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, $(\Omega, \mathcal{A}, P)$, junto con el concepto de variable aleatoria definido sobre él.

1.1.1. Conceptos básicos del Cálculo de Probabilidad

- *Espacio de probabilidad o probabilístico*: $(\Omega, \mathcal{A}, P)$

  - Ω: conjunto arbitrario.
  - $\mathcal{A}$: clase de subconjuntos de Ω, con estructura de $\sigma-$álgebra,
    1. $\Omega \in \mathcal{A}$,
    2. $A \in \Omega / A \in \mathcal{A} \Longrightarrow \overline{A} \in \mathcal{A}$,
    3. Sean $A_n \subset \Omega / A_n \in \mathcal{A}, \forall n \Longrightarrow \bigcup_{i=1}^{\infty} A_n \in \mathcal{A}$ 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: $(\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathbb{B}, P_X)$, 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: $F_X : \mathbb{R} \to [0, 1]$

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

  Propiedades:

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

  Teorema de correspondencia: Toda función $F : \mathbb{R} \to [0, 1]$ 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

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

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

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

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 el estudio científico que se encarga de la recopilación, clasificación y análisis de datos numéricos sobre fenómenos específicos, buscando obtener conclusiones a partir de ellos, frecuentemente mediante el uso de la probabilidad".

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 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.

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.

Éste 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 esta 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 $x_1, x_2, \ldots, x_n$ que constituyen lo que se denomina una muestra de $\mathcal{X}$, 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 $(x_1, x_2, \ldots, x_n)$, es decir un vector de dimensión n, se pueden ver como la observación del vector aleatorio $(X_1, X_2, \ldots, X_n)$ compuesto por n v.a. independientes e idénticamente distribuidas a la v.a. X.

Al conjunto de los posibles valores del vector $(x_1, x_2, \ldots, x_n)$ 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 $(X_1, \ldots, X_n)$ formado por n variables aleatorias independientes e idénticamente distribuidas, con distribución común F.

Al ser las v.a. $X_1, \ldots, X_n$ 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 $(X_1, \ldots, X_n)$ es

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

Si X es una v.a. discreta, entonces la *función masa de probabilidad conjunta* de $(X_1, \ldots, X_n)$ es

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

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

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

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, $(X_1, \ldots, X_n)$, de una v.a. X con función de distribución F, una *distribución muestral* $F^*_{X_1,\ldots,X_n}$, 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

$$F^*_{X_1,\ldots,X_n}(x) = \frac{\text{n}^{\text{o}} \text{ de variables } X_i \leq x}{n}, \quad \forall x \in \mathbb{R}.$$

La función de distribución muestral se puede escribir también como

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

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, $(x_1, \ldots, x_n) \in \mathcal{X}^n$, $F^*_{x_1,\ldots,x_n}$ 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.

- $\forall x \in \mathbb{R}$, $F^*_{X_1,\ldots,X_n}(x)$ es una variable aletoria tal que $nF^*_{X_1,\ldots,X_n}(x) \rightsquigarrow B(n, F(x))$:

$$E[F^*_{X_1,\ldots,X_n}(x)] = F(x), \qquad Var[F^*_{X_1,\ldots,X_n}(x)] = \frac{F(x)(1 - F(x))}{n}.$$

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

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

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, $F^*_{x_1,\ldots,x_n}(x)$ 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 $F^*_{X_1,\ldots,X_n}$ es la función muestral asociada a la m.a.s. $(X_1, \ldots, X_n)$, se verifica que $F^*_{X_1,\ldots,X_n}$ converge casi seguramente y uniformemente a la función de distribución de X, F.

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

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

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

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

Como T es una función medible, el estadístico muestral $T(X_1, \ldots, X_n)$ 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: $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$
2. Varianza muestral: $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$
3. Cuasivarianza muestral: $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$
4. Estadísticos ordenados:

   $X_{(1)} = \text{mín}(X_1, \ldots, X_n)$

   $X_{(n)} = \text{máx}(X_1, \ldots, X_n)$

   El resto de los estadísticos ordenados, $X_{(2)}, \ldots, X_{(n-1)}$ 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 $X_1, \ldots, X_n$.

