Inferencia Estadística

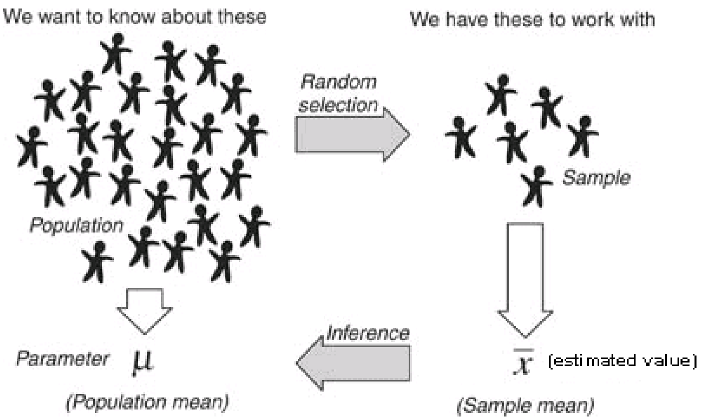
Principios de Reducción

Gabriel Martos Venturini
gmartos@utdt.edu

UTDT

UNIVERSIDAD
TORCUATO DI TELLA

Recapitulación



- Modelos estadísticos (familia exponencial y localización–escala).
- Estimador/estadístico: Principios y propiedades.

- Un estadístico $T_n(\underline{X})$ (donde $\underline{X} \equiv \{X_1, \ldots, X_n\}$) es una v.a. con la que típicamente intentamos estimar cierto parámetro de interés:

$$\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Bern}(\theta), \text{ y consideremos } T_n(\underline{X}) = \sum_{i=1}^{n} X_i/n.$$

- T_n constituye una *reducción* de $\underline{X}$ y debería cumplir ciertos principios:
    - *Suficiencia*: No queremos *perder información* al pasar de $\{X_1, \ldots, X_n\}$ a T_n cuando intentamos estimar (en cualquier sentido) a θ.
    - *Verosimilitud*: Una vez realizada la muestra, toda la información relevante sobre θ está contenida en la función de verosimilitud.
    - *Invarianza*: Las conclusiones que sacamos de $\{X_1 = x_1, \ldots, X_n = x_n\}$ no deberían cambiar cuando la escala en la que observamos los datos y los parámetros del modelo estadístico cambian de manera compatible.

Agenda

1. Principio de Suficiencia
   - Definición y algunos ejemplos
   - Ancillaridad y Completitud (Teorema de Basu)
2. Principio de Verosimilitud
3. Principio de Invarianza
4. Apéndice

Agenda

- Sea $\underline{X} = \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ y $T_n \equiv T(\underline{X})$.

Informal: T_n es suficiente para el parámetro θ si **contiene toda la información** que hay en la muestra aleatoria $\underline{X}$ respecto de θ.

- Si T_n es un estadístico suficiente para θ, entonces la información respecto de θ que tenemos en $\underline{x}$ y $T_n(\underline{x}) = t$ es equivalente.

- El estimador de θ debería ser función de un estadístico suficiente.

Definition (Suficiencia)

Un estadístico T_n es suficiente para el parámetro θ si la distribución condicional de $\underline{X} \mid T_n = t$ no depende de θ. Es decir se cumple que:

$$P\big(\underline{X} = \underline{x} \mid T_n(\underline{x}) = t\big) \text{ es constante respecto de } \theta.$$

- La suficiencia es una propiedad asociada al modelo estadístico.

- En la practica no es sencillo determinar la distribución de $\underline{X} \mid T_n = t$.

Condición a verificar para la suficiencia

- Si $\underline{X} = \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ luego $\underline{X} \sim f(\underline{x};\theta) \overset{iid}{=} \Pi_{i=1}^{n} f(x_i;\theta)$.

Teorema (Factorización de Fisher–Neyman)

Un estadístico T_n es suficiente para θ si y solo si existen funciones $g(T_n(\underline{x});\theta)$ y $h(\underline{x})$ tales que podemos factorizar $f(\underline{x};\theta)$ como:

$$f(\underline{x};\theta) \equiv g(T_n(\underline{x});\theta)\, h(\underline{x}) \text{ para todo } \underline{x} \text{ y } \theta \in \Theta.$$

Remark: Notar que $g(T_n(\underline{x});\theta) \geq 0$ solo puede depender de los datos a través de T_n y que $h(\underline{x}) \geq 0$ no puede depender del parámetro.

- Ejemplo 1: Cantidad de éxitos cuando $X \sim \text{Bern}(\theta)$.
- Ejemplo 2: Media muestral cuando $X \sim N(\theta, \sigma_0^2)$.
- Ejemplo 3: El máximo en la muestra cuando $X \sim U(0,\theta)$.

Suficiencia en familias exponenciales de 1 parámetro

- Sea $\underline{X} := \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta) = h(x)c(\theta)\exp\Big(w(\theta)t(x)\Big)$.
- El estadístico $T_n \equiv \sum_{i=1}^n t(X_i)$ es suficientes para θ.

$$
\begin{aligned}
\underline{X} \sim f(\underline{x};\theta) &= \Pi_{i=1}^n h(x_i)c(\theta)\exp\big(t(x_i)w(\theta)\big), \\
&= \Big(\underbrace{\Pi_{i=1}^n h(x_i)}_{h(\underline{x})}\Big)\underbrace{c^n(\theta)\exp\Big(\sum_{i=1}^n t(x_i)w(\theta)\Big)}_{g(T_n(\underline{x});\theta)}.
\end{aligned}
$$

- Ejemplo: Modelo Geométrico en la familia exponencial.

$$f(x;\theta) = \mathbb{1}_{[\mathbb{N}]}(x)\theta(1-\theta)^{x-1}, \text{ luego } t(X) = X.$$

  - Por lo tanto $T_n = \sum_{i=1}^n X_i$ es suficiente para θ.

