Auxiliar N°11. Estimación e Intervalos de Confianza

Probabilidades y Estadística - MA3403 - Primavera 2009
Profesor: Fernando Lema
Auxiliares: Abelino Jiménez - Benjamín Palacios

RESUMEN.
ESTIMACIÓN.

Definición. Muestra Aleatoria.

Sea X v.a. con cierta distribución de probabilidad. Sean X_1, $X_2, ..., X_n$ n variables aleatorias independientes, cada una con la misma distribución de X. Se dice que $(X_1, X_2, ..., X_n)$ es una muestra aleatoria de X.

Definición. Estadístico.

Sea X_1, $X_2, ..., X_n$ una muestra aleatoria de una v.a. X y sean $x_1, x_2, ..., x_n$ los valores tomados por la muestra. Sea H una función definida para $(x_1, ..., x_n)$. Se dice que $Y = H(X_1, ..., X_n)$ es un estadístico que toma el valor $y = H(x_1, ..., x_n)$

Observaciones.

1. Una muestra aleatoria se puede considerar como n mediciones de la variable X.
2. Un Estadístico es una variable aleatoria!!!

Teorema

Sea X v.a. con $\mathbb{E}(X) = \mu$, y $Var(X) = \sigma^2$. Sea $\overline{X}$ el promedio muestral de una muestra aleatoria simple de tamaño n. Entonces

- $\mathbb{E}(\overline{X}) = \mu$
- $Var(\overline{X}) = \frac{\sigma^2}{n}$
- Para n grande $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ tiene aproximadamente la distribución $N(0, 1)$. (En caso que n sea pequeño, se debe calcular la distribución de $\overline{X}$.)

Definición. Estimador.

Sea X v.a. con cierta distribución de probabilidad dependiente de un parámetro θ. Sea $X_1, ..., X_n$ una muestra de X y sean $x_1, ..., x_n$ los valores muestrales. Si $g(X_1, ..., X_n)$ es una función de la muestra que va a ser usada para estimar θ, nos referimos a g como un estimador de θ. Llamaremos a $\hat{\theta} = g(X_1, ..., X_n)$ estimación de θ.

Observación.

θ es un número real desconocido (pero no aleatorio), mientras que $\hat{\theta}$ es una variable aleatoria.

Definición. Estimador Insesgado.

Sea $\hat{\theta}$ una estimación del parámetro θ asociado a la distribución de la v.a. X. Entonces $\hat{\theta}$ es un estimador insesgado para θ si $\mathbb{E}(\hat{\theta}) = \theta$ para cualquier θ.

Definición. Estimador de Varianza Mínima.

Sea $\hat{\theta}$ una estimación insesgada de θ. Diremos que $\hat{\theta}$ es una estimación de varianza mínima de θ si para todas las estimaciones $\theta*$ tales que $\mathbb{E}(\theta*) = \theta$, tenemos $Var(\hat{\theta}) \leq Var(\theta*)$ para cualquier θ. Es decir, entre todas las estimaciones insesgadas de θ, $\hat{\theta}$ tiene la varianza más pequeña.

Definición. Estimador Convergente.

Sea $\hat{\theta}$ una estimación θ. Se dice que $\hat{\theta}$ es una estimación convergente a θ si:

$$\lim_{n\to\infty} \mathbb{P}\left(\mid \hat{\theta}-\theta \mid > \epsilon\right) = 0 \ \ \forall \epsilon > 0$$

o equivalentemente

$$\lim_{n\to\infty} \mathbb{P}\left(\mid \hat{\theta}-\theta \mid \leq \epsilon\right) = 1 \ \ \forall \epsilon > 0$$

En general, decir a partir de la definición, si el estimador es convergente es algo difícil. Para facilitar las cosas se tiene el siguiente teorema.

Teorema.

Sea $\hat{\theta}$ una estimación θ. Si $\lim\limits_{n\to\infty} \mathbb{E}(\hat{\theta}) = \theta$ y si $\lim\limits_{n\to\infty} Var(\hat{\theta}) = 0$, entonces $\hat{\theta}$ es una estimación convergente a θ.

Corolario.

Si X es una v.a. tal que $\mathbb{E}(X) = \mu$ y $Var(X) = \sigma^2$. Se tiene que $\overline{X}$ es un estimador insesgado y convergente a μ.

Definición. Error Cuadrático Medio.

