

# Fundamentos de Inferencia Estadística

## Contenidos de teoría

Francisco Javier Mercader Martínez

## Tema 1: Distribución en el muestreo

### 1) Definición de muestra aleatoria simple

Sea una variable aleatoria  $X$ . Consideramos  $n$  variables aleatorias independientes e idénticamente distribuidas  $X_1, X_2, \dots, X_n$ , que se distribuyen como  $X$ . La variable aleatoria multidimensional  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de  $X$ .

### 2) Definición de estadístico

Cualquier cantidad calculada a partir de las observaciones de una muestra.

### 3) Demostración de la ley de los grandes números

Consideremos una m.a.s  $X_1, X_2, \dots, X_n$  de una distribución  $X$ . Para cualquier función  $g$  que cumple  $E[g^2(X)] < +\infty$ , tenemos que, con probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(X_i) = E[g(X)].$$

### 4) Concepto de distribución muestral de un estadístico

El **concepto de distribución de un estadístico** se refiere al comportamiento probabilístico de un estadístico (como la media, varianza, etc.) calculado a partir de muestras repetidas de una población. Describe valores que puede tomar el estadístico y con qué probabilidad, formando una distribución teórica que depende del tamaño de la muestra y la distribución de la población.

### 5) Obtención de la esperanza y la varianza de la media muestral, para cualquier población.

Llamamos  $\mu = E[X]$  y  $\sigma^2 = \text{Var}(X)$ .

- Tenemos  $E[\bar{X}] = \mu$ , es decir, que el centro de la distribución muestral de  $\bar{X}$  coincide con el centro de la distribución  $X$ .
- Tenemos,  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$ , es decir, la dispersión de la distribución muestral de  $\bar{X}$  es  $\sqrt{n}$  veces más pequeña que la dispersión inicial de  $X$ .

### 6) Obtención de la esperanza de la varianza muestral.

Si  $(X_1, X_2, \dots, X_n)$  es una muestra aleatoria simple de  $X$  con varianza  $\sigma_X^2$ ,

$$\mathbb{E}[S_n^2] = \sigma_X^2.$$

### 7) Demostración del teorema de Fisher.

Consideramos una m.a.s de una v.a  $X$  con distribución normal  $\mathcal{N}(\mu, \sigma^2)$ , entonces se verifica:

- 1)  $\bar{X}_n$  y  $S_n^2$  son dos vv.aa. independientes.

$$2) \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

$$3) (n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2.$$

### 8) Obtención de la distribución muestral de la proporción muestral.

Tenemos que  $N = X_1 + X_2 + \dots + X_n$ , por lo tanto

$$\hat{p} = \frac{N}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}.$$

Por el Teorema Central del Límite:

$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \text{ aproximadamente.}$$

## Tema 2: Estimación

### 1) Descripción del método de los momentos para la estimación de parámetros.

- Experimento con una variable aleatoria  $X$ , suponemos  $f_X \in \{x \mapsto f_\theta(x), \theta \in \Theta\}$ .
- El parámetro  $\theta$  tiene dimensión  $p$ .
- Consideramos una muestra aleatoria simple  $X_1, \dots, X_n$  de  $X$ .

### 2) Definición de la función de verosimilitud.

Sea  $X$  una v.a. con distribución  $x \mapsto f_X(x; \theta)$  (función de densidad o función puntual de probabilidad), donde  $\theta$  es de dimensión  $p : \theta \in \Theta \subset \mathbb{R}^p$ .

Para un valor concreto de una m.a.s.  $(X_1, \dots, X_n)$ , que denotamos  $(x_1, \dots, x_n)$ , consideramos la función de  $\theta$ :

$$L_n : \begin{cases} \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}^+ \\ \theta \mapsto L_n(\theta) = F_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta). \end{cases}$$

La función  $L_n$  es la **función de verosimilitud**. Nos dice lo creíbles (verosímiles) que son las observaciones para ese valor del parámetro.

### 3) Descripción del método de máxima verosimilitud para la estimación de parámetros.

El **método de máxima verosimilitud** es un procedimiento estadístico para estimar los parámetros de un modelo basado en datos observados. Consiste en contrar los valores de los parámetros que maximizan la función de verosimilitud, es decir, que hacen más probable observar los datos bajo el modelo asumido.

