

Tema 1

Introducción a la Inferencia Estadística. Estadísticos muestrales

El objetivo de este tema es introducir una serie de conceptos básicos que serán fundamentales para el desarrollo y estudio posterior, a lo largo de los siguientes temas, de la Inferencia Estadística. Previo a ello se va a hacer una pequeña introducción donde se recordarán los principales elementos asociados al estudio del Cálculo de Probabilidades, herramienta imprescindible para el estudio de la Inferencia.

1.1. Introducción

El Cálculo de Probabilidades proporciona una teoría matemática que permite analizar las propiedades de los fenómenos en los que interviene el azar. Para ello utiliza como modelo básico, común a cualquier situación aleatoria, el concepto de espacio probabilístico, (Ω, \mathcal{A}, P) , junto con el concepto de variable aleatoria definido sobre él.

1.1.1. Conceptos básicos del Cálculo de Probabilidad

- *Espacio de probabilidad o probabilístico:* (Ω, \mathcal{A}, P)
 - Ω : conjunto arbitrario.
 - \mathcal{A} : clase de subconjuntos de Ω , con estructura de σ -álgebra,
 1. $\Omega \in \mathcal{A}$,
 2. $A \in \Omega/A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$,
 3. Sean $A_n \subset \Omega/A_n \in \mathcal{A}, \forall n \implies \bigcup_{i=1}^{\infty} A_n \in \mathcal{A}$ y $\bigcap_{i=1}^{\infty} A_n \in \mathcal{A}$.
 - $P: \mathcal{A} \longrightarrow [0, 1]$ función de probabilidad,
 1. $\forall A \in \mathcal{A}, P(A) \geq 0$,

2. $P(\Omega) = 1$,

3. Sean $A_n \subset \mathcal{A}$ incompatibles dos a dos, entonces $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{i=1}^{\infty} P(A_n)$.

■ *Variable aleatoria*

$X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B}, P_X)$, función medible definida sobre un espacio de probabilidad. La función de probabilidad P_X es lo que se denomina distribución de probabilidad de X .

■ *Función de distribución de X : $F_X: \mathbb{R} \rightarrow [0, 1]$*

$$F_X(x) = P_X((-\infty, x]) = P[X^{-1}((-\infty, x])] = P[X \leq x]$$

Propiedades:

1. No decreciente
2. Continua a la derecha
3. $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x) = 0$ y $F_X(+\infty) = \lim_{x \rightarrow +\infty} F_X(x) = 1$

Teorema de correspondencia: Toda función $F: \mathbb{R} \rightarrow [0, 1]$ cumpliendo las tres propiedades anteriores es la función de distribución de una variable aleatoria.

■ *Clasificación de variables aleatorias*

- *Discretas:* toman valores en un conjunto finito o infinito numerable.

La función de distribución es creciente a saltos.

Función masa de probabilidad: $p_X(x) = P[X = x]$

- *Continuas:* toman valores en un conjunto infinito no numerable.

La función de distribución es creciente y continua.

Función de densidad: $f_X = F'_X(x)$

■ *Independencia de variables aleatorias*

X_1, \dots, X_n , v.a. definidas sobre (Ω, \mathcal{A}, P) , son independientes si y sólo si

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n),$$

$$\forall x_1, \dots, x_n \in \mathbb{R}$$

El disponer de un conjunto de observaciones acerca del fenómeno considerado, en lugar del espacio probabilístico, hace abandonar el Cálculo de Probabilidades, para introducirse en el terreno de la Estadística Matemática o Inferencia Estadística, cuya finalidad es obtener información acerca de la ley de probabilidad de un fenómeno, a partir de una observación no exhaustiva del mismo.

1.1.2. Definiciones

Existen multitud de definiciones distintas para el concepto *Estadística*, como por ejemplo: “proceso de recolección de datos u observaciones, así como su tratamiento numérico; etapas que comportan considerable complicaciones técnicas”.

Otra opción es la definición que da Vic Barnett (1982): “la Estadística es la ciencia que nos indica como debe ser utilizada la información, en orden a reflejar y dar una guía de acción en situaciones prácticas que envuelven incertidumbre”.

Por último, según la RAE: “la Estadística es una rama de las Matemáticas que utiliza grandes conjuntos de datos numéricos para obtener inferencias basadas en el cálculo de probabilidades”.

Por lo tanto, la Estadística no sólo debe encargarse de recoger la información (datos u observaciones), sino que además debe obtener conclusiones basándose en ella. A la parte de la Estadística que se encarga de esa última tarea se la denomina *Inferencia Estadística*.

