Tema 10. Estimación Puntual.

Presentación y Objetivos.

1. Comprender el concepto de estimador y su distribución.
2. Conocer y saber aplicar el método de los momentos y el de máxima verosimilitud para obtener estimadores.

Esquema Inicial.

1. Introducción.
2. Estadísticos y estimadores.
3. Métodos de obtención de estimadores.
4. Obtención de estimadores en la distribución normal. Teorema de Fisher.

Desarrollo del Tema.

1. Introducción.

Supóngase que se observa una muestra aleatoria simple de una variable aleatoria X que, se sabe, sigue una distribución conocida, por ejemplo, Poisson, Normal o cualquiera de las ya estudiadas en los temas 6 y 7. Se sabe la forma de esa distribución pero se ignora el valor de alguno(s) o todos sus parámetros. La pregunta fundamental que se plantea es: ¿cómo se puede utilizar la información muestral para aproximar o estimar esos parámetros desconocidos de una distribución? La respuesta inmediata es mediante la **estimación**, que consiste en asignar valores concretos a los parámetros desconocidos. Existen dos tipos principales de estimación:

- Estimación puntual: se da un único valor aproximado para el parámetro desconocido, un solo punto o número real.
- Estimación por intervalos: se da un intervalo de valores posibles para el parámetro desconocido. En ese intervalo se cree que se encuentra el verdadero valor del parámetro con cierta seguridad que habrá que especificar. Este tipo de estimación se dará en el tema 11.

A continuación se recuerda lo que se conoce como parámetro de una distribución. Un **parámetro** es un valor utilizado para representar una característica concreta de la distribución de una variable aleatoria X. En inferencia estadística esta variable aleatoria representará a la población bajo estudio. Así, si $X \sim \mathcal{N}(\mu, \sigma)$, μ será el parámetro que representa la media poblacional y σ será el parámetro que representa la desviación típica, si $X \sim \mathcal{P}(\lambda)$, λ representa la media poblacional y si $X \sim \mathcal{E}xp(\lambda)$, el parámetro representa la tasa de la distribución.

2. Estadísticos y estimadores.

Dada $X_1, X_2, \dots, X_n$ muestra aleatoria simple de una variable aleatoria X, se llama **estadístico** a toda función (medible) de los elementos de la muestra, es decir,

$$Estadístico = T(X_1, \dots, X_n)$$

Un **estimador** de un parámetro será un estadístico cuya finalidad es aproximar el verdadero valor del parámetro desconocido. Si queremos estimar un parámetro genérico θ se utilizará la notación $\hat{\theta}$ para designar su estimador.

Los estadísticos son variables aleatorias y son funciones que no dependen del parámetro desconocido. Se observa la muestra $X_1 = x_1, \dots, X_n = x_n$ y el estadístico resume la información contenida en la muestra para dar información sobre θ.

Ejemplo 1: Dada $X_1, \dots, X_n$ muestra aleatoria simple de una variable aleatoria $X \sim \mathcal{N}(\mu, \sigma)$ un estadístico sería:

$$T(X_1, \dots, X_n) = X_1^2 + \dots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

No sería un estadístico la función:

$$T(X_1, \dots, X_n) = \mu X_1^2 + \dots + \mu X_n^2 = \mu \sum_{i=1}^{n} X_i^2$$

ya que depende del parámetro desconocido μ.

Ejemplo 2: Se estudia la variable aleatoria X = *número de trabajos enviados a una impresora en un día* suponiendo que $X \sim \mathcal{P}(\lambda)$. Si se observa el número de trabajos enviados en n días, se pueden tomar como estimador de λ:

$$\hat{\lambda} = \bar{X} \quad o \quad \hat{\lambda} = mín\,(X_i) \quad o \quad \hat{\lambda} = máx\,(X_i)$$

Ejemplo 3: Si $X \sim \mathcal{N}(\mu, 1)$ y se toma una muestra aleatoria simple $X_1, \dots, X_n$, se pueden tomar como estimador del parámetro μ los estadísticos:

$$\hat{\mu} = \bar{X} \quad o \quad \hat{\mu} = Med\,(X_i)$$

ya que sabemos que, al ser la distribución normal simétrica, la media y la mediana coinciden.

