varepsilon > 0$ . Sabemos que

$$\left( \bar{X} - t_{n-1, 1-\frac{1-\gamma}{2}} \frac{S^2}{\sqrt{n}}, \bar{X} + t_{n-1, 1-\frac{1-\gamma}{2}} \frac{S^2}{\sqrt{n}} \right)$$

es un intervalo de confianza para  $\mu$ . Entonces

$$|\bar{X} - \mu| < t_{n-1, 1-\frac{1-\gamma}{2}} \frac{S^2}{\sqrt{n}} < \varepsilon \iff t_{n-1, 1-\frac{1-\gamma}{2}} \frac{S^2}{\varepsilon} < \sqrt{n} \implies n > \left[ t_{n-1, 1-\frac{1-\gamma}{2}} \frac{S^2}{\varepsilon} \right]$$

Si  $n$  es lo suficientemente grande, podemos aproximar  $t_{n-1}$  por  $Z \sim N(0, 1)$ .

## Capítulo 5

# Contrastes de hipótesis

### 5.1. Hipótesis estadística

**Definición 5.1.1.** Decimos que una afirmación o hipótesis sobre la población es:

- Simple, si determina la distribución.
- Compuesta, si no determina la distribución.

**Definición 5.1.2.** Un test de una hipótesis estadística es una regla para aceptar o rechazar una hipótesis.

**Definición 5.1.3.** Sea  $\mathcal{T}$  un test para la hipótesis  $H$ . Decimos que  $C_{\mathcal{T}} \subset \mathfrak{X}$  es una región crítica si rechazamos la hipótesis  $H$  si y solo si la realización de la muestra  $(x_1, \dots, x_n)$  está en  $C_{\mathcal{T}}$ .

**Observación 5.1.4.** Sean  $H_0$  una hipótesis nula y  $H_1$  una hipótesis alternativa. Entonces:

- Si  $H_0$  es cierta y acepto  $H_0$ , entonces hemos acertado.
- Si  $H_0$  es cierta y rechazo  $H_0$ , entonces hemos cometido un error de tipo I.
- Si  $H_0$  es falsa y acepto  $H_0$ , entonces hemos cometido un error de tipo II.
- Si  $H_0$  es falsa y rechazo  $H_0$ , entonces hemos acertado.

Denotaremos:

- $P(\text{Rechazar } H_0 \mid H_0 \text{ cierta}) = \alpha$  (significación).
- $P(\text{No rechazar } H_0 \mid H_0 \text{ falsa}) = \beta$  (riesgo).
- $P(\text{Rechazar } H_0 \mid H_0 \text{ falsa}) = 1 - \beta$  (potencia).
- $P(\text{No rechazar } H_0 \mid H_0 \text{ cierta}) = 1 - \alpha$  (confianza).

**Definición 5.1.5.** Función de potencia de un test  $\mathcal{T}$ :

$$\Pi_{\mathcal{T}}(\theta) = P_{\theta}((X_1, \dots, X_n) \in C_{\mathcal{T}}).$$

## 5.2. Criterio de elección del test más potente

**Definición 5.2.1.** Sea  $\mathcal{T}$  un test de  $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$ . Fijamos  $\alpha \in (0, 1)$ . Decimos que  $\mathcal{T}$  es de máxima potencia para  $\alpha$  cuando:

- $\Pi_{\mathcal{T}}(\theta_0) = \alpha$ .
- $\Pi_{\mathcal{T}}(\theta_1) \geq \Pi_{\mathcal{T}^*}(\theta_1)$  para todo test  $\mathcal{T}^*$  con  $\Pi_{\mathcal{T}^*}(\theta_0) = \alpha$ .

**Definición 5.2.2.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable  $X$  con densidad  $f(x; \theta)$ . Consideramos  $\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$  con  $\Theta = \Theta_0 \dot{\cup} \Theta_1$ . Definimos la razón de verosimilitudes como

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(x_1, \dots, x_n; \theta)}{\sup_{\theta \in \Theta} L(x_1, \dots, x_n; \theta)}$$

Un test de razón de verosimilitudes es aquel con región crítica  $C_{\mathcal{T}} = \{x \in \mathfrak{X} : \lambda(x) \leq c\}$  para cierto  $c \in [0, 1)$ .

