Teoria-tema-1.pdf

Apuntes completos tema 1

KEEP
CALM
AND
ESTUDIA
UN POQUITO

Tema 1: Introducción a la Inferencia Estadística. Estadísticos muestrales. [1]

Índice

Introducimos en este tema algunos conceptos básicos de Inferencia Estadística. En particular, estudiaremos los elementos principales de un problema de Inferencia Paramétrica tales como muestra aleatoria simple y estadístico muestral.

[1] Si encuentras algún error en el documento, escribe un correo electrónico a *melferizq@gmail.com*.

1. Introducción.

El Cálculo de Probabilidades proporciona una teoría matemática que permite analizar propiedades de fenómenos (o experimentos) en los que interviene el azar. En general, se trabajará en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, construido a partir del conjunto de los posibles resultados del fenómeno (espacio muestral), sobre el que se define una σ-álgebra de sucesos, y una función de probabilidad adecuada. Así, una vez el espacio probabilístico se encuentra totalmente especificado, se podrá estudiar la ocurrencia de los sucesos incluidos en Ω.

Los espacios muestrales asociados a fenómenos aleatorios podrán ser cualitativos o cuantitativos, siendo evidente que el tratamiento matemático de una variable cualitativa puede ser complejo. Por ello, con frecuencia en Probabilidad se trabajará con variables aleatorias puesto que facilitan el tratamiento numérico de los datos obtenidos en el estudio de cualquier experimento aleatorio.

Sin embargo, para poder construir una variable aleatoria que modele una situación real será necesario tener construido previamente su espacio de probabilidad asociado, y se podrá dar el caso en el que, por cualquier razón, se disponga de un conjunto de observaciones acerca del fenómeno considerado en lugar de un espacio probabilístico bien definido. Al tener algún desconocimiento sobre las variables de estudio o algunas de sus características, no podremos construir completamente la variable aleatoria necesaria para el cálculo de probabilidades, y por tanto no podremos obtener conclusiones sobre los datos a analizar.

Con el fin de obtener conclusiones a partir de unos pocos datos obtenidos en el estudio de un fenómeno aleatorio cuya variable aleatoria modelo es, de algún modo, desconocida, se desarrollan métodos que se engloban dentro de lo que se llama *Inferencia Estadística* o *Estadística Matemática*.

Se plantea el problema de cómo obtener información acerca de la ley de probabilidad de un fenómeno a partir de una observación no exhaustiva del mismo. La Inferencia Estadística trata de analizar e interpretar las observaciones de las que se dispone como método para obtener conclusiones sobre la ley de probabilidad del fenómeno de estudio.

Toda variable aleatoria X sigue una *distribución*, que es una función que indica la probabilidad de que X tome o bien un valor concreto, o bien un valor dentro de un intervalo (o un conjunto de intervalos). La falta del conocimiento total de X se presenta cuando no se conoce su distribución, bien porque no se sabe nada acerca de su expresión matemática, bien porque aunque se sepa algo, esta depende de un parámetro desconocido.

Ejemplos de la primera situación ocurren cuando, a pesar de no saber nada sobre la expresión matemática de la distribución, se saben algunas de sus propiedades, como por ejemplo que es simétrica, o continua, o absolutamente continua, o que existen sus momentos centrados hasta orden 3. En este caso, conocer la distribución de una variable a partir de algunas de sus propiedades será un problema de *Inferencia no paramétrica* (tema 9).

Para la segunda situación, lo que se hace es obtener una aproximación del parámetro que se desconoce. De este tipo de problemas se encarga la *Inferencia Paramétrica* (temas 1-8). Dependiendo de la forma en que queremos obtener el valor de tal parámetro, podemos dividir la Inferencia Paramétrica en:

- *Estimación puntual.* Se busca obtener un pronóstico numérico único acerca del parámetro.
- *Intervalos de confianza.* Se da un margen de variación sobre la acotación del parámetro; en concreto, se precisa un intervalo numérico en el que pueda afirmarse, razonablemente, que este varía.

- *Contrastes de hipótesis.* Se busca corroborar o invalidar una determinada afirmación acerca del parámetro desconocido.

1.1. Introducción a la Inferencia Paramétrica.

De forma genérica, la distribución desconocida F de la variable aleatoria X involucrada en un problema de Inferencia Estadística recibe el nombre de *distribución teórica.* El mayor o menor grado de desconocimiento acerca de la distribución teórica F se refleja mediante la familia $\mathcal{F}$ de distribuciones, candidatas a ser realmente la distribución de la población.