Dicho de otra forma, el objetivo de la Inferencia Estadística es el análisis e interpretación de la información que aportan las observaciones, como método para obtener conclusiones sobre la ley de probabilidad asociada al fenómeno en estudio.

La Inferencia Estadística se subdivide en dos áreas de estudio denominadas Inferencia Clásica o Frecuentista e Inferencia Bayesiana. La principal diferencia conceptual entre ambas es su interpretación de lo que significa una probabilidad. Por otro lado, la Inferencia Bayesiana, en contraste a la Clásica, le da la posibilidad al investigador de incorporar la información previa que puede poseer sobre el fenómeno.

A lo largo de esta asignatura se estudiará principalmente la Inferencia Clásica con idea de que el alumno adquiera el conocimiento de los fundamentos básicos de esta materia a un nivel elemental, aunque con una perspectiva amplia (Temas del 1 al 9). Finalmente, se introducirá al alumno en la Inferencia Bayesiana para completar sus conocimientos sobre la Inferencia Estadística (Tema 10).

1.2. Planteamiento de un problema de inferencia

Bajo el punto de vista clásico, un problema de Inferencia Estadística consiste en obtener conclusiones acerca del comportamiento de una o varias características de una determinada población, basándose en la observación de las mismas en un subconjunto de la población.

En todo problema de inferencia estadística existen una serie de elementos base del problema:

- La población bajo estudio (*Población*): Conjunto de elementos en el que se pretende estudiar una determinada característica. (Ω)

- La característica que se desea estudiar (*Característica*): Se suele representar por la v.a. que la cuantifica. (X)
- La muestra de que se dispone para el estudio (*Muestra*): Subconjunto de la población sobre el que se va a estudiar la característica para inferir las conclusiones sobre la misma a la población total.

1.2.1. Modelización estadística

De una manera genérica la distribución desconocida F de la variable aleatoria involucrada en un problema de Inferencia Estadística, recibe el nombre de *distribución teórica* o *distribución de la población*. Según el conocimiento que se tenga sobre la distribución teórica, se pueden planear dos situaciones:

- La forma de la función de distribución teórica es conocida, salvo el valor de uno o varios parámetros, es decir, se sabe que F pertenece a una familia de funciones de distribuciones

$$F \in \{F_\theta, \theta \in \Theta\}$$

compuesta por distribuciones de forma funcional fija y conocida, dependientes de uno o varios parámetro θ , que varían dentro de un subconjunto Θ de \mathbb{R}^k , denominado *espacio paramétrico*.

- No se conoce nada acerca de la función de distribución teórica, salvo cosas de tipo muy general, como por ejemplo que la v.a. es discreta o continua, o que existen o no momentos, pero se desconoce la forma de F .

En el primer caso, se estaría considerando un problema de *Inferencia estadística paramétrica* ya que se asume un modelo paramétrico.

En el segundo caso, se estaría considerando un modelo no paramétrico, y por tanto, sería un problema de *Inferencia estadística no paramétrica*.

Dentro del estudio de la Inferencia Clásica, se estudiará principalmente la Inferencia paramétrica, es decir, se hará inferencia sobre el/los parámetros θ , ya que es lo único que falta por conocer para determinar la distribución teórica (Temas del 1 al 8). En el tema 9 se estudiará una introducción a la Inferencia Clásica no paramétrica.

1.3. Muestra aleatoria simple

Como ya se ha comentado anteriormente el objetivo de la inferencia estadística es inferir sobre una población el estudio de una característica a partir del mismo realizado sobre una muestra o subcolección de la población. Para poder realizar correctamente dicha

extensión a toda la población, la muestra considerada para el estudio debe ser aleatoria y representativa de toda la población.

El procedimiento de selección de la muestra puede conducir a diferentes tipos de muestras, pero en esta asignatura sólo se va a considerar el muestreo con reemplazamiento.

Este tipo de muestreo consiste en seleccionar, por mecanismos aleatorios, los elementos de la población que entran a formar parte de la muestra, pero de tal forma que cuando se observa la característica que se está investigando del elemento seleccionado, éste se devuelve a la población y se selecciona el siguiente elemento de entre todos los elementos de la población. Este procedimiento permite que un elemento de la población puede ser seleccionado en más de una ocasión para formar parte de una muestra.

Para llegar al concepto de *muestra aleatoria simple* y poder dar una definición rigurosa de la misma se va a considerar una población, con función de distribución teórica F de la variable aleatoria X .