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 $T(X_1, \ldots, X_n)$ 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. $T(X_1, \ldots, X_n)$.

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 $(X_1, \ldots, , X_n)$ una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

Estadísticos ordenados

En el caso de los estadísticos ordenados,

$$X_{(r)} = \begin{cases} \text{mín}(X_1, \ldots, X_n) & r = 1; \\ \text{mín}(\{X_1, \ldots, X_n\} - \{X_{(1)}, \ldots, X_{(j-1)}\}), & r = j; \\ \text{máx}(X_1, \ldots, X_n), & r = n \end{cases}$$

se tiene que la función de distribución muestral es:

$$F_{X_{(r)}}(x) = P[X_{(r)} \leq x] = P[\text{al menos } r \text{ elementos muestrales sean } \leq x] =$$

$$= \sum_{i=r}^{n} \binom{n}{i} (P[X \leq x])^i \, (P[X > x])^{n-i}, \quad x \in \mathbb{R}.$$

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 $(X_1, \ldots, X_n)$ 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,

$$E[A_1] = \mu$$

.

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

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1} B_2 \Rightarrow \\ E[S^2] &= E\left[\frac{n}{n-1} B_2\right] = \frac{n}{n-1} E[B_2] = \frac{n}{n-1}\frac{(n-1)\sigma^2}{n} = \sigma^2. \end{aligned}$$

Tema 2

Distribuciones en el muestreo de poblaciones normales.

2.1. Introducción

En el tema 1 se obtuvieron algunos resultados relativos al comportamiento de algunas características muestrales como los momentos. En particular, se estudiaron:

- Resultados "exactos", acerca de momentos de la distribución en el muestreo de algunas características muestrales.

- Resultados "de aproximación" asintóticos, sobre el comportamiento límite (para "muestras grandes") de ciertas características.

Exceptuando los ejemplos específicos analizados, todas las consideraciones se hicieron sin presuponer ninguna distribución concreta en la población.

En este tema, se estudian algunos resultados fundamentales de la Inferencia Estadística relativos al muestreo de poblaciones normales, que se refieren a la distribución exacta (para muestras de cualquier tamaño) de estadísticos que surgen de forma natural en problemas concretos básicos de inferencia.

En este sentido, se analizan la distribución χ^2 de Pearson, t de Student y F de Snedecor, desde una doble aproximación:

1. Desde el punto de vista analítico, por particularización o construcción basada en distribuciones conocidas (normal y gamma).

2. Desde el punto de vista muestral, como distribuciones exactas relacionadas con el muestreo de poblaciones normales.

2.2. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor

2.2.1. Distribución χ^2 de Pearson

Se dice que la v.a. X tiene una distribución χ^2 con n ($n \in \mathbb{N}$) grados de libertad si su función de densidad de probabilidad viene dada por

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

y se denota $X \rightsquigarrow \chi^2(n)$[1].

La distribución χ^2 es un caso particular de la distribución Gamma, $\Gamma(p, a)$, cuya densidad es:

$$f(x) = \frac{a^p}{\Gamma(p)} \, x^{p-1} e^{-ax}, \quad x > 0$$

En particular se verifica: $X \rightsquigarrow \chi^2(n) \Leftrightarrow X \rightsquigarrow \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$.