La situación que estudiamos en primer lugar es aquella en que la familia $\mathcal{F}$ está compuesta por distribuciones que se expresan a partir de una función fija y conocida, dependientes de un parámetro θ, de una o más dimensiones, que varía dentro de un subconjunto Θ de $\mathbb{R}^n$, denominado espacio paramétrico:

$$\mathcal{F} = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^n\}$$

2. Muestra aleatoria simple.

Nos planteamos de qué forma debemos obtener observaciones a partir de las cuales se ha de intentar disminuir el desconocimiento de la distribución teórica F de la variable aleatoria X en estudio. En principio, tales observaciones pueden obtenerse llevando a cabo repeticiones del experimento aleatorio que da lugar a X y anotando los valores de la variable en cada una de ellas. Se tendrá así un conjunto de valores numéricos $(x_1, \ldots, x_n)$ que se denomina una *muestra aleatoria* de X. El número n de repeticiones efectuadas, y de observaciones obtenidas, se denomina *tamaño de la muestra.*

Observación 1. *Los procedimientos posibles para obtener una muestra aleatoria pueden ser diversos, dependiendo de las condiciones en que se efectúen las repeticiones del experimento aleatorio. Todo problema de Inferencia Estadística, para estar bien formulado, tiene que incluir la definición precisa del procedimiento de muestreo con el que se obtienen las observaciones.*

Cuando ya se ha obtenido una muestra aleatoria de tamaño n, $(x_1, \ldots, x_n)$, se dispone simplemente de un conjunto de n variables numéricas con las cuales tratar de cumplir el objetivo de precisar la distribución teórica. Sin embargo, cuando se está planificando la obtención de la muestra aleatoria y diseñando el procedimiento que se empleará para intentar conocer tal distribución, todavía no se sabe qué valores numéricos $(x_1, \ldots, x_n)$ resultarán, de manera que deben considerarse como n variables aleatorias, $(X_1, \ldots, X_n)$. Así, la muestra aleatoria, como conjunto de n variables aleatorias, tiene una distribución n-dimensional (supuesto que las observaciones individuales son unidimensionales), que determina las probabilidades de aparición de cada muestra efectiva.

Observación 2.

- *Al conjunto de valores $(x_1, \ldots, x_n)$ obtenidos en la selección de una muestra particular se le llama* realización muestral.

  A efectos prácticos, en un problema real se estudiará la muestra aleatoria $(X_1, \ldots, X_n)$ a partir de una realización muestral suya, es decir, a partir de valores concretos $(x_1, \ldots, x_n)$, donde cada valor x_i se obtiene al observar la variable aleatoria X_i correspondiente.

- *Al conjunto de todas las realizaciones muestrales, es decir, el conjunto de los posibles valores del vector $(x_1, \ldots, x_n)$, se le denomina* espacio muestral *o conjunto de posibles valores de la muestra, y se le denota por χ^n.*

En el caso en que se efectúen repeticiones independientes del experimento aleatorio (es decir, con reemplazamiento; una vez observado un individuo lo devolvemos a la población), cabe destacar que lo que se está haciendo es observar de forma independiente, para la obtención de cada valor x_i, la variable X en estudio, cuya distribución teórica es F, luego cada uno de los valores que se obtengan, al ser extraídos con repetición, tendrán la misma distribución.

Muestra aleatoria simple.

Definición 1. *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 variables aleatorias $X_1, \dots, X_n$ 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. Es decir, la función de distribución conjunta de una muestra aleatoria simple $(X_1, \dots, X_n)$ correspondiente a una distribución de la población F es:

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

- Para el caso de variables aleatorias discretas, se cumple que la función masa de probabilidad conjunta es el producto de las funciones masa de probabilidad marginales:

$$P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] = P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \dots \cdot P[X_n = x_n]$$

- Para el caso de variables aleatorias continuas, se cumple que la función de densidad de probabilidad conjunta es el producto de las funciones de densidad marginales:

$$f_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

Ejemplo 1.

a) Calcular la función masa de probabilidad conjunta de una muestra aleatoria simple de tamaño arbitrario de una variable $X \rightsquigarrow B(k_0, p)$.

b) Calcular la función de densidad de probabilidad conjunta de una muestra aleatoria simple de tamaño arbitrario de una variable $X \rightsquigarrow U(a, b)$.

