Resumen de Estadística MA34B

April 8, 2006

Profesor: Alexis Peña. Realizado por el auxiliar Diego Díaz Espinoza

1 Definiciones fundamentales

Población: Conjunto de variables aleatorias observables (estatura, tiempo de duración, posición en el espacio, etc.). Cada una de ellas tiene asociada una función de probabilidad (para el caso entero) o función densidad de probabilidad (para el caso real). Esta función depende de ciertos parámetros que pueden o no ser conocidos, pero todos observables. En este curso sólo se considerarán poblaciones cuyas variables son iid (independientes e idénticamente distribuidas, es decir, todas siguen la misma función de probabilidad o densidad de probabilidad.

Muestra aleatoria: Es un vector - de cualquier tamaño - cuyos componentes pertenecen a la población. Cada una de estas componentes es iid y todas siguen la misma distribución. El espacio donde asume valores la muestra aleatoria se denomina *Espacio Muestral*.

Función densidad conjunta: Es la distribución que siguen todos los componentes de la muestra aleatoria en conjunto. Recordar que estas funciones son familias determinadas por un parámetro; por ejemplo: exp(p), es una exponencial de parámetro p, de esta forma:

$$exp(x/p) = \lambda e^{-\lambda x} \tag{1}$$

Es una función en x pero tambieén puede ser vista como función en p.

Espacio de parámetros: Es el conjunto de todos los valores posibles para el parámetro de la función densidad de probabilidad o probabilidad.

Estadístico o Estadígrafo: Es una función sobre una muestra aleatoria que devuelve valores en algún espacio. Si los valores devueltos pertenecen al espacio de parámetros entonces se le llama Estimador. Notar que un estadístico es una variable alaeatoria, pues contiene sólo variables aleatorias más operaciones entre ellas. Se les escribirá como $\widehat{\theta}$

Estimación puntual: Escoger un valor para el parámetro buscado. Formalmente, es un estaístico que entrega valores pertenecientes al espacio de parámetros.

2 Problema general de Estadística

Se desea saber la distribución conjunta que sigue una población. Como generalmente no se pueden conocer todas las variables que componen una población, se extrae una muestra aleatoria de ella. Nótese que de conocerse todas las variables - y sus valores - que componen la población el problema está resuelto.

3 Método general de solución

Se extrae una muestra aletatoria de la población, de la cual se conocen la función densidad de probabilidad para cada componente. Se calcula la función densidad de probabilidad acumulada. Se aplica un estimador a la muestra obteniendo un valor determinado para un parámetro. Se analiza el resultado y se extrae información. En este curso se utilizará principalmente este enfoque, denominado *Clásico*; otro tipo de enfoque es el *Bayesiano*.

4 Estimadores.

Definición 4.1 Función de Verosimilitud
Se define como:

$$L(\theta) \equiv L(\theta, \vec{x}) = f(\vec{x}/\theta) = \prod_{i=1}^{n} f(x_i/\theta) \tag{2}$$

$$\theta \in \Theta \qquad (espacio\,de\,parámetros)$$

$$\vec{x} \in \Xi \qquad (espacio\,muestral)$$

Notar que en esta función θ es el parámetro y x una constante.

Principio 4.1 Principio de la Verosimilitud
Toda la información acerca del parámetro (θ) de la función densidad de probabilidad de una muestra aleatoria ($\vec{x}$) está contenida en la función de verosimilitud ($L(\theta, \vec{x})$).

<u>Definición</u> 4.2 Estadístico suficiente:
Una estadístico es suficiente si observado $T = t$, la muestra $\vec{x}$ no entrega más información acerca de θ [1]

Teorema 4.2 Teorema de Factorización de Neyman
Un estadístico es suficiente - para θ -, si y sólo si, la función de verosimilitud puede escribirse como un producto entre: una función que depende sólo de la muestra aleatoria - $h(\vec{x})$ - y una función que depende del parámetro y el estadístico en cuestión - $g(T(\vec{x}, \theta))$ -. Formalmente:

T es estaístico suficiente para θ si y sólo si:

$$L(\theta, \vec{x}) = g(T(\vec{x}), \theta)h(\vec{x}) \tag{3}$$

Con $g, h \geq 0$.

4.1 Propiedades deseables para un estimador:

Sesgo: Un estimador es insesgado si la esperanza del mismo es igual al parámetro del la función densidad de probabilidad conjunta. i.e. $E(\widehat{\theta}) = \theta$. Esperamos que nuestro estimador sea *insesgado*.

Error cuadrático medio: Se define como:
$ECM(\widehat{\theta}) = E[(\widehat{\theta} - \theta)^2]$
Es una medida de cuán *lejos* se está del *verdadero* valor de θ.