Se define al error cuadrático medio de un estimador $\hat{\theta}$ de θ como

$$\begin{aligned} ECM(\hat{\theta}) &= \mathbb{E}\left(\left(\hat{\theta}-\theta\right)^2\right) \\ &= Var(\hat{\theta}) - Sesgo^2 \end{aligned}$$

donde $Sesgo = \mathbb{E}(\hat{\theta}) - \theta$.

Definición. Estimador de Máxima Verosimilitud.

El estimador de máxima verosimilitud de θ, llamado $\hat{\theta}$, basado en una muestra $X_1, ..., X_n$, es el valor de θ que maximiza a $L(X_1, ..., X_n; \theta)$, considerado como una función de θ para la muestra dada. L está definida como

$$L(X_1, X_2, ..., X_n; \theta) = f(X_1; \theta) \cdot f(X_2; \theta) \cdot ... \cdot f(X_n; \theta)$$

Observaciones y Propiedades

1. El estimador de máxima verosimilitud no necesariamente va a ser insesgado.
2. Se tiene la propiedad de invarianza. Es decir si $\hat{\theta}$ es el EMV de θ, entonces $g(\hat{\theta})$ es el EMV de $g(\theta)$

INTERVALOS DE CONFIANZA.

Protocolo para la construcción de un intervalo de confianza para un parámetro θ.

1. Encontrar un estimador $\hat{\theta}$ de θ. Ojalá insesgado.
2. Encontrar la distribución de $\hat{\theta}$ o de alguna función del estimador.
3. Encontrar un estadístico, digamos Z, tal que dependa de $\hat{\theta}$ y de θ, pero cuya distribución no dependa de θ.
4. Encontra a y b, relacionados bajo algun criterior, tal que $\mathbb{P}(a < Z < b) = 1 - \alpha$, donde $1 - \alpha$ se denomina, nivel de confianza, que es fijado a priori.
5. En base a los cálculos anteriores construir un intervalo para θ.

EJERCICIOS.

1.- Sea $X \sim Poisson(\lambda)$. Sea desea construir un estimador insesgador para $e^{-2\lambda}$ a partir de una muestra de tamaño uno. Demuestre que este estimador es:

$$\delta(X) = \begin{cases} 1 & si\, X\, par \\ -1 & si\, X\, impar \end{cases}$$

2.- Se podría suponer (erradamente) que siempre puede encontrarse un estimador insesgado para un parámetro desconocido. Que esto no es así se ilustra con el ejemplo siguiente. Sea p el parámetro de una distribución Binomial. Muestre que no existe un estimador insesgado de

$$\theta = \frac{p}{1-p}$$

3.- Una variable aleatoria X tiene distribución $N(\mu, 1)$. Se toman veinte observaciones de X, pero, en vez de anotar su valor observamos sólo si X era negativa o no. Suponiendo que el suceso $\{X < 0\}$ ocurrió exactamente 14 veces, utilizar esta información para obtener la estimación de Máxima Verosimilitud de μ.

4.- Suponga que T, el tiempo para fallar (en horas) de un instrumento electrónico, tiene la siguiente fdp:

$$\begin{aligned} f(t) &= \beta e^{-\beta(t-t_0)} \;\; t > t_0 > 0 \\ &= 0 \;\; para\, cualquier\, otro\, valor \end{aligned}$$

Suponga que se prueban n artículos y que se anotan los tiempos para fallar $T_1, ..., T_n$.
a) Suponiendo que t_0 es conocido, obtener el estimador de Máxima Verosimilitud de β.
b) Suponiendo que t_0 es deconocido, encontrar el estimador de Máxima Verosimilitud de t_0.

5.- Suponga una muestra de tamaño n de $X \sim N(\mu, \sigma^2)$.
a) Encuentre un intervalo de confianza para μ suponiendo σ desconocido.
b) Se dispone de 10 muestras de sangre tomadas en las mismas condiciones a una misma persona. Se obtiene para cada una la dosis de Colesterol (en gramos) 245, 248, 250, 247, 249, 247, 247, 246, 246, 248. Cada medida puede considerarse como una realización particular de la variable "tasa de Colesterol" $X \sim N(\mu, \sigma^2)$.
i) Dé un intervalo de confianza para μ al 95 % suponiendo $\sigma^2 = 1,5$
ii) Dé un intervalo de confianza para μ al 95 % suponiendo σ^2 desconocido.