Solución.

1. Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow B(k_0, p)$.

$$P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] \underbrace{=}_{\text{independientes}} P[X_1 = x_1]P[X_2 = x_2]\cdots P[X_n = x_n]$$

$$\underbrace{=}_{\text{identicamente distribuidas a } X} P[X = x_1]P[X = x_2]\cdots P[X = x_n]$$

$$= \binom{k_0}{x_1} p^{x_1}(1-p)^{k_0-x_1} \binom{k_0}{x_2} p^{x_2}(1-p)^{k_0-x_2} \cdots \binom{k_0}{x_n} p^{x_n}(1-p)^{k_0-x_n}$$

$$= \prod_{i=1}^{n} \binom{k_0}{x_i} p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n (k_0-x_i)} = \prod_{i=1}^{n} \binom{k_0}{x_i} p^{\sum_{i=1}^n x_i}(1-p)^{nk_0-\sum_{i=1}^n x_i}$$

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U(a,b)$. Por el mismo razonamiento que antes, pero ahora con la función de densidad, se tiene:

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = f_X(x_1)\cdots f_X(x_n) = \frac{1}{(b-a)^n}, a < x_i < b, \forall i$$

3. Función de distribución muestral.

En los problemas de Inferencia Estadística, en general se usará la función de densidad muestral para variables aleatorias continuas y la función masa de probabilidad muestral para variables aleatorias discretas. Sin embargo, asociada a la muestra también existirá una función de distribución, que tendrá propiedades de convergencia que aseguran que su uso es conveniente para aproximar la función de distribución de la variable aleatoria de partida. Así, vamos a asociar a cada muestra aleatoria simple $(X_1, \ldots, X_n)$ de una variable aleatoria 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.

Función de distribución muestral.

Definición 2. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X con función de distribución F. Se define la función de distribución muestral de X asociada a la muestra $(X_1, \ldots, X_n)$ como*

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

Observación 3. *La función de distribución muestral se puede escribir como*

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

donde el sumando $I_{(-\infty,x]}(X_i)$ es la variable aleatoria definida como

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

Esta puede verse como una variable donde con probabilidad 1 se tiene el suceso éxito *(donde el éxito se resume en $X_i \leq x$) y con probabilidad 0 se tiene el suceso* fracaso *(es decir, $X_i > x$). Por tanto, es una Bernoulli, y sabemos que la suma de n Bernoullis es una binomial. Además, se verifica:*

$$P[I_{(-\infty,x]}(X_i) = 1] = P[X_i \leq x] = F(x)$$

$$P[I_{(-\infty,x]}(X_i) = 0] = P[X_i > x] = 1 - F(x)$$

Propiedades de la función de distribución muestral.

Proposición 1.

i) Para cada realización muestral, $(x_1, \ldots, x_n) \in \chi^n$, $F^*_{X_1,\ldots,X_n}(x)$ 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.*

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

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

iii) Para valores grandes de n*, en virtud del Teorema Central del Límite:*

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

Demostración.

i) Asumiendo $x_1 < x_2 < \cdots < x_n$ (en otro caso, basta con que la función que se construye tenga saltos múltiples), se tiene:

$$F^*_{X_1,\ldots,X_n}(x) = \begin{cases} 0/n & x < x_1 \\ 1/n & x_1 \leq x < x_2 \\ \ldots & \ldots \\ 1 & x \geq x_n \end{cases}$$

Así, F^* es una función de distribución en $\mathbb{R}$ pues tiene límite 0 a la izquierda, límite 1 a la derecha, es monótona no decreciente y continua a la derecha.

ii) Por la Observación 3, se tiene:

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

Teniendo en cuenta la linealidad de la esperanza:

$$\mathrm{E}[nF^*_{X_1,\ldots,X_n}(x)] = nF(x) \Rightarrow n\mathrm{E}[F^*_{X_1,\ldots,X_n}(x)] = nF(x) \Rightarrow \mathrm{E}[F^*_{X_1,\ldots,X_n}(x)] = F(x)$$