Para seleccionar una muestra aleatoria de esta población se van llevando a cabo repeticiones del experimento aleatorio que da lugar a la variable aleatoria X , siempre bajo las mismas condiciones, y anotando en cada una de ellas los valores de X . Se obtendrá así un conjunto de valores numéricos x_1, x_2, \dots, x_n que constituyen lo que se denomina una muestra aleatoria simple de Ω , el conjunto de posibles valores de la v.a. X . El número n de repeticiones efectuadas, y de observaciones obtenidas, se denomina *tamaño de la muestra*.

Si se consideran las n observaciones de X como (x_1, x_2, \dots, x_n) , es decir un vector de dimensión n , se pueden ver como la observación del vector aleatorio (X_1, X_2, \dots, X_n) compuesto por n v.a. independientes e idénticamente distribuidas a la v.a. X .

Al conjunto de los posibles valores del vector (x_1, x_2, \dots, x_n) se le denomina *espacio muestral* o conjunto de posibles valores de la muestra y se le denota por \mathcal{X}^n .

Una **muestra aleatoria simple**, de tamaño n , de una variable aleatoria X con distribución teórica F , es un vector (X_1, \dots, X_n) formado por n variables aleatorias independientes e idénticamente distribuidas, con distribución común F .

Al ser las v.a. X_1, \dots, X_n independientes, se tiene que la función de distribución conjunta del vector aleatorio formado por dichas variables será igual al producto de las distribuciones marginales de cada una de ellas. Además, al ser todas las variables de la muestra idénticamente distribuidas a X se tiene que la *función de distribución conjunta* de (X_1, \dots, X_n) es

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n) = \prod_{i=1}^n F_X(x_i), \quad (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Si X es una v.a. discreta, entonces la *función masa de probabilidad conjunta* de

(X_1, \dots, X_n) es

$$P[X_1 = x_1, \dots, X_n = x_n] = \prod_{i=1}^n P[X = x_i], \quad (x_1, \dots, x_n) \in \chi^n.$$

Si X es una v.a. continua, entonces la *función de densidad conjunta* de (X_1, \dots, X_n) es

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i), \quad (x_1, \dots, x_n) \in \chi^n.$$

Ejemplo: Obtener la función masa de probabilidad conjunta de una m.a.s. de $X \rightsquigarrow B(k_0, p)$ y la función de densidad conjunta de una m.a.s. de $X \rightsquigarrow U(a, b)$.

1.4. Función de distribución muestral

Se puede asociar a cada muestra aleatoria simple, (X_1, \dots, X_n) , de una v.a. X con función de distribución F , una *distribución muestral* F_{X_1, \dots, X_n}^* , que emule a F a partir únicamente de la información contenida en la muestra.

Puesto que $F(x) = P[X \leq x]$, para cada $x \in \mathbb{R}$, la distribución muestral se define como

$$F_{X_1, \dots, X_n}^*(x) = \frac{\text{n}^\circ \text{ de variables } X_i \leq x}{n}, \quad \forall x \in \mathbb{R}.$$

La función de distribución muestral se puede escribir también como

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

donde sumando $I_{(-\infty, x]}(X_i)$ es la siguiente v.a.

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

Propiedades:

- Para cada realización muestral, $(x_1, \dots, x_n) \in \mathcal{X}^n$, F_{x_1, \dots, x_n}^* es una función de distribución en \mathbb{R} . En particular es una función a saltos, con saltos de amplitud $1/n$ en los sucesivos valores muestrales ordenados de menor a mayor, supuestos que sean distintos, y de saltos múltiples en el caso de que varios valores muestrales coincidieran.

- $\forall x \in \mathbb{R}$, $F_{X_1, \dots, X_n}^*(x)$ es una variable aleatoria tal que $nF_{X_1, \dots, X_n}^*(x) \rightsquigarrow B(n, F(x))$:

$$E[F_{X_1, \dots, X_n}^*(x)] = F(x), \quad \text{Var}[F_{X_1, \dots, X_n}^*(x)] = \frac{F(x)(1 - F(x))}{n}.$$

- Para valores grandes de n , en virtud del Teorema Central del Límite:

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

Ejemplo: Dada una muestra aleatoria formada por las observaciones (3, 8, 5, 4, 5), obtener su función de distribución muestral y realizar la representación gráfica.