En los últimos ejemplos hemos visto que para un mismo parámetro se pueden proponer diferentes estimadores. A continuación se presentan los métodos para obtener buenos estimadores.

3. Métodos de obtención de estimadores.

Se estudiarán dos métodos, el método de los momentos y el método de máxima verosimilitud.

3.1. Método de los momentos.

Su principio es sencillo, se estiman los momentos poblacionales por los correspondientes momentos muestrales. Así, la media poblacional se estimará por medio de la media muestral y la varianza poblacional por medio de la varianza muestral. Se ilustra el método con un ejemplo para luego desarrollarlo de forma teórica:

Ejemplo 4: Sea X variable aleatoria con distribución $\gamma(\lambda, p)$. Obtener estimadores de los parámetros λ y p por el método de los momentos. Primero se buscan relaciones funcionales entre los momentos respecto del origen de X y los parámetros a estimar:

$$E(X) = \alpha_1 = \frac{p}{\lambda}$$

$$E(X^2) = \alpha_2 = V(X) + E(X^2) = \frac{p}{\lambda^2} + \left(\frac{p}{\lambda}\right)^2 = \frac{p(p+1)}{\lambda^2}$$

Se despejan los parámetros desconocidos, con lo que:

$$\lambda = \frac{\alpha_1}{\alpha_2 - \alpha_1^2} \qquad y \qquad p = \frac{\alpha_1^2}{\alpha_2 - \alpha_1^2}$$

A continuación se estiman los momentos poblacionales α (ver definición en el tema 5) por medio de sus correspondientes momentos muestrales a (ver definición en el tema 2) y los estimadores serían:

$$\hat{\lambda} = \frac{a_1}{a_2 - a_1^2} \qquad y \qquad \hat{p} = \frac{a_1^2}{a_2 - a_1^2}$$

siendo $a_1 = \bar{x}$ y $a_2 = \frac{\sum_{i=1}^n x_i^2}{n}$.

Por tanto, si se quiere estimar el parámetro k-dimensional $(\theta_1, \ldots, \theta_k)$, el método de los momentos consiste en:

1. Encontrar relaciones funcionales, tantas como parámetros se quieren estimar, entre los momentos con respecto del origen de la variable aleatoria de la que proviene la muestra y los parámetros.

$$\begin{cases} \theta_1 = g_1(\alpha_1, \dots, \alpha_k) \\ \vdots \qquad\quad \vdots \\ \theta_k = g_2(\alpha_1, \dots, \alpha_k) \end{cases}$$

2. Estimar dichos momentos, $\alpha_1, \alpha_2, \dots, \alpha_k$, por medio de los momentos muestrales $a_1, a_2, \dots, a_k$, obteniendo los estimadores.

$$\begin{cases} \hat{\theta}_1 = g_1(a_1, \dots, a_k) \\ \vdots \qquad\quad \vdots \\ \hat{\theta}_k = g_2(a_1, \dots, a_k) \end{cases}$$

Ejemplo 5: Dada $X \sim \mathcal{N}(\mu, \sigma)$, encontrar los estimadores para los parámetros μ y σ por el método de los momentos.

Se quiere estimar el parámetro bidimensional $\theta = (\mu, \sigma)$. El sistema a resolver es:

$$\begin{cases} E(X) = \alpha_1 = \mu \\ E(X^2) = \alpha_2 = \sigma^2 + \mu^2 \end{cases}$$

Despejando μ y σ^2 se tiene:

$$\begin{cases} \hat{\mu} = a_1 = \bar{x} \\ \widehat{\sigma^2} = a_2 - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \end{cases}$$

3.2. Método de máxima verosimilitud.

3.2.1. Función de verosimilitud.

Dada X variable aleatoria discreta con función de probabilidad $P_\theta(X)$ y una muestra aleatoria simple de X, $(X_1, \dots, X_n)$, se definía (tema 9) la función de probabilidad conjunta de la muestra como:

$$P_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} p_\theta(x_i).$$

Ésta puede verse como función de la muestra aleatoria simple y como función del parámetro.

