Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática
MA3403 - Probabilidades y Estadística

Estimación Puntual

Alberto Vera Azócar, albvera@ing.uchile.cl

1. Modelos Paramétricos

El contexto que nos interesa ahora es que estamos observando una variable aleatoria y sabemos la forma de su densidad, pero ésta depende de un parámetro θ que desconocemos.

El ejemplo clásico es que estamos observando como se lanza una moneda y podemos tomar todos los datos que queramos (como número de caras y sellos obtenidos), además de ésto estamos todos convencidos que el lanzamiento de una moneda sigue un modelo Bernoulli(p), pero no sabemos el valor del parámetro p, por supuesto que nos gustaría saber cuanto vale p y es el trabajo de la estadística estimarlo adecuadamente.

Definición 1.1 (**Modelo Paramétrico**). *Dado $\Theta \subseteq \mathbb{R}^k$, diremos que la familia dada por $\{f(x|\theta) : \theta \in \Theta\}$ constituye un modelo paramétrico cuando $\forall\, \theta \in \Theta$ se tiene que $f(x|\theta)$ es función de densidad (o bien ley de probabilidad).*

Ejemplo 1.1. *Considerando el ejemplo de la moneda, identificando $x = 1$ por cara y $x = 0$ por sello, tenemos que*

$$f(x|p) = p^x(1-p)^{1-x} \qquad x \in \{0, 1\}$$

Por supuesto acá el conjunto de los valores posibles de p es $\Theta = [0, 1]$.

Ya tenemos definida una familia de funciones que estudiaremos, para poder deducir propiedades necesitaremos datos. Nos concentraremos solo en aquellos datos que son obtenidos mediante un muestreo aleatorio simple (m.a.s.). La forma más intuitiva de pensar en un m.a.s. es que tenemos una población de N individuos y queremos muestrear n datos de una característica de ellos, para eso ponemos N bolitas en una urna y extraemos de a una **con reposición** n bolitas.

Definición 1.2 (**Muestreo Aleatorio Simple**). *Una colección $X_1, \cdots, X_n$ de variables aleatorias constituyen un m.a.s. si provienen de la misma distribución y son independientes.*

Si tenemos datos $X_1, \cdots, X_n$ que entendemos como "realizaciones de la v.a." queremos estimar el valor de θ. A una función de aquellos datos (y sólo de datos) la llamaremos "estimador".

Definición 1.3 (**Estimador**). *Dados los valores muestrales $X_1, \cdots, X_n$ correspondientes al modelo paramétrico $\{f(x|\theta) : \theta \in \Theta\}$, un estimador es cualquier función de los valores muestrales $\hat{\theta}(X_1, \cdots, X_n)$ tal que $\hat{\theta} \in \Theta$ casi seguramente.*

Ejemplo 1.2. *Supongamos el modelo exponencial, donde $f(x|\lambda) = \lambda e^{-\lambda x}\mathbb{1}_{\{x\geq 0\}}$ y $\Theta = (0, \infty)$. Luego según la definición, cualquier función de los datos $X_1, \cdots, X_n$ tal que sea positiva con probabilidad 1 se llamará estimador.*

En particular $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^n X_i$ es estimador, pues como cada X_i es positivo el promedio también lo será. Así también $\hat{\lambda}_2 = \max\{X_1, \cdots, X_n\}$ es estimador. Por otro lado $\hat{\lambda}_3 = -X_1$ no es estimador.

2. Propiedades de los Estimadores

Por la definición de estimador casi cualquier cosa puede ser un estimador, por esto es que formalizaremos la idea de un "buen estimador" (en el ejemplo anterior nos gustaría saber cuál es mejor de entre $\hat{\lambda}_1$ y $\hat{\lambda}_2$). Para esto pensaremos de ahora en más que el modelo paramétrico está fijo y que tenemos una muestra $X_1, \cdots, X_n$ de tal modelo.