### 4) Obtención del estimador de máxima verosimilitud de $(\mu, \sigma^2)$ para una m.a.s de una $\mathcal{N}(\mu, \sigma^2)$ (solo se pide ver cuáles son las soluciones de las ecuaciones de las primeras derivadas parciales igualadas a cero).

La función de densidad de una variable aleatoria  $X$  con distribución normal es:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Si tenemos una muestra de tamaño  $n$ , es decir,  $X_1, \dots, X_n$  que son observaciones independientes e idénticamente distribuidas como  $\mathcal{N}(\mu, \sigma^2)$ , entonces la función de verosimilitud  $L(\mu, \sigma^2)$  es el producto de las funciones de densidad de cada observación

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Para simplificar los cálculos se toma el logaritmo de la función de verosimilitud, lo que da la función de log-verosimilitud  $\ell(\mu, \sigma^2)$ :

$$\begin{aligned}\ell(\mu, \sigma^2) &= \log L(\mu, \sigma^2) = \sum_{i=1}^n \log \left( \frac{1}{2\pi\sigma^2} \right) + \sum_{i=1}^n \log \left( \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right) \\ &= \sum_{i=1}^n \log \left( (2\pi\sigma^2)^{-\frac{1}{2}} \right) + \sum_{i=1}^n -\frac{(x_i - \mu)^2}{2\sigma^2} = \sum_{i=1}^n -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (x_i - \mu)^2 \\ &= -\frac{n}{2} \cdot \log(\sigma^2) - \frac{n}{2} \cdot \log(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\end{aligned}$$

Para encontrar los estimadores de máxima verosimilitud, derivamos  $\ell(\mu, \sigma^2)$  con respecto a  $\mu$  y  $\sigma^2$ , e igualamos a cero:

$$\begin{aligned}\frac{\partial}{\partial \mu} \ell(\mu, \sigma^2) &= -\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \longrightarrow \sum_{i=1}^n (x_i - \mu) = 0 \longrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ \frac{\partial}{\partial \sigma^2} \ell(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \longrightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \longrightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2\end{aligned}$$

##### 5) Definición de estimador insesgado.

Si el sesgo  $(E_\theta[\hat{\theta}] - \theta)$  de un estimador es nulo para todo valor de  $\theta$ , decimo que el estimador es **insesgado**.

##### 6) Definición del error cuadrático medio para un estimador.

El **error cuadrático medio del estimador**  $\hat{\theta}$  es la función de  $\theta$  definida por

$$\theta \mapsto E_\theta[(\hat{\theta} - \theta)^2].$$

##### 7) Definición de estimador consistente.

Una sucesión de estimadores  $(\hat{\theta}_n)_n$  es consistente si converge hacia  $\theta$  en probabilidad, para todo  $\theta$ . Es decir:

$$\forall \epsilon > 0, \quad P_\theta \left[ |\hat{\theta}_n - \theta| > \epsilon \right] \xrightarrow{n \rightarrow \infty} 0$$

##### 8) Definición del estimador tipo núcleo.

Cualquier estimador de una función de densidad, con una función núcleo  $K$ , recibe el nombre de **estimador tipo núcleo**.

##### 9) Descripción del cálculo del error estándar mediante Bootstrap.

- A partir de la muestra original de tamaño  $n$ , se generan múltiples muestra  $B$  mediante remuestreo con reemplazo (pueden aparecer valores repetidos). Cada muestra tiene el mismo tamaño  $n$  que la original.
- Para cada muestra Bootstrap, se calcula el estadístico de interés (la media, la mediana, etc.).
- Se obtiene la distribución de los valores del estadístico calculados en las  $B$  muestras Bootstrap.
- El error estándar del estadístico se calcula como la desviación estándar de los valores del estadístico en la distribución Bootstrap:

$$\hat{se} = \sqrt{\frac{1}{B} \sum_{i=1}^B (\hat{\theta}_i - \bar{\theta})^2},$$

donde  $\hat{\theta}_i$  es el estadístico calculado para la  $i$ -ésima muestra Bootstrap y  $\bar{\theta}$  es su promedio.