Propiedades:

- FGM: $M(t) = \frac{1}{(1-2t)^{n/2}}, \quad t < 1/2$
- Momentos: $E[X^k] = 2^k \dfrac{\Gamma\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)}$. En particular, se tiene: $E[X] = n$, $E[X^2] = n^2 + 2n$, $Var[X] = 2n$.
- Reproductividad: Si $X_1, \ldots, X_n$ son v.a. independientes tales que $X_i \rightsquigarrow \chi^2(k_i)$. Entonces
$$\sum_{i=1}^{n} X_i \rightsquigarrow \chi^2\left(\sum_{i=1}^{n} k_i\right).$$
- Relación con la distribución normal $\mathcal{N}(0,1)$: Si $X_1, \ldots, X_n$ son v.a.i.i.d. con distribución común $\mathcal{N}(0,1)$. Entonces
$$\sum_{i=1}^{n} X_i^2 \rightsquigarrow \chi^2(n)$$

Tablas y aproximaciones: La distribución χ^2 está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(n, 2n)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

[1] La función Γ se define como $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$. Esta función cumple un par de propiedades: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$; si α es un número entero, $\Gamma(\alpha) = (\alpha-1)!$.

a) $P[\chi^2(10) \geq k] = 0.005$

b) $P[\chi^2(45) \leq k] = 0.005$

c) $P[\chi^2(14) \geq 21.06]$

d) $P[\chi^2(20) \leq 12.44]$

Gráfica de la función de densidad de $\chi^2(n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $n = 1$, $\lim_{x \to 0} f(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $n = 2$, $f(0) = 1/2$ y la función de densidad es estrictamente decreciente.
- Para $n \geq 3$, $f(0) = 0$, crece hasta la moda y luego decrece.

2.2.2. Distribución t de Student

Sean X e Y variables aleatorias independientes con distribuciones, $X \rightsquigarrow \mathcal{N}(0,1)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$T = \frac{X}{\sqrt{Y/n}}$$

se dice que tiene una distribución t de Student con n grados de libertad. Se denota $T \rightsquigarrow t(n)$.

Propiedades:

- Función de densidad:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

  Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $T = X/\sqrt{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $t(n)$, con $n > 1$. Entonces se tiene que existen los momentos $E[X^r]$ para $r < n$ y se verifica que,

  - Si $r < n$ e impar, se tiene que $E[X^r] = 0$.

- Si $r < n$ y par, $E[X^r] = n^{r/2} \dfrac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$.

En particular, para $n > 2$, existen los momentos de primer y segundo orden, $E[X] = 0$ y $E[X^2] = Var[x] = \dfrac{n}{n-2}$.

Tablas y aproximaciones: La distribución t está tabulada para valores de n pequeños. Para n grandes, su distribución se puede aproximar por $\mathcal{N}(0,1)$.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

a) $P[t(26) \geq k] = 0.05$

b) $P[t(20) \leq k] = 0.25$

c) $P[t(26) \geq k] = 0.9$

d) $P[t(21) \geq 1.721]$

e) $P[t(11) \leq 0.697]$

f) $P[t(8) \leq -2.306]$

Gráfica de la función de densidad de $t(n)$: Cumple las siguientes propiedades

- Es similar a la de la $\mathcal{N}(0,1)$, es decir, simétrica alrededor del cero y unimodal.
- Para $n \to +\infty$ se aproxima a la gráfica de la $\mathcal{N}(0,1)$.
- Tiene colas mayores que las de la normal que van reduciéndose y aproximándose a las de la normal conforme n crece.
- Es más aplastada que la de la normal, es decir es platicúrtica, y su zona central va creciendo y aproximándose a la de la normal conforme n crece.

2.2.3. Distribución F de Snedecor

Sean X e Y variables aleatorias independientes con distribuciones $X \rightsquigarrow \chi^2(m)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la v.a.

$$F = \frac{X/m}{Y/n}$$

se dice que tiene una distribución F de Snedecor con (m,n) grados de libertad. Se denota $F \rightsquigarrow F(m,n)$.

Propiedades:

- Función de densidad:
$$g(f) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} f^{m/2-1} \left(1+\frac{m}{n}f\right)^{-\frac{m+n}{2}}, \quad f>0.$$
Para obtener esta densidad se aplica el cambio de variable indicado en la propia construcción de la distribución, $F = \frac{X/m}{Y/n}$, tomando como variable auxiliar $U = Y$.