- Si θ fuera conocido, esta función determina la probabilidad de obtención de cada muestra.
- Si θ es desconocido y se conoce un valor concreto y fijo de la muestra $(x_1, \dots, x_n)$, esta función daría, para cada valor de θ, la probabilidad de obtener esa muestra observada.

La función de probabilidad o de densidad conjunta de la muestra aleatoria simple $(X_1, \dots, X_n)$, vista como función de θ, se **denomina función de verosimilitud $\boldsymbol{L(\theta)}$.** Es decir, para $(x_1, \dots, x_n)$ fijo, se tiene:

- $L(\theta) = \prod_{i=1}^{n} P_\theta(X_i = x_i) = \prod_{i=1}^{n} p_\theta(x_i)$ si X es discreta.
- $L(\theta) = \prod_{i=1}^{n} f_\theta(x_i)$ si X es continua.

3.2.2. Metodología.

El método de máxima verosimiltud selecciona como estimador del parámetro desconocido θ, aquel valor que maximiza la probabilidad de la muestra aleatoria observada, es decir:

$$\hat{\theta} = \max_{\theta} L(\theta)$$

Se maximiza la función $L(\theta)$ manteniendo $(x_1, \dots, x_n)$ fijo.

Sea L función de verosimilitud diferenciable, cuyo máximo no se alcanza en un extremo de su recorrido. Si el parámetro θ se supone k-dimensional, $\theta = (\theta_1, \dots, \theta_k)$, para encontrar el estimador de máxima verosimilitud $\hat{\theta}$ debe resolverse el sistema:

$$\begin{cases} \dfrac{\partial L(\theta)}{\partial \theta_1} = 0 \\ \quad \vdots \\ \dfrac{\partial L(\theta)}{\partial \theta_k} = 0 \end{cases}$$

Habrá que comprobar que, efectivamente, el $\hat{\theta}$ así obtenido es un máximo de la función. Para ello habrá que comprobar que la matriz de segundas derivadas parciales o matriz Hessiana es definida negativa.

Observación: en la práctica, los estimadores de máxima verosimilitud (E.M.V.), se obtienen derivando la función $ln\, L(\theta)$, también llamada **función soporte**. Al ser el logaritmo una función monótona, las funciones $L(\theta)$ y $ln\, L(\theta)$ alcanzan el máximo en el mismo punto. La ventaja es que, al tomar logaritmos, las constantes multiplicativas se transforman en aditivas y desaparecen al derivar, facilitando los cálculos.

Ejemplo 6: Supóngase que $X \sim \mathcal{P}(\lambda)$. Se toma una muestra aleatoria simple $(X_1, \ldots, X_n)$. Encontrar el E.M.V. del parámetro λ.

Lo primero que se hace es construir la función de verosimilitud. Dada $(X_1, \ldots, X_n)$, se tiene:

$$L(\lambda) = \prod_{i=1}^{n} P_\lambda(x_i) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i!}$$

A continuación se calcula $ln\, L(\lambda)$:

$$ln\, L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) ln\, \lambda - ln\left(\prod_{i=1}^{n} x_i\right).$$

Se deriva e, igualando a cero se obtiene $\hat{\lambda}$:

$$\frac{\partial lnL(\lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

$$\frac{\partial lnL(\lambda)}{\partial \lambda} = 0 \Longrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}$$

Se comprueba que, efectivamente, es un máximo:

$$\frac{\partial^2 lnL(\lambda)}{\partial \lambda^2} = -\frac{\sum x_i}{\lambda^2}\bigg]_{\lambda=\hat{\lambda}} = -\frac{1}{\bar{X}} < 0$$

Esta última expresión es menor que cero ya que los elementos de una muestra aleatoria simple que proviene de una distribución de Poisson son todos positivos y, por tanto, su media es también positiva.

4. Teorema de Fisher.

El Teorema de Fisher nos dice que:

Si $(X_1, \ldots, X_n)$ es una muestra aleatoria simple de $X \sim \mathcal{N}(\mu, \sigma)$, se verifica que:

1. $\bar{X}\ y\ S^2$ son independientes.
2. $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$