## Tema 3: Estimación por intervalos

### 1) Interpretación del nivel de confianza para la estimación por intervalos.

- El intervalo de confianza obtenido, es un intervalo **aleatorio**.
- Al extraer una muestra, tengo una probabilidad  $\alpha$  de que, al afirmar que  $\mu$  se encuentra en dicho intervalo de confianza, me equivoque.

### 2) Factores que afectan a la precisión del intervalo.

Utilizando, como ejemplo, el margen de error  $z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$  de una distribución Normal:

- Al aumentar el valor de  $n$ , la precisión aumenta.
- Al aumentar el valor de  $\sigma$ , la precisión disminuye.
- Al incrementar el intervalo de confianza, la precisión disminuye.

### 3) Construcción de un intervalo de confianza a partir de un estadístico pivotal.

### 4) Intervalo de confianza basado en percentiles Bootstrap

## Tema 4: Contrastes de hipótesis

### 1) Definición de un test.

Un **test de hipótesis** es una regla que nos lleva a decidir si lo observado en la muestra es compatible con una hipótesis sobre los parámetros del modelo.

### 2) Definición de los errores de tipo I y II.

- Error de tipo I: rechazar la hipótesis nula  $H_0$  cuando es cierta.
- Error de tipo II: aceptar la hipótesis nula  $H_0$  cuando es falsa.

### 3) Definición de potencia de un test.

La **función potencia** de un test  $\delta$  asocia a cada  $\theta$  la probabilidad de rechazar la hipótesis nula:

$$\theta \mapsto \pi_\delta(\theta) = P_\theta(X \in S_1), \quad \text{con } \theta \in \Theta.$$

### 4) Construcción de contrastes de hipótesis uniformes de máxima potencia para hipótesis nula y alternativa simples mediante el teorema de Neyman-Pearson.

### 5) Demostración del teorema de Neyman-Pearson.

Sea  $\delta^*$  el test definido anteriormente.

- La condición

$$\alpha = P(\mathbf{X} \in S_1 | \theta = \theta_0)$$

nos asegura que  $\delta^* \in T_\alpha$ . Sea ahora  $\delta \in T_\alpha$  cualquiera. Tenemos que probar que:

$$\pi_{\delta^*}(\theta_1) \geq \pi_\delta(\theta_1)$$

supondremos  $X$  variable aleatoria continua (en el caso discreto se razonará de forma análoga).

- Sea  $\mathbf{X} \in \text{sop}(X)$  se verifica que:

$$(\delta^*(\mathbf{X}) - \delta(\mathbf{X}))(L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0)) \geq 0.$$

- Si  $\delta(\mathbf{X})^* = 1$  entonces:

$$\frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)} \geq k \longleftrightarrow L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0) \geq 0$$

y  $\delta^*(\mathbf{X}) - \delta(\mathbf{X}) \geq 0$  ( $\delta(\mathbf{X})$  sólo puede ser 0 o 1) y el producto es no negativo.

- Si  $\delta(\mathbf{X})^* = 0$  entonces:

$$\frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)} < k \longleftrightarrow L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0) < 0$$

y  $\delta^*(\mathbf{X}) - \delta(\mathbf{X}) \leq 0$  ( $\delta(\mathbf{X})$  sólo puede ser 0 o 1) y el producto es no negativo.

- La integral de la expresión anterior sobre el  $\text{sop}(\mathbf{X})$  será no negativa:

$$\int_{\text{sop}(\mathbf{X})} (\delta^*(\mathbf{X}) - \delta(\mathbf{X}))(L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0)) d\mathbf{x} \geq 0$$

$$\int_{\text{sop}(\mathbf{X})} \delta^*(\mathbf{X})L(\mathbf{x}, \theta_1) d\mathbf{x} - \int_{\text{sop}(\mathbf{X})} \delta(\mathbf{X})L(\mathbf{x}, \theta_1) d\mathbf{x} - k \int_{\text{sop}(\mathbf{X})} \delta^*(\mathbf{X})L(\mathbf{x}, \theta_0) d\mathbf{x} - \int_{\text{sop}(\mathbf{X})} \delta(\mathbf{X})L(\mathbf{x}, \theta_0) d\mathbf{x} \geq 0$$

notar que ambos contrastes solo toman el valor 1 en sus respectivas regiones de rechazo (y cero en las regiones de aceptación) con lo que al desarrollar los terminos de la integral tenemos:

$$\pi_{\delta^*} - \pi_{\delta}(\theta_1) - k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \geq 0.$$