- Momentos: Sea X una v.a. con distribución $F(m,n)$. Entonces se verifica que,
$$E[X^r] = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\frac{m}{2}+r\right)\Gamma\left(\frac{n}{2}-r\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}, \quad \text{para } 0<r<\frac{n}{2}.$$
En particular, $E[X] = \dfrac{n}{n-2}$ si $n>2$, $E[X^2] = \dfrac{n^2(m+2)}{m(n-4)(n-2)}$ si $n>4$ y $Var[X] = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$ si $n>4$.

- Mediante un cambio de variable se puede comprobar que:
  - $X \rightsquigarrow F(m,n) \Leftrightarrow X^{-1} \rightsquigarrow F(n,m)$.
  - $X \rightsquigarrow t(n) \Leftrightarrow X^2 \rightsquigarrow F(1,n)$.

Tablas y aproximaciones: La distribución F está tabulada para valores de m y n pequeños. Las tablas que se os han proporcionado incluyen aproximaciones para valores grandes de m y n.

Ejemplo: Calcula el valor de k o la probabilidad indicada:

a) $P[F(7,3) \leq k] = 0.95$

b) $P[F(8,4) \geq k] = 0.01$

c) $P[F(2,2) \leq 19]$

d) $P[F(3,5) \geq 12.1]$

e) $P[F(60,40) \leq k] = 0.05$

Gráfica de la función de densidad de $F(m,n)$: Cumple las siguientes propiedades

- Es asimétrica a la derecha y unimodal.
- Para $m=1$, $\lim\limits_{f\to 0} g(x) = +\infty$ y la función de densidad es estrictamente decreciente.
- Para $m=2$, $g(0)=1$ y la función de densidad es estrictamente decreciente.
- Para $m\geq 3$, $g(0)=0$, crece hasta la moda y luego decrece.

2.3. Muestreo en una población normal unidimensional

Sea X una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Se considera una m.a.s. de tamaño n, $X_1, \ldots, X_n$. Sea $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ la media muestral y $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ la cuasivarianza muestral. El objetivo de este apartado es dar las distribuciones de ambos estadísticos muestrales, para lo cual es necesario el siguiente resultado:

Teorema: Sea $X_1, \ldots, X_n$ una m.a.s. de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Entonces, $\bar{X}$ y $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ son independientes.

Corolarios:

1. $\bar{X} \rightsquigarrow \mathcal{N}(\mu, \frac{\sigma^2}{n})$
2. (Lema de Fisher): $\bar{X}$ y S^2 son independientes.
3. $\dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$
4. $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$

2.3.1. Esquemas de resultados para una muestra y uso en inferencia

$$\text{- Inferencia sobre } \mu : \begin{cases} \sigma_0^2 \text{ conocida} & \dfrac{\bar{X} - \mu}{\sigma_0/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1) \\[2ex] \sigma^2 \text{ desconocida} & \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n-1) \end{cases}$$

$$\text{- Inferencia sobre } \sigma^2 : \begin{cases} \mu_0 \text{ conocida} & \dfrac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\sigma^2} \rightsquigarrow \chi^2(n) \\[2ex] \mu \text{ desconocida} & \dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1) \end{cases}$$

2.4. Muestreo en dos poblaciones normales unidimensionales

Sea X una v.a. con distribución $\mathcal{N}(\mu_1, \sigma_1^2)$, $\mu_1 \in \mathbb{R}$ y $\sigma_1^2 > 0$. Se considera una m.a.s. de tamaño n_1, $X_1, \ldots, X_{n_1}$. Sea $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ la media muestral y $S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i -$

$\bar{X})^2$ la cuasivarianza muestral.

Sea Y otra v.a. con distribución $\mathcal{N}(\mu_2, \sigma_2^2)$, $\mu_2 \in \mathbb{R}$ y $\sigma_2^2 > 0$. Se considera una m.a.s. de tamaño n_2, $Y_1, \ldots, Y_{n_2}$. Sea $\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$ la media muestral y $S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$ la cuasivarianza muestral.