1. Para estimadores insesgados se tiene:

   $ECM(\widehat{\theta}) = Var(\widehat{\theta})$

2. En general:

   $ECM(\widehat{\theta}) = Var(\widehat{\theta}) + sesgo(\widehat{\theta})^2$

   Donde:

   $sesgo(\widehat{\theta}) = E(\widehat{\theta}) - \theta$

[1] Para una mejor definición ver *Probabilidad y Estadística, Morris H. DeGroot, pág. 340.*

Eficiencia relativa: Es la eficiencia entre dos estimadores. ie:

$$ef(\widehat{\theta},\widetilde{\theta}) = \frac{Var(\widehat{\theta})}{Var(\widetilde{\theta})}$$

Así, $\widehat{\theta}$ es más eficiente que $\widetilde{\theta}$ si: $Var(\widehat{\theta}) < Var(\widetilde{\theta})$

<u>**Definición**</u> **4.3** Información de Fisher
Se define como:

$$I_x(\theta) = E[\Big(\frac{\partial log(f(x/\theta))}{\partial\theta}\Big)^2]$$
$$I_n(\theta) = I_x(\theta)n$$

Lo que entrega la cantidad de información esperada de $\vec{x}$ acerca de θ
Bajo ciertas condiciones de regularidad (BCR) [2] *se tiene:*

1. $$E[\frac{\partial log(f(\vec{x},\theta))}{\partial\theta}] = 0$$

2. $$I_n(\theta) = -E[\frac{\partial^2 log(f(\vec{x},\theta))}{\partial\theta^2}]$$

Eficiencia de un estimador: Se dice que un estimador es eficiente si

$$Var(\widehat{\theta}) = \frac{1}{I_n(\theta)}$$

Cota de Crámer-Rao: Si un estimador es insesgado entonces

$$Var(\widehat{\theta}) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI_x(\theta)}$$

[2] Ver Crámer: Métodos Matemáticos de Estadística.

4.2 Métodos Clásicos de Estimación Puntual.

Algunos métodos para estimaciónes:

1. Estimación de momentos EMM:

$$\mu_k = E[x^k], \qquad k = 1, 2.... \tag{4}$$

$$m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \qquad k = 1, 2.... \tag{5}$$

   Así las ecuaciones de momento se definen por:

$$\mu_k(\theta) = m_k = \frac{\sum_{i=1}^{n} x_i^k}{n}, \qquad k = 1, 2.... \tag{6}$$

   En general estos estimadores no son insesgados.

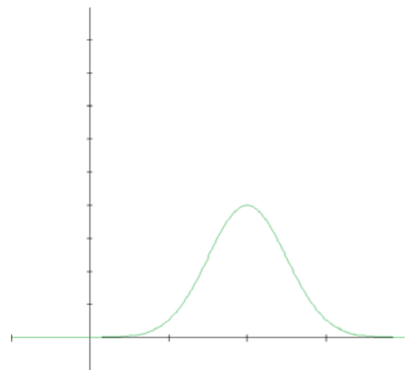

Figure 1: Función densidad de probabilidad conjunta siguiendo una distribución normal. El máximo está en dos unidades.

2. Estimador Máximo Verosimil (EMV):
   Sabemos que la función de verosimilitud es igual a la función densidad conjunta o acumulada de $\vec{x}$ dado θ; esto es la probabilidad de sacar el vector $\vec{x}$ dado que nuestro parámetro es θ. Supongamos ahora dos parámetros θ_0 y θ_1 y que $f_n(\vec{x}, \theta_1) > f_n(\vec{x}, \theta_0)$, esto indicaría que la probabilidad de encontrar el vector $\vec{x}$ dado que el parámetro es igual a θ_1 es mayor que si el parámetro fuese θ_0. Por lo cual sería mucho más atinado escoger a θ_1 como estimación de θ. La idea del estimador máximo verosimil es esa; encontrar el máximo de la funcón verosimilitud respecto al parámetro θ, dejando constante el vector $\vec{x}$.

4.2.1 Propiedades de un Estimador Máximo verosimil

1. De existir un estadístico suficiente para θ el EMV de θ depende de la muestra $\vec{x}$ através de dicho estadístico.

2. Los EMV no son necesariamente insesgados, pero sí asintóticamente insesgados.

3. Los EMV son asíntóticamente insesgados. Además siguen la siguiente distribucón:
   $$\widetilde{\theta} \sim N\Big(\theta, \frac{1}{I_n(\theta)}\Big)$$

4. *Propiedad de invarianza:*
   Si $\widetilde{\theta}$ es el EMV de θ entonces $\varphi(\widetilde{\theta})$ es el EMV de $\varphi(\theta)$

5. Usando 3 y 4 se obtiene:
   Si $\widetilde{\theta}$ es el EMV de θ entonces $\varphi(\widetilde{\theta}) \sim N(\varphi(\theta), \frac{(\varphi(\theta)')^2}{I_n(\theta)})$. Y análogamente $\varphi(\widetilde{\theta})$. Además $\varphi(\widetilde{\theta})$ alcanza asintóticamente la cota de Crámer-Rao, es decir es asintóticamente eficiente.

4.2.2 Estimadores consistentes.

Si:

$$E(\widetilde{\theta_n}) \stackrel{n\to\infty}{\longrightarrow} \theta$$

y

$$Var(\widetilde{\theta_n}) \stackrel{n\to\infty}{\longrightarrow} 0$$

Entonces se dice que $\widetilde{\theta_n}$ es un estimador consistente de θ
Estimador consisten implica que:

$$\lim_{n\to\infty} P_r\{|\widetilde{\theta_n} - \theta| > \epsilon\} = 0, \qquad \epsilon > 0$$

ó, equivalentemente:

$$\lim_{n\to\infty} P_r\{|\widetilde{\theta_n} - \theta| < \epsilon\} = 1, \qquad \epsilon > 0$$