Por otro lado, puesto que $\mathrm{Var}[nX] = n^2\mathrm{Var}[X]$:

$$\mathrm{Var}[nF^*_{X_1,\ldots,X_n}(x)] = n^2\mathrm{Var}[F^*_{X_1,\ldots,X_n}(x)] = nF(x)(1-F(x))$$
$$\Rightarrow \mathrm{Var}[F^*_{X_1,\ldots,X_n}(x)] = \frac{F(x)(1-F(x))}{n}$$

iii) Como $nF^*_{X_1,\ldots,X_n}(x)$ es suma de variables aleatorias independientes e idénticamente distribuidas (ind. e i.d), el Teorema Central del Límite permite afirmar que, cuando $n \to +\infty$,

$$\frac{nF^*_{X_1,\ldots,X_n}(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} = \sqrt{n}\frac{F^*_{X_1,\ldots,X_n}(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \rightsquigarrow \mathcal{N}(0,1)$$

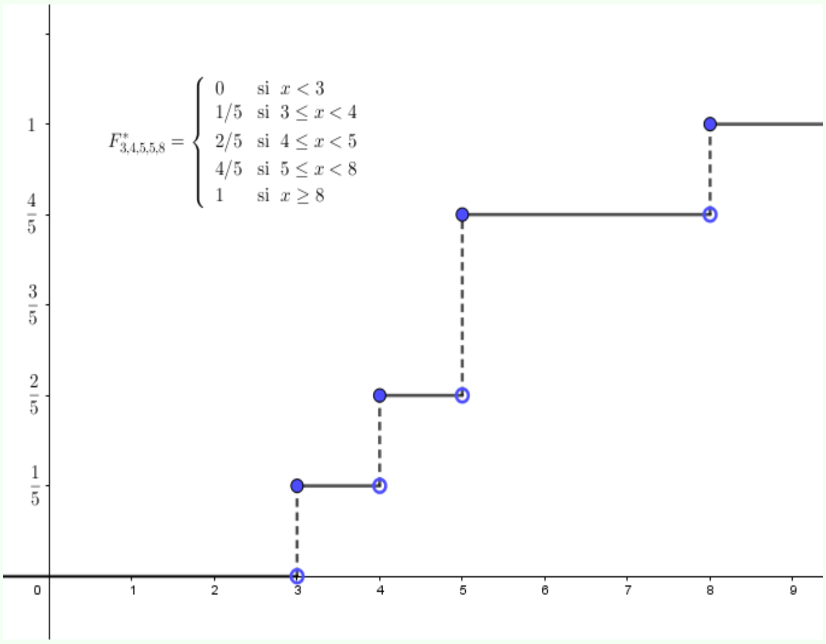
de manera que, cuando el tamaño muestral es grande, se tiene

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

■

Ejemplo 2. *Dada una muestra aleatoria simple formada por las observaciones* $(3, 8, 5, 4, 5)$, *obtener su función de distribución muestral y realizar la representación gráfica.*

Solución. Ordenamos los valores muestrales: $3 < 4 < 5 \leq 5 < 8$. Por tanto, en $x = 5$, la función de distribución muestral tendrá un salto de amplitud $2/5$, puesto que al haber cinco observaciones muestrales, la amplitud de cada salto en caso de valores no repetidos es $1/5$.



En el punto iii) de la Proposición 1 se ha visto que, a medida que n crece (aumenta el tamaño muestral), la distribución muestral $F^*_{X_1,\ldots,X_n}(x)$ se encuentra cada vez más concentrada alrededor del valor de $F(x)$, es decir, con un tamaño muestral grande, es probable que la muestra seleccionada proporcione un valor de $F^*_{X_1,\ldots,X_n}(x)$ muy próximo al valor desconocido de $F(x)$.

Nos interesa por tanto poder asegurar que dicha aproximación se produce simultáneamente para todos los valores de x, lo cual no es inmediato, pues aunque el suceso relativo a x se produzca con probabilidad 1, el número de valores de x no es numerable. En ese sentido, el siguiente teorema prueba que $F^*_{X_1,\ldots,X_n}(x)$ converge a $F(x)$, uniformemente en x, con probabilidad 1.

Teorema de Glivenko-Cantelli.

Teorema 1. *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución común* F. *Si* $F^*_{X_1,\ldots,X_n}(x)$ *es la función de distribución muestral asociada a la muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de* $X \rightsquigarrow F$, *se verifica que* $F^*_{X_1,\ldots,X_n}(x)$ *converge casi seguramente y uniformemente a la función de distribución de* X, F.