Se supone que $(X_1, \ldots, X_{n_1})$ e $(Y_1, \ldots, Y_{n_2})$ son independientes.

Extensión del Lema de Fisher: Los vectores $(\bar{X}, \bar{Y})$ y (S_1^2, S_2^2) son independientes.

Corolarios:

1. $\dfrac{n_2}{n_1}\dfrac{\sigma_2^2}{\sigma_1^2}\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} \rightsquigarrow F(n_1, n_2)$

2. $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1, n_2-1)$

   En particular, si $\sigma_1 = \sigma_2$, se tiene que $\dfrac{S_1^2}{S_2^2} \rightsquigarrow F(n_1-1, n_2-1)$

3. $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2}}}\sqrt{\dfrac{n_1+n_2-2}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$

   En particular, si $\sigma_1 = \sigma_2$, se tiene que $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}\sqrt{\frac{n_1+n_2-2}{\frac{1}{n_1}+\frac{1}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$

2.4.1. Esquemas de resultados para dos muestra y uso en inferencia

- Inferencia sobre $\mu_1 - \mu_2$ (comparación de medias):

$$\begin{cases} \sigma_1^2, \sigma_2^2 \text{ conocidas} & \dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1) \\ \\ \sigma_1^2 = \sigma_2^2 = \sigma^2 \text{ desconocidas} & \dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}\sqrt{\dfrac{n_1+n_2-2}{\frac{1}{n_1}+\frac{1}{n_2}}} \rightsquigarrow t(n_1+n_2-2) \end{cases}$$

- Inferencia sobre $\dfrac{\sigma_1^2}{\sigma_2^2}$ (comparación de varianzas):

$$\begin{cases} \mu_1, \mu_2 \text{ conocidas} & \dfrac{n_1}{n_2}\dfrac{\sigma_1^2}{\sigma_2^2}\dfrac{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{\sum_{i=1}^{n_1}(X_i-\mu_1)^2} \rightsquigarrow F(n_2, n_1) \\ \\ \mu_1, \mu_2 \text{ desconocidas} & \dfrac{\sigma_1^2}{\sigma_2^2}\dfrac{S_2^2}{S_1^2} \rightsquigarrow F(n_2-1, n_1-1) \end{cases}$$

Tema 3

Suficiencia y completitud

3.1. Estadísticos suficientes

Antes de dar una definición del concepto de estadístico suficiente, se va a dar una idea intuitiva del concepto de suficiencia que facilite su comprensión.

Sea X una v.a. con distribución en una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$ dependientes de un parámetro desconocido θ. El objetivo es inferir el valor de θ, para lo cual se considera una m.a.s. para, en base a la información que proporciona, estimar el valor del parámetro.

Sin embargo, en lugar de considerar la muestra, se va a trabajar con un estadístico de la misma, $T(X_1, \ldots, X_n)$, que resumen la información que proporciona la muestra. Puede suceder que al resumir la información de la muestra se pierda parte relevante de la misma. Por ejemplo, si $T(X_1, \ldots, X_n) = X_1$ se pierde la información que proporcionan $X_2, \ldots, X_n$. En ocasiones dicha pérdida puede no ser relevante, según la distribución de X.

En ese sentido de no perder información relevante surge el concepto de suficiencia gracias a Fisher (1922): "Un estadístico es suficiente cuando contiene toda la información contenida en la muestra sobre el parámetro que se esta considerando"(es decir, basta con usar el estadístico para inferir el valor del parámetro).

Un estadístico puede ser suficiente en una familia de distribuciones y no para otras.

Definición: Sea $(X_1, \ldots, X_n)$ una m.a.s. de $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Un estadístico $T(X_1, \ldots, X_n)$ es *suficiente* para la familia de distribuciones considerada (o suficiente para θ) si la distribución de la muestra condicionada a cualquier valor del estadístico, $T(X_1, \ldots, X_n) = t$, es independiente de θ.