La distribución muestral asociada a una realización, $F_{x_1, \dots, x_n}^*(x)$ refleja exclusivamente los valores numéricos contenidos en la muestra concreta a la que está asociada y no tiene relación directa con la distribución de la población F , en el sentido de que una muestra determinada tendría la misma distribución muestral fuese cual fuese la población de origen. Pese a ello, es muy razonable esperar que la distribución muestral proporcione una imagen aproximada de la distribución de la población de la que se haya extraído la muestra. Es más, la función de distribución muestral verifica el siguiente teorema:

Teorema de Glivenko-Cantelli: Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de v.a. independientes e idénticamente distribuidas (i.i.d.) con función de distribución común F . Si F_{X_1, \dots, X_n}^* es la función muestral asociada a la m.a.s. (X_1, \dots, X_n) , se verifica que F_{X_1, \dots, X_n}^* converge casi seguramente y uniformemente a la función de distribución de X , F .

$$P \left\{ \lim_{n \rightarrow +\infty} \sup_{x \in \mathbb{R}} |F_{X_1, \dots, X_n}^*(x) - F(x)| = 0 \right\} = 1.$$

Con probabilidad 1, al tomar sucesivas observaciones independientes de la variable y considerar las correspondientes funciones de distribución muestrales:

$$\forall \epsilon > 0 \exists n_\epsilon / n > n_\epsilon \Rightarrow F(x) \in (F_{X_1, \dots, X_n}^*(x) - \epsilon, F_{X_1, \dots, X_n}^*(x) + \epsilon)$$

Lo cual, a efectos prácticos, implica que cuando el tamaño de la muestra crece la gráfica de la función de distribución empírica se aproxima bastante a la de la función de distribución de la población.

1.5. Estadístico muestral

Anteriormente se ha indicado que para conocer más acerca de la función de distribución de la variable X que se este estudiando, se toma una m.a.s. de la misma, (X_1, \dots, X_n) y

se observan los valores que toma la muestra. Puede resultar, en ocasiones, más cómodo anotar, en vez de todo el vector de la muestra, una función de dichos resultados, como por ejemplo la suma de los mismo, $\sum_{i=1}^n X_i$. A dicha función de la m.a.s. es lo que se va a denominar *estadístico muestral* cuya ventaja es que en vez de trabajar con la muestra, que es un vector aleatorio de dimensión n , se trabaja con una variable aleatoria o con un vector de menor dimensión. Más formalmente su definición sería:

Sea (X_1, \dots, X_n) una m.a.s. de una v.a. X . Sea $T : (\mathbb{R}^n, \mathcal{B}^n) \longrightarrow (\mathbb{R}^k, \mathcal{B}^k)$ una función medible ($T^{-1}(B) \in \mathcal{B}^n$, $\forall B \in \mathcal{B}^k$) e independiente de cualquier parámetro desconocido. A $T(X_1, \dots, X_n)$ se le denomina **estadístico muestral** asociado a la v.a. X .

Como T es una función medible, el estadístico muestral $T(X_1, \dots, X_n)$ es una v.a. de dimensión k . La función T basta con que este definida en el espacio muestral \mathcal{X}^n y no en todo \mathbb{R}^n .

Ejemplos:

1. Media muestral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
2. Varianza muestral: $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$
3. Cuasivarianza muestral: $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
4. Estadísticos ordenados:
 $X_{(1)} = \min(X_1, \dots, X_n)$
 $X_{(n)} = \max(X_1, \dots, X_n)$

El resto de los estadísticos ordenados, $X_{(2)}, \dots, X_{(n-1)}$ son los estadísticos muestrales que se quedan con el valor que ocupa la posición que indica su subíndice después de ordenar de menor a mayor todos los valores, es decir, por ejemplo $X_{(2)}$ dará como resultado el segundo valor más pequeño de X_1, \dots, X_n .

Ejemplo: Sea X una v.a. con distribución $B(1, p)$ con $p \in (0, 1)$. Se toma una muestra de tamaño 5, $(X_1, X_2, X_3, X_4, X_5)$, y se obtiene la siguiente observación $(0, 1, 1, 0, 0)$. Determinar el valor de los estadísticos estudiados en la observación.

1.5.1. Distribución en el muestreo de un estadístico

Ya que el estadístico muestral $T(X_1, \dots, X_n)$ es una v.a. de dimensión k , tiene una función de distribución asociada, la cual no siempre será fácil de conocer.

Se denomina *distribución en el muestreo de un estadístico* T , definido en el espacio muestral $(\mathcal{X}^n, \mathcal{B}^n)$, a la distribución de la v.a. $T(X_1, \dots, X_n)$.

Media muestral

En general, determinar la distribución del estadístico no será fácil, pero en casos en los que el estadístico sea suma de las variables, como el estadístico media muestral, puede resultar asequible. Para ello se aplica que la función generatriz de momentos, que define de forma única a una v.a., verifica la propiedad de adición, es decir, que la f.g.m. de la suma de n v.a. independientes coincide con el producto de las n f.g.m. Por lo tanto, para el estadístico media muestral se verifica:

$$M_{\bar{X}}(t) = (M_X(t/n))^n$$

teniendo en cuenta que las n v.a. de la muestra son independientes e idénticamente distribuidas a X . Identificando la distribución que tiene dicha f.g.m. se tiene la distribución del estadístico.