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

Es decir, con probabilidad 1, al tomar sucesivas observaciones independientes de la variable y considerar las correspondientes funciones de distribución muestrales, se verifica:

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

4. Estadístico muestral.

Para obtener información acerca de la función de distribución de la variable X en estudio (que, a priori, depende de ciertos parámetros que no se conocen), se toma una muestra aleatoria simple de la misma, $(X_1, \ldots, X_n)$, y se observan los valores que esta toma. Puede resultar, en ocasiones, más cómodo analizar, en lugar de todos los valores del vector muestral, una función suya, como por ejemplo la suma de los mismos. A dicha función de la muestra aleatoria simple, siempre que sea independiente de los parámetros desconocidos, le llamaremos *estadístico*, y la ventaja de su uso reside en que, en lugar de trabajar con la muestra, que es un vector aleatorio de dimensión n, se trabaja con una variable aleatoria unidimensional o con un vector de menor dimensión (en general, los estadísticos buscan reducir la dimensión con la que estamos trabajando.)

Estadístico.

Definición 3. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Un estadístico muestral asociado a X es una función $T : (\mathbb{R}^n, \mathcal{B}^n) \longrightarrow (\mathbb{R}^k, \mathcal{B}^k)$ medible, es decir, cumple $T^{-1}(B) \in \mathcal{B}^n, \forall B \in \mathcal{B}^k$, e independiente de cualquier parámetro desconocido. Lo denotamos por $T(X_1, \ldots, X_n)$, o cuando se sobreentienda que es un estadístico, simplemente T o $T(X)$.*

Observación 4. *Como T es una función medible, el estadístico muestral $T(X_1, \ldots, X_n)$ es una variable aleatoria de dimensión k. Por tanto, la función T basta que esté definida en el espacio muestral χ^n y no en todo $\mathbb{R}^n$.*

Ejemplo 3. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Algunos de los estadísticos muestrales más frecuentes en su uso son los siguientes:*

- *Media muestral:* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$
- *Varianza muestral:* $Var[X] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$
- *Cuasivarianza muestral:* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$
- *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 4. *Sea X una variable aleatoria con distribución $B(1, p)$ con $p \in (0, 1)$. Se toma una muestra aleatoria simple de tamaño 5, $(X_1, X_2, X_3, X_4, X_5)$, y se obtiene la siguiente observación: $(0, 1, 1, 0, 0)$. Determina el valor de los estadísticos estudiados en el Ejemplo 3.*

Solución. Basta con sustituir en las fórmulas del Ejemplo 3 para obtener los siguientes valores:

$$\bar{X} = \frac{2}{5} \qquad \text{Var}[X] = \frac{6}{25} \qquad S^2 = 0{,}3 \qquad X_{(1)} = 0 \qquad X_{(n)} = 1$$

4.1. Distribución en el muestreo y características de un estadístico.

Por ser $T(X_1, \dots, X_n)$ una varable aleatoria, es entonces evidente que tendrá una función de distribución y unas características numéricas asociadas tales como esperanza, varianza o existencia de momentos. Así, como en el caso de una variable aleatoria cualquiera X, tener completamente determinado al estadístico T implica conocer todas sus características.

Distribución en el muestro.

Definición 4. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. La distribución en el muestreo de un estadístico T definido en el espacio muestral $(\chi^n, \mathcal{B}^n)$ es la distribución de la variable aleatoria $T(X_1, \dots, X_n)$.*

Proposición 2. *La función generatriz de momentos del estadístico media muestral viene dada por:*

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

Demostración. Recordemos que la función generatriz de momentos de una variable aleatoria es aditiva-multiplicativa en el sentido de que la función generatriz de momentos de la suma de n variables aleatorias independientes coincide con el producto de las n funciones generatrices de momentos.

Sea X una variable aleatoria, y $(X_1, \dots, X_n)$ una muestra aleatoria simple de X de tamaño n.

$$M_{\bar{X}}(t) = \mathrm{E}\left[e^{t\bar{X}}\right] = \mathrm{E}\left[e^{t\frac{\sum X_i}{n}}\right] = \mathrm{E}\left[e^{\frac{\sum X_i t}{n}}\right] = \mathrm{E}\left[\prod e^{\frac{tX_i}{n}}\right]$$
$$\underbrace{=}_{\text{ind}} \prod_{i=1}^{n} \mathrm{E}\left[e^{\frac{t}{n}X_i}\right] = \prod_{i=1}^{n} M_{X_i}(t/n) \underbrace{=}_{\text{i.d}} (M_X(t/n))^n$$

■

Importante para un tipo test.