Notas:

- Si $T(X_1, \ldots, X_n) = (X_1, \ldots, X_n)$ no hay pérdida de información. Luego siempre hay un estadístico suficiente trivial, la propia muestra.

- Si se encuentra un estadístico suficiente que no sea la muestra, a partir de entonces se trabajará con él, porque es más fácil de manejar ya que resumen la información de la muestra sin perder información sobre θ.

Ejemplos:

1. Sea X_1, X_2 variables aleatorias independientes con distribución de Poisson de parámetro λ. Probar que $X_1 + X_2$ es suficiente para λ y que $X_1 + 2X_2$ no lo es.
2. En n lanzamientos de una moneda, el número de caras es suficiente para el parámetro p.

En ocasiones puede ocurrir que se disponga de una muestra y se pierda el valor de sus datos. A partir de un estadístico suficiente se puede reconstruir la muestra. Intuitivamente la idea de *reconstrucción de la muestra* es la siguiente:

Sea X una v.a. con distribución en $\{F_\theta, \theta \in \Theta\}$ y $(X_1, \ldots, X_n)$ una m.a.s. de X. Se realiza un experimento que consiste en observar la muestra y se obtienen unos valores concretos u observaciones $(x_1, \ldots, x_n)$. Asociada a la variable X se tiene un estadístico $T(X_1, \ldots, X_n)$, que al aplicarlo a los datos observados se obtiene un valor, $T(x_1, \ldots, x_n) = t$.

Si se pierden los datos de la m.a.s., pero se conoce el valor del estadístico y no se puede volver a observar la variable para obtener datos equivalentes a los anteriores, para reconstruir la muestra se realiza lo siguiente: Se considera una v.a. X^* que va a tener la distribución de la v.a. $X/T = t$, o más concretamente, una m.a.s $(X_1^*, \ldots, X_n^*)$ con la misma distribución que $(X_1, \ldots, X_n)/T = t$.

Si la distribución de $(X_1, \ldots, X_n)/T = t$ es independiente del parámetro θ, por tanto es conocida, se pueden observar las variables $(X_1^*, \ldots, X_n^*)$, que es una muestra equivalente a $(X_1, \ldots, X_n)$, porque las dos muestras conducen al mismo valor de estadístico:

$$T(X_1^*, \ldots, X_n^*) = T(X_1, \ldots, X_n)$$

y para cada valor de t, las distribuciones son las mismas.

Ejemplo: Se supone que se ha lanzado una moneda 100 veces, y se sabe que se han obtenido 60 caras pero se han perdido los datos originales de si salió cara o cruz en cada tirada. Reconstruir la muestra.

3.1.1. Teorema de factorización de Neyman-Fisher

La definición de estadístico suficiente no es constructiva, en el sentido de que sirve para comprobar si un estadístico dado es o no suficiente, pero no indica cómo buscarlo.

El teorema de Factorización de Neyman-Fisher establece un criterio útil para la búsqueda de estadísticos suficientes, así como para probar, con mayor facilidad, si un estadístico es suficiente.

Teorema de factorización:

Sea $(X_1, \ldots, X_n)$ una m.a.s de $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Sea f_θ la función masa de probabilidad o función de densidad de X bajo F_θ y sea f_θ^n la f.m.p. o f.d.d. de la muestra bajo F_θ.

Un estadístico $T(X_1, \ldots, X_n)$ se dice que es suficiente si y sólo si, para cualquier valor de $\theta \in \Theta$,

$$f_\theta^n(x_1, \ldots, x_n) = h(x_1, \ldots, x_n)\, g_\theta(T(x_1, \ldots, x_n)), \quad (x_1, \ldots, x_n) \in \mathcal{X}^n$$

donde h es independiente de θ y g_θ depende de $(x_1, \ldots, x_n)$ sólo a través de $T(x_1, \ldots, x_n)$.