**Observación 5.2.3.** Si  $\hat{\theta}$  es el estimador de máxima verosimilitud para  $\theta$  en  $\Theta$  y  $\hat{\theta}_0$  es el estimador de máxima verosimilitud para  $\theta$  en  $\Theta_0$ , entonces:

$$\lambda(x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n; \hat{\theta}_0)}{L(x_1, \dots, x_n; \hat{\theta})}$$

**Ejemplo 5.2.4.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X \sim N(\theta, 1)$ . Calulemos un test de razón de verosimilitudes para el contraste  $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$ . Sea  $x = (x_1, \dots, x_n)$  una realización de la muestra. El estimador de máxima verosimilitud para  $\theta$  en  $\Theta = \mathbb{R}$  es  $\bar{X}$ . El único posible valor que maximize  $L(x; \theta)$  en  $\Theta_0 = \{\theta_0\}$  es el propio  $\theta_0$ . Entonces

$$\begin{aligned} \lambda(x) &= \frac{L(x; \theta_0)}{L(x; \bar{x})} = \frac{(2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2}}{(2\pi)^{-n/2} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2}} = \exp \left[ \frac{1}{2} \left( \sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (x_i - \theta_0)^2 \right) \right] \\ &= \dots = \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^n (\bar{x} - \theta_0)^2 \right) \right] \end{aligned}$$

Definimos el test  $\mathcal{T}$  de razón de verosimilitudes como: Rechazo  $H_0$  si algún valor de realización de la muestra se encuentra en:

$$C_{\mathcal{T}} = \{x \in \mathfrak{X} : \lambda(x) \leq c\} = \left\{ x \in X : |\bar{x} - \theta_0| \geq \sqrt{-\frac{2}{n} \log c} \right\}$$

**Teorema 5.2.5.** Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con densidad  $f(x; \theta)$ . Sea  $T(X_1, \dots, X_n)$  un estadístico suficiente para  $\theta$  y sean  $\lambda(t)$  y  $\lambda^*(t)$  radios de verosimilitudes para  $T(X_1, \dots, X_n)$  y  $X$  respectivamente. Entonces

$$\lambda^*(T(x)) = \lambda(x), \quad \text{para cada } x \in \mathfrak{X}$$

*Demostración.* Usando el Teorema de Factorización sabemos que  $f(x; \theta) = g(T(x); \theta) \cdot h(x)$ , donde



$h$  no depende de  $\theta$ . Entonces

$$\begin{aligned}\lambda(x) &= \frac{\sup_{\Theta_0} L(x; \theta)}{\sup_{\Theta} L(x; \theta)} = \frac{\sup_{\Theta_0} \prod_{i=1}^n f(x_i; \theta)}{\sup_{\Theta} \prod_{i=1}^n f(x_i; \theta)} = \frac{\sup_{\Theta_0} \prod_{i=1}^n g(T(x_i); \theta) \cdot h(x)}{\sup_{\Theta} \prod_{i=1}^n g(T(x_i); \theta) \cdot h(x)} \\ &= \frac{\sup_{\Theta_0} \prod_{i=1}^n g(T(x_i); \theta)}{\sup_{\Theta} \prod_{i=1}^n g(T(x_i); \theta)} = \frac{\sup_{\Theta_0} L^*(T(x); \theta)}{\sup_{\Theta} L^*(T(x); \theta)} = \lambda^*(T(x))\end{aligned}$$

□

**Definición 5.2.6.** Sea  $\alpha \in (0, 1)$  y sea  $\mathcal{T}$  un test con  $\Pi_{\mathcal{T}}(\theta)$  su función de potencia. Decimos que  $\mathcal{T}$  es de tamaño  $\alpha$  si  $\sup_{\Theta_0} \Pi_{\mathcal{T}}(\theta) = \alpha$ .

**Definición 5.2.7.** Consideremos el contraste  $\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$ . Un test  $\mathcal{T}$  para dicho contraste se dirá que es uniformemente más potente si  $\Pi_{\mathcal{T}}(\theta) \geq \Pi_{\mathcal{T}^*}(\theta)$  para todo  $\theta \in \Theta_1$  y cualquier test  $\mathcal{T}^*$ .

**Lema 5.2.8** (de Neymann-Pearson). Sea  $X_1, \dots, X_n$  una muestra aleatoria simple de una variable aleatoria  $X$  con densidad  $f(x; \theta)$ ,  $\theta \in \{\theta_1, \theta_2\}$ . Sean  $\alpha \in (0, 1)$ ,  $k > 0$  constante y  $C \subset \mathfrak{X}$  una región tales que

$$\blacksquare P_{\theta}((X_1, \dots, X_n) \in C) = \alpha.$$

■

$$\begin{cases} \frac{L(x_1, \dots, x_n; \theta_0)}{L(x_1, \dots, x_n; \theta_1)} \leq k & \text{si } (x_1, \dots, x_n) \in C \\ \frac{L(x_1, \dots, x_n; \theta_0)}{L(x_1, \dots, x_n; \theta_1)} \geq k & \text{si } (x_1, \dots, x_n) \notin C \end{cases}$$

Entonces un test  $\mathcal{T}$  con región crítica  $C$  es un test uniformemente de tamaño  $\alpha$  para  $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$ .