Ejemplo: Obtener la distribución muestral de \bar{X} para (X_1, \dots, X_n) una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Estadísticos ordenados

En el caso de los estadísticos ordenados,

$$X_{(r)} = \begin{cases} \min(X_1, \dots, X_n) & r = 1; \\ \min(\{X_1, \dots, X_n\} - \{X_{(1)}, \dots, X_{(j-1)}\}), & r = j; \\ \max(X_1, \dots, X_n), & r = n \end{cases}$$

se tiene que la función de distribución muestral es:

$$\begin{aligned} F_{X_{(r)}}(x) &= P[X_{(r)} \leq x] = P[\text{al menos } r \text{ elementos muestrales sean } \leq x] = \\ &= \sum_{i=r}^n \binom{n}{i} (P[X \leq x])^i (P[X > x])^{n-i}, \quad x \in \mathbb{R}. \end{aligned}$$

Por lo tanto, según la v.a. X sea discreta o continua, se tiene que:

- Discreta:

$$P[X_{(r)} = x_{(r)}] = P[X_{(r)} \leq x_{(r)}] - P[X_{(r)} < x_{(r)}] = F_{X_{(r)}}(x_{(r)}) - F_{X_{(r)}}^-(x_{(r)}).$$

- Continua:

$$g_r(x_{(r)}) = \frac{n!}{(r-1)!(n-r)!} [F(x_{(r)})]^{r-1} [1 - F(x_{(r)})]^{n-r} f(x_{(r)}).$$

Ejemplo: Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para $X \rightsquigarrow U(a, b)$.

1.5.2. Estadísticos de Interés

Sea (X_1, \dots, X_n) una m.a.s. de X , v.a. con función de distribución F . Anteriormente se ha estudiado que asociada a la muestra existe una distribución denominada muestral o empírica, $F_n^*(x)$. Además de dicha distribución, sobre la muestra se pueden estudiar una serie de características:

- Momentos muestrales centrados y no centrados:

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \equiv \text{momento no centrado de orden } k \in \mathbb{N}$$

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \equiv \text{momento centrado de orden } k \in \mathbb{N}$$

En particular se tiene:

$$A_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \text{ media muestral}$$

$$B_1 = 0$$

$$B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \text{ (Varianza muestral).}$$

- Cuantiles muestrales: Para cada $p \in (0, 1)$, el cuantil de orden p , c_p , es un valor real tal que

$$F_n^*(c_p) \geq p \text{ y } F_n^*(c_p^-) \leq p$$

Se pueden expresar de la siguiente forma en función de los elementos de la muestra ordenada:

- Si $np \in \mathbb{N}$, $c_p = \frac{X_{(np)} + X_{(np+1)}}{2}$.

- En otro caso, sea $[np]$ la parte entera de np , entonces $c_p = X_{([np]+1)}$

- Función generatriz de momentos muestral:

$$M^*(t) = \frac{1}{n} \sum_{i=1}^n e^{tX_i}$$

Esta función se usa para obtener los momentos no centrados, $\left[\frac{\partial^k M^*(t)}{\partial t^k} \right]_{t=0} = A_k$.

Para los momentos no centrados se tiene

- $E[A_k] = \frac{1}{n} \sum_{i=1}^n E[X_i^k] = E[X^k]$, es decir, coincide con el momento de orden k respecto al origen de la distribución teórica. En particular,

$$E[A_1] = \mu$$

- $Var[A_k] = \frac{1}{n^2} \sum_{i=1}^n Var[X_i^k] = \frac{1}{n} Var[X^k] = \frac{1}{n} (E[X^{2k}] - E[X^k]^2)$. En particular,

$$Var[A_1] = \frac{\sigma^2}{n}.$$

Se ha denotado por $\mu = E[X]$ y por $\sigma^2 = Var[X]$.

Para los momentos centrados tiene especial interés el caso de la varianza muestral, $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$:

$$E[B_2] = \frac{(n-1)\sigma^2}{n}.$$

Por tanto, la esperanza de la varianza muestral no coincide con la varianza teórica. Debido a este resultado es frecuente considerar la cuasivarianza muestral, S^2 , en lugar de la varianza, ya que ella si cumple que su esperanza coincide con la varianza teórica:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} B_2 \Rightarrow$$

$$E[S^2] = E\left[\frac{n}{n-1} B_2\right] = \frac{n}{n-1} E[B_2] = \frac{n}{n-1} \frac{(n-1)\sigma^2}{n} = \sigma^2.$$