Propiedades

1. Si $T(X_1, \ldots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta\}$, entonces $T(X_1, \ldots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta'\}$, para cualquier $\Theta' \subseteq \Theta$, es decir, si un estadístico es suficiente para una familia de distribuciones, lo es también para cualquier subfamilia suya.

2. Si $T(X_1, \ldots, X_n)$ es suficiente para una familia $\{F_\theta, \theta \in \Theta\}$ y $U(X_1, \ldots, X_n)$ es un estadístico tal que $T = h(U)$, entonces U es suficiente para la misma familia. (Esto no indica que una función de un estadístico suficiente sea suficiente, sino que si tengo un estadístico suficiente que es función de otro estadístico, entonces el otro estadístico también es suficiente).

3. Si $T(X_1, \ldots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta\}$, toda transformación biunívoca de T proporciona también un estadístico suficiente para la misma familia.

El teorema de factorización también se verifica si el parámetro θ es multidimensional, en cuyo caso el estadístico también es multidimensional. En particular, aunque el parámetro sea de dimensión 1, el estadístico suficiente puede tener dimensión mayor que 1. En general se verifica:

$$\text{Dimensión estadístico suficiente} \geq \text{Dimensión parámetro}$$

Ejemplo: Sea X una v.a. con distribución en $\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0\}$. Estudiar la suficiencia en los casos de :

- σ_0^2 conocida.
- μ_0 conocida.
- μ y σ^2 desconocidas.

Suficiencia minimal

Dada una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$, generalmente existirán varios estadísticos suficientes para θ.

Un criterio de selección consiste en elegir estadísticos suficientes minimales (máxima reducción de los datos, sin pérdida de información sobre θ).

Va a existir siempre un estadístico suficiente que es función de todos los demás, ese estadístico es el *estadístico suficiente minimal*.

Una aplicación adecuada del teorema de factorización nos llevará a él (es único) y siempre existe en distribuciones discretas y continuas.

3.2. Familias de distribuciones completas. Estadísticos completos

El concepto de suficiencia se usa frecuentemente en conjunción con el concepto de completitud, sin embargo dicho concepto es menos intuitivo que el de suficiencia. En próximos temas se verá su utilidad cuando se estudien las propiedades de los denominados "estimadores".

Antes de dar la definición de estadístico completo, se va a definir el concepto de familia de distribuciones completa:

Definición: Sea $\{F_\theta, \theta \in \Theta\}$ una familia de distribuciones con f.d.d. o f.m.p. $\{f_\theta(x), \theta \in \Theta\}$. Se dice que dicha *familia* es *completa* si para cualquier función medible unidimensional, g, tal que

$$E_\theta[g(X)] = 0, \quad \forall \theta \in \Theta$$

se tiene que

$$P_\theta[g(X) = 0] = 1, \quad \forall \theta \in \Theta$$

es decir, una familia completa es aquella que no admite funciones medibles no nulas con esperanza cero.

Definición: Un estadístico $T(X_1, \ldots, X_n)$ se dice que es *completo* para la familia de distribuciones de X si para cualquier función medible unidimensional, g, se tiene:

$$E_\theta[g(T(X_1, \ldots, X_n))] = 0, \quad \forall \theta \in \Theta \Rightarrow P_\theta[g(T(X_1, \ldots, X_n)) = 0] = 1, \quad \forall \theta \in \Theta.$$

Como ya se ha comentado, el concepto de estadístico completo suele ir asociado al concepto de estadístico suficiente. Es más, si se pide dar un estadístico suficiente y completo, se suele tomar el estadístico suficiente que proporciona el teorema de factorización y se comprueba si dicho estadístico es completo. Además se tiene el siguiente resultado:

Ejemplos:

1. Sea $(X_1, \ldots, X_n)$ una m.a.s de la v.a. $X \rightsquigarrow \{B(1,p), p \in (0,1)\}$. Encontrar un estadístico suficiente y completo asociado a la muestra.