- Como  $\delta \in T_{\alpha}$  tenemos que  $\pi_{\delta}(\theta_0) \leq \alpha$  con lo que:

$$0 \leq \alpha - \pi_{\delta}(\theta_0) = \pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)$$

y por lo tanto:

$$k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \geq 0$$

con lo que:

$$0 \leq \pi_{\delta^*}(\theta_1) - \pi_{\delta}(\theta_1) - k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \leq \pi_{\delta^*}(\theta_1) - \pi_{\delta}(\theta_1)$$

de donde se deduce:

$$\pi_{\delta^*}(\theta_1) \geq \pi_{\delta}(\theta_1)$$

que era lo que queríamos probar.

## 6) Definición del p-valor. Uso correcto del p-valor e interpretación erróneas.

Consideramos un contraste, un test, y hemos observado una muestra concreta  $\mathbf{x}$ .

- El **p-valor** es el nivel de significación más pequeño que nos lleve, para esa muestra  $\mathbf{x}$ , a rechazar la hipótesis nula.
- El p-valor está asociado a la región  $S_1$  más pequeñas que nos lleve, para esa muestra, a rechazar  $H_0$ .

## 7) Definición del estadístico chi-cuadrado de Pearson, en el caso en que las frecuencias esperadas son conocidas.

Para constatar

$$H_0 : P(A_i) = p_i, i = 1, \dots, k$$

$$H_1 : \text{Existe } i \text{ tal que } P(A_i) \neq p_i.$$

Karl Pearson propuso el estadístico

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}.$$

Si  $n$  es suficientemente grande,

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} \text{ es aproximadamente } \chi_{k-1}^2.$$

## Tema 5: Inferencia Bayesiana

### 1) Descripción del enfoque Bayesiano para la estimación de un parámetro. Distribución a priori y a posteriori.

- La inferencia bayesiana asume que modelizamos la información que podemos tener sobre el parámetro a través de una distribución de probabilidad.
- Formulamos como parte del modelo una **distribución a priori** sobre  $\theta$ . Su función de densidad se denota por  $\pi(\theta)$ .
- La inferencia bayesiana actualiza la información a priori, usando para ello la muestra  $\mathbf{x}$  observada, para obtener la **distribución a posteriori**  $\pi(\theta|\mathbf{x})$ .

### 2) Demostración de que las distribuciones beta y Bernoulli son conjugadas.

Supongamos que el parámetro  $p$  tiene una distribución Beta prior:

$$f(p) = \frac{1}{\text{Be}(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}.$$

Dado que  $X_1, X_2, \dots, X_n$  son observaciones independientes y siguen una distribución Bernoulli, la función de verosimilitud es:

$$P(\mathbf{X}|p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i},$$

donde  $x_i \in \{0, 1\}$ .

Esto se puede reescribir como:

$$P(\mathbf{X}|p) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i}.$$

La distribución a posteriori es proporcional al producto de la verosimilitud y la función a priori:

$$f(p|\mathbf{X}) \propto P(\mathbf{X}|p)f(p).$$

Sustituyendo las expresiones:

$$f(p|\mathbf{X}) \propto p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \cdot p^{\alpha-1} (1-p)^{\beta-1}.$$

Combinando los exponentes de  $p$  y  $1-p$ :

$$f(p|\mathbf{X}) \propto p^{\alpha + \sum_{i=1}^n x_i - 1} (1-p)^{\beta + n - \sum_{i=1}^n x_i - 1}.$$

Por lo tanto:

$$f(p|\mathbf{X}) = \text{Be} \left( \alpha + \sum_{i=1}^n x_i, \beta + n - \sum_{i=1}^n x_i \right).$$