Definición 2.1 (**Sesgo**). *Sea $\hat{\theta}$ un estimador de θ, definimos $Sesgo(\hat{\theta}) := \mathbb{E}(\hat{\theta}) - \theta$. Además diremos que $\hat{\theta}$ es insesgado si $Sesgo(\hat{\theta}) = 0$.*

Definición 2.2 (**Consistencia**). *Sea $\hat{\theta}$ un estimador de θ, diremos que $\hat{\theta}$ es consistente si converge en probabilidad a θ, esto es: $\forall \varepsilon > 0$, $\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \to 0$ cuando $n \to \infty$.*

En la práctica la definición anterior no es muy fácil de aplicar, por eso siempre que nos interese saber la consistencia de un estimador usaremos el teorema siguiente que aceptaremos sin demostración.

Teorema 2.1. *Sea $\hat{\theta}$ un estimador de θ, se tiene que si $\mathbb{V}ar(\hat{\theta}) \to 0$ y $\mathbb{E}(\hat{\theta}) \to \theta$ cuando $n \to \infty$, entonces $\hat{\theta}$ es estimador consistente.*

Definición 2.3 (**Error Cuadrático Medio**). *Sea $\hat{\theta}$ un estimador de θ, definimos el error cuadrático medio (e.c.m.) de $\hat{\theta}$ como*

$$ECM(\hat{\theta}) := \mathbb{E}\left((\hat{\theta} - \theta)^2\right)$$

Las definiciones que hemos dado dependen claramente del valor de θ y además de n (el tamaño muestral). Obviamente los buenos estimadores serán aquellos consistentes, es decir, los que tienen sesgo y varianza pequeña, veamos que esto equivale a tener e.c.m. convergente a cero:

Proposición 2.1. *Sea $\hat{\theta}$ estimador de θ, se tiene que*

$$ECM(\hat{\theta}) = \mathbb{V}ar(\hat{\theta}) + Sesgo(\hat{\theta})^2$$

Demostración.

$$\begin{aligned}
ECM(\hat{\theta}) &\triangleq \mathbb{E}\left((\hat{\theta} - \theta)^2\right) \\
&= \mathbb{E}\left(([\hat{\theta} - \mathbb{E}(\hat{\theta})] - [\theta - \mathbb{E}(\hat{\theta})])^2\right) \\
&= \mathbb{E}\left([\hat{\theta} - \mathbb{E}(\hat{\theta})]^2 - 2[\hat{\theta} - \mathbb{E}(\hat{\theta})][\theta - \mathbb{E}(\hat{\theta})] + [\theta - \mathbb{E}(\hat{\theta})]^2\right) \\
&= \mathbb{E}\left([\hat{\theta} - \mathbb{E}(\hat{\theta})]^2\right) - 2[\theta - \mathbb{E}(\hat{\theta})]\mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) + [\theta - \mathbb{E}(\hat{\theta})]^2 \\
&= \mathbb{V}\mathrm{ar}(\hat{\theta}) - 2[\theta - \mathbb{E}(\hat{\theta})]\mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) + Sesgo(\hat{\theta})^2 \\
&= \mathbb{V}\mathrm{ar}(\hat{\theta}) - 2[\theta - \mathbb{E}(\hat{\theta})]\left(\mathbb{E}(\hat{\theta}) - \mathbb{E}(\hat{\theta})\right) + Sesgo(\hat{\theta})^2 \\
&= \mathbb{V}\mathrm{ar}(\hat{\theta}) + Sesgo(\hat{\theta})^2
\end{aligned}$$

□

Ejemplo 2.1. *Sea $X_1, \cdots, X_n$ un m.a.s. del modelo uniforme en $[0, \theta]$, con $\theta > 0$ desconocido. Se pide estudiar la consistencia del estimador $\hat{\theta} := \frac{2}{n}\sum_{i=1}^{n} X_i$.*