Ejemplo 5. *Obtener la distribución muestral de $\bar{X}$ para $(X_1, \dots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.*

Solución. Aplicando la Proposición 2, se tiene:

$$M_{\bar{X}}(t) = (M_X(t/n))^n = \left(\exp\left(\frac{t}{n}\mu + \left(\frac{t}{n}\right)^2 \frac{\sigma^2}{2}\right)\right)^n = \exp\left(\cancel{n}\left(\frac{t}{\cancel{n}}\mu + \frac{t^2}{n^{\cancel{2}}}\frac{\sigma^2}{2}\right)\right) \Rightarrow \bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Consideremos los estadísticos ordenados, definidos como:

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

Proposición 3. *La distribución muestral general de los estadísticos de orden, según la variable aleatoria sea discreta o continua, es:*

- *X discreta:*

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

- *X continua:*

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

Observación 5. *Las fórmulas obtenidas en la Proposición 3 cuentan de cuántas formas se pueden seleccionar permutaciones de r elementos menores o iguales que un cierto x.*

Importante para un tipo test.

Ejemplo 6.

a) Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para X una variable aleatoria cualquiera y $(X_1, \ldots, X_n)$ una muestra aleatoria simple suya.

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

Solución.

a) Hallamos las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para una muestra aleatoria simple de tamaño n. de una variable aleatoria X con función de distribución F_X.

- Para la distribución del mínimo $X_{(1)}$, tengamos en cuenta que si el mínimo de los valores es mayor que un cierto x, entonces el resto de valores también serán mayores que x. Así:

$$\begin{aligned} &F_{X_{(1)}}(x) = P[X_{(1)} \leq x] = 1 - P[X_{(1)} > x] = 1 - P[X_1 > x, \ldots, X_n > x] \\ &\underbrace{=}_{\text{ind}} 1 - P[X_1 > X] \cdots P[X_n > x] \underbrace{=}_{\text{i.d}} 1 - (P[X > x])^n = 1 - (1 - P[X \leq x])^n \end{aligned}$$

$$\boxed{F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n} \tag{1}$$

- Para la distribución del máximo $X_{(n)}$, se tiene que si el máximo de los valores es menor que un cierto x, entonces el resto de valores son también menores que x. Por tanto:

$$F_{X_{(n)}}(x) = P[X_{(n)} \leq x] = P[X_1 \leq x, \ldots, X_n \leq x] \underbrace{=}_{\text{ind}} P[X_1 \leq x] \cdots P[X_n \leq x]$$

$$\underbrace{=}_{\text{i.d}} (P[X \leq x])^n = (F_X(x))^n$$

$$\boxed{F_{X_{(n)}}(x) = (F_X(x))^n} \qquad (2)$$

b) Para obtener las funciones de distribución muestrales de $X_{(1)}$ y $X_{(n)}$ para el caso de una variable continua uniforme, basta con sustituir su función de distribución en las fórmulas (1) y (2) y su función de densidad en ambas derivadas. En este caso, nos centraremos en calcular las funciones de densidad muestrales, que se obtienen derivando las funciones de distribución muestrales:

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \Rightarrow f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x)$$

$$F_{X_{(n)}}(x) = (F_X(x))^n \Rightarrow f_{X_{(n)}}(x) = n(F_x(x))^{n-1} f_X(x)$$

Así, teniendo en cuenta que las funciones de distribución y de densidad de una variable $U(a,b)$ vienen dadas, respectivamente, por

$$F_X(x) = \frac{x-a}{b-a}; \quad f_X(x) = \frac{1}{b-a}, \ a < x < b$$

se tiene entonces:

$$f_{X_{(1)}}(x) = n\left(1 - \frac{x-a}{b-a}\right)^{n-1} \frac{1}{b-a}; \quad f_{X_{(n)}}(x) = n\left(\frac{x-a}{b-a}\right)^{n-1} \frac{1}{b-a}, \quad a < x < b$$

Otros ejemplos de estadísticos menos frecuentes (de los cuales ya hemos mencionado algunos casos particulares) son los siguientes.

Momentos muestrales centrados y no centrados.

Definición 5. *Se definen los momentos centrados y no centrados de una muestra aleatoria simple de tamaño n como:*

- *Momento no centrado de orden $k \in \mathbb{N}$:* $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.