2. Sea $(X_1, \ldots, X_n)$ una m.a.s de la v.a. $X \rightsquigarrow \{U(0,\theta), \theta > 0\}$. Encontrar un estadístico suficiente y completo asociado a la muestra.

3.3. Suficiencia y completitud en familias exponenciales

Sea $\{F_\theta, \theta \in \Theta\}$ una familia de distribuciones de probabilidad paramétricas con f.d.d. o f.m.p. $\{f_\theta(x), \theta \in \Theta\}$.

3.3.1. Familia exponencial uniparamétrica

Se dice que la familia de distribuciones es *exponencial uniparamétrica* si se cumple que:

1. El espacio paramétrico, Θ, es un intervalo real ($\Theta \subseteq \mathbb{R}$).

2. El conjunto de valores de la variable no depende de θ:
$$\{x, f_\theta(x) > 0\} = \chi, \ \forall \theta \in \Theta.$$

3. Existen funciones real-valuadas $Q(\theta)$ y $D(\theta)$, definidas sobre Θ, y existen funciones medibles Borel T y S, también real valuadas, tales que
$$\forall \theta \in \Theta, \ f_\theta(x) = \exp\left[Q(\theta)T(x) + D(\theta) + S(x)\right], \ x \in \chi.$$

Ejemplos:

1. La $\{\mathcal{P}(\lambda), \lambda > 0\}$ es una familia exponencial uniparamétrica.

2. La $\{B(k_0, p), p \in (0,1)\}$ es una familia exponencial uniparamétrica.

3.3.2. Familia exponencial k-paramétrica

Se dice que la familia de distribuciones es *exponencial k-paramétrica* si se cumple que:

1. El espacio paramétrico, Θ, es un intervalo de $\mathbb{R}^k$, ($\Theta \subseteq \mathbb{R}^k$).

2. El conjunto de valores de la variable no depende de θ:
$$\{x, f_\theta(x) > 0\} = \chi, \ \forall \theta \in \Theta.$$
3. Existen funciones real-valuadas $Q_1(\theta), \ldots, Q_k(\theta)$ y $D(\theta)$, definidas sobre Θ, y existen funciones medibles Borel $T_1, \ldots, T_k$ y S, también real valuadas, tales que
$$\forall \theta \in \Theta, \ f_\theta(x) = \exp\left[\sum_{h=1}^{k} Q_h(\theta) T_h(x) + D(\theta) + S(x)\right], \ x \in \chi.$$

Ejemplo: La $\{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ es una familia exponencial uniparamétrica en cada parámetro y bi-paramétrica en los dos parámetros.

Teorema: Si se tiene una v.a. $X \rightsquigarrow \{F_\theta, \theta \in \Theta\}$ con familia de funciones asociadas, $\{f_\theta(x), \theta \in \Theta\}$, siendo f.d.d. o f.m.p. según el caso, donde la familia de distribuciones es exponencial k-paramétrica, entonces la familia de distribuciones asociadas a $(X_1, \ldots, X_n)$ (que es una m.a.s. de X) es también exponencial k-paramétrica:
$$f_\theta^n(x_1, \ldots, x_n) = \exp\left\{\sum_{h=1}^{k} Q_h(\theta) \left(\sum_{i=1}^{n} T_h(x_i)\right) + \sum_{i=1}^{n} S(x_i) + nD(\theta)\right\}, \ (x_1, \ldots, x_n) \in \chi^n,$$

y se tiene:

- El estadístico $(\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_k(X_i))$ es suficiente para θ.
- Si $k \leq n$ y el conjunto imagen de la función $Q(\theta) = (Q_1(\theta), \ldots, Q_k(\theta))$ contiene a un abierto de $\mathbb{R}^k$, el estadístico $(\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_k(X_i))$ es también completo.

Ejemplo: Sea X una v.a. con distribución $\mathcal{N}(\mu, \sigma^2)$ con $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Buscar un estadístico suficiente y completo basado en una muestra de tamaño n.