Solución: Observamos que si $X \sim \mathcal{U}(0,\theta)$, entonces $\mathbb{E}(X) = \frac{\theta}{2}$, por lo que $2\mathbb{E}(X) = \theta$, es decir, el promedio es la mitad del intervalo, por lo que si lo multiplicamos por 2 se tiene el extremo del intervalo. Tiene entonces sentido el estimador $\hat{\theta}$ que es 2 veces el promedio empírico.

$$\begin{aligned}
\mathbb{E}(\hat{\theta}) &= \mathbb{E}\left(\frac{2}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{2}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) \\
&= \frac{2}{n}\sum_{i=1}^{n} \frac{\theta}{2} \\
&= \theta
\end{aligned}$$

Sigue que $\hat{\theta}$ es insesgado, luego la consistencia depende sólo de la varianza. Usaremos que la varianza de una uniforme en $[a,b]$ es conocida y vale $\frac{(b-a)^2}{12}$

$$\begin{aligned}
\mathbb{V}ar(\hat{\theta}) &= \mathbb{V}ar\left(\frac{2}{n}\sum_{i=1}^{n} X_i\right) \\
\textit{(por independencia)} \quad &= \left(\frac{2}{n}\right)^2 \sum_{i=1}^{n} \mathbb{V}ar(X_i) \\
&= \frac{4}{n^2}\sum_{i=1}^{n} \frac{\theta^2}{12} \\
&= \frac{\theta^2}{3n}
\end{aligned}$$

Sigue que $\mathbb{V}ar(\hat{\theta}) \to 0$ cuando $n \to \infty$, por lo tanto el estimador es consistente. ¿Tiene sentido que la varianza crezca con el valor de θ?

3. Máxima Verosimilitud

Ahora lo que buscamos es un método para obtener estimadores, por supuesto hay varias técnicas, pero veremos solo una que es la más usada.

Definición 3.1 (**Función de Verosimilitud**). *Dada una muestra de v.a. $X_1, \cdots, X_n$ del modelo paramétrico $f(x|\theta)$, definimos la función de verosimilutd del parámetro θ como*

$$L(\theta|X_1, \cdots, X_n) := \prod_{i=1}^{n} f(X_i|\theta)$$

Observemos primero que una densidad depende de θ y es función de x, pero la función de verosimilitud no. Por el contrario, tomamos θ como variable y fijamos los valores X_i.

Por otro lado, si las v.a. $X_1, \cdots, X_n$ fueran independietes y cada una con densidad $f(x_i)$, entonces escribimos la densidad conjunta $f_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$. Por lo tanto la función de verosimilitud no es nada más que la densidad conjunta (asumiendo independencia).

Definición 3.2 (**Estimador Máximo Verosímil**). *Diremos que $\hat{\theta}$ es un estimador de máxima verosimilitud de θ si:*

$$\hat{\theta} \in \underset{\theta \in \Theta}{argmáx}\, L(\theta | X_1, \cdots, X_n)$$

Ejemplo 3.1. *Sea $X_1, \cdots, X_n$ un m.a.s. del modelo exponencial, con λ desconocido. Se pide estimar mediante máxima verosimilitud el parámetro.*

Solución: Se tiene que $f(x|\lambda) = \lambda e^{-\lambda x}$, luego la función de verosimilitud es

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \lambda^n \prod_{i=1}^{n} e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}$$

Denotando $\bar{X}$ como el promedio empírico, tenemos que $\sum_{i=1}^{n} X_i = n\bar{X}$ por lo que podemos escribir

$$L(\lambda) = \lambda^n e^{-n\lambda \bar{X}}$$

Ahora buscamos el $\lambda \in (0, \infty)$ que maximice L, entonces derivamos para obtener las condiciones de primer orden

$$n\lambda^{n-1} e^{-n\lambda \bar{X}} - n\lambda^n \bar{X} e^{-n\lambda \bar{X}} = 0$$

Despejando la ecuación para λ se obtiene $\lambda = \bar{X}^{-1}$, es decir, el estimador máximo verosímil es el inverso del promedio empírico.