  En particular, se tiene $A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} =$ *media muestral.*

- *Momento centrado de orden $k \in \mathbb{N}$:* $B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k$.

  En particular, se tiene $B_1 = 0$ (por las propiedades de la media de una distribución), y definimos $B_2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 = Var[X] =$ *Varianza muestral.*

Proposición 4. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X.*

i) Para los momentos no centrados, se tiene:

$$E[A_k] = E[X^k] \qquad Var[A_k] = \frac{1}{n}(E[X^{2k}] - E[X^k]^2)$$

En particular, $E[A_1] = \mu$, $Var[A_1] = \dfrac{\sigma^2}{n}$, donde μ y σ^2 son, respectivamente, la media y la varianza poblacional.

ii) Para el momento centrado de orden 2, se tiene

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

En particular, se verifica $E[S^2] = \sigma^2$, donde $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$.

A S^2 así definida le llamamos cuasivarianza muestral.

Demostración.

i)

$$\mathrm{E}[A_k] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X^k\right] = \frac{1}{n}\mathrm{E}[nX^k] = \mathrm{E}[X^k]$$

En particular, $\mathrm{E}[A_1] = \mathrm{E}[X] = \mu$.

$$\mathrm{Var}[A_k] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i^k\right] \underbrace{=}_{\text{ind}} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i^k]$$
$$= \frac{1}{n}\mathrm{Var}[X^k] = \frac{1}{n}(\mathrm{E}[X^{2k}] - \mathrm{E}[X^k]^2)$$

En particular, $\mathrm{Var}[A_1] = \dfrac{1}{n}(\mathrm{E}[X^2] - (\mathrm{E}[X])^2) = \dfrac{\sigma^2}{n}$.

ii) El teorema de König tiene su versión para la varianza poblacional, cumpliéndose que:

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

Por tanto:

$$\mathrm{E}[B_2] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i^2] - \mathrm{E}[\bar{X}^2] \underbrace{=}_{i.d} \frac{1}{\cancel{n}}\cancel{n}\mathrm{E}[X^2] - \mathrm{E}[\bar{X}^2] = \mathrm{E}[X^2] - (\mathrm{E}[\bar{X}])^2 = (1)$$

Utilizando que

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \Rightarrow \mathrm{E}[X^2] = \mathrm{Var}[X] + (\mathrm{E}[X])^2$$
$$\mathrm{Var}[\bar{X}] = \mathrm{E}[\bar{X}^2] - (\mathrm{E}[\bar{X}])^2 \Rightarrow \mathrm{E}[\bar{X}^2] = \mathrm{Var}[\bar{X}] + (\mathrm{E}[\bar{X}])^2$$

y que $\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i\right] = \dfrac{1}{n^2}\mathrm{Var}\left[\displaystyle\sum_{i=1}^{n} X_i\right] \underbrace{=}_{\text{i.i.d}} \dfrac{1}{n^{\cancel{2}}}\cancel{n}\mathrm{Var}[X] = \dfrac{1}{n}\mathrm{Var}[X]$, se llega a:

$$(1) = \mathrm{Var}[X] + (\mathrm{E}[X])^2 - \frac{\mathrm{Var}[X]}{n} - \underbrace{(\mathrm{E}[\bar{X}])^2}_{(\mathrm{E}[X])^2} = \frac{(n-1)\mathrm{Var}[X]}{n} = \frac{(n-1)\sigma^2}{n} = \mathrm{E}[B_2]$$

En particular, podemos reescribir S^2 como

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

Por tanto:

$$\mathrm{E}[S^2] = \frac{\cancel{n}}{\cancel{n-1}}\cdot\frac{\cancel{(n-1)}\sigma^2}{\cancel{n}} = \sigma^2$$

■

Cuantil muestral.

Definición 6. *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 puede expresar de la siguiente forma en función de los elementos de la muestra ordenada:

- *Si $np \in \mathbb{N}$, $c_p = \dfrac{X_{(np)} + X_{(np+1)}}{2}$.*
- *En otro caso, sea $[np]$ la parte entera de np, entonces $c_p = X_{([np]+1)}$.*

Función generatriz de momentos muestral.

Definición 7. *Se define la función generatriz de momentos muestral como*

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

Proposición 5. *La función generatriz de momentos muestral se utiliza para obtener los momentos no centrados, pues se verifica*

$$\left.\frac{\partial^k M^*(t)}{\partial t^k}\right|_{t=0} = A_k$$