Familias exponenciales de k parámetros

- Sea $\underline{X} := \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x; \boldsymbol{\theta})$, con $f(x; \boldsymbol{\theta})$ en la familia exponencial y $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, entonces:

$$f(x; \boldsymbol{\theta}) = h(x) c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right), \text{ donde } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k).$$

- Notar que X_1 puede ser una variable o un vector aleatorio.
- Luego los estadísticos:

$$\mathbf{T}_n(\underline{X}) = \left(T_{n,1}(\underline{X}) = \sum_{j=1}^{n} t_1(X_j), \ldots, T_{n,k}(\underline{X}) = \sum_{j=1}^{n} t_k(X_j)\right),$$

son suficientes para $\boldsymbol{\theta}$.

- Ejemplo: Modelo normal ($k = 2$).

Recapitulación

- *Principio de Suficiencia*: **Es deseable que la inferencia sobre el parámetro de interés** esté basada en un **estadístico suficiente**.
- Criterio de factorización de Fisher–Neyman / Familia Exponencial.
- La suficiencia es una propiedad vinculada al modelo estadístico.
- Una transformación biyectiva de un estadístico suficiente da como resultado otro estadístico suficiente, en otras palabras no son únicos.
  - Ejemplo: $T_n = \sum_{i=1}^{n} X_i$ es suficiente para θ cuando $X \sim \text{Bern}(\theta)$ entonces también lo es $\overline{X}_n = T_n/n$ (transformación biyectiva).
- En familias exponenciales los estimadores máximo verosímiles son funciones de los estadísticos suficientes (lo formalizamos en S3).
  - En caso de existir un estimador UMVUE, éste es función de un estadístico suficiente (ver Lehmann—Scheffé en S4).

Estadístico suficiente minimal

Para $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$:

- En general existen muchos estadísticos suficientes, el mismo vector aleatorio $\underline{X}$ es un estadístico suficiente (pero poco reductor) para θ.
- La suficiencia minimal apunta a identificar estadísticos que compacten al máximo la información de la muestra sin perder información de θ.

Definition (Estadístico suficiente minimal)

T_n es **minimal suficiente** para θ si para cualquier otro estadístico suficiente T_n', se cumple que si $T_n'(\underline{x}) = T_n'(\underline{y})$ entonces $T_n(\underline{x}) = T_n(\underline{y})$.

- $T_n(\underline{X}) = \underline{X}$ NO es minimal suficiente cuando $X \sim \text{Bern}(\theta)$.
  - Para $\underline{x} = \{1,1,0\}$ e $\underline{y} = \{1,0,1\}$: $T_n(\underline{x}) = \underline{x} \neq T_n(\underline{y}) = \underline{y}$.
  - Sin embargo: $T'(\underline{x}) = \sum_{i=1}^{3} x_i = 2 = T'(\underline{y})$ (T' es suficiente para θ).

Criterio para suficiencia minimal

Sea $f_{\underline{X}}(\underline{x};\theta)$ la densidad de $\underline{X}$, T_n es minimal suficiente para θ cuando:

$$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} \text{ es constante para todo } \theta \in \Theta \iff T_n(\underline{x}) = T_n(\underline{y})$$

Es decir que se cumplen, simultáneamente, las siguientes dos condiciones:

- Si para un par de muestras (datos) $\underline{x}$ e $\underline{y}$ se cumple que:

  $$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} \text{ es constante para todo } \theta \in \Theta,$$

  entonces debe verificarse que $T_n(\underline{x}) = T_n(\underline{y})$.

- Si para un par de muestras $\underline{x}$ e $\underline{y}$ se cumple que $T_n(\underline{x}) = T_n(\underline{y})$, entonces tiene que verificarse que:

  $$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} \text{ es constante para todo } \theta \in \Theta.$$

Ejemplo: $X \sim \text{Bern}(\theta)$ ($\theta \in (0,1)$) y $T_n = \sum_{i=1}^n X_i$

$$\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} = \frac{\theta^{T_n(\underline{x})}(1-\theta)^{n-T_n(\underline{x})}}{\theta^{T_n(\underline{y})}(1-\theta)^{n-T_n(\underline{y})}} = \frac{\left(\frac{\theta}{1-\theta}\right)^{T_n(\underline{x})}}{\left(\frac{\theta}{1-\theta}\right)^{T_n(\underline{y})}} = \left(\frac{\theta}{1-\theta}\right)^{T_n(\underline{x})-T_n(\underline{y})}.$$

- $\frac{f_{\underline{X}}(\underline{x};\theta)}{f_{\underline{X}}(\underline{y};\theta)} = 1 \iff T_n(\underline{x}) = T_n(\underline{y})$, para todo $\theta \in (0,1)$.
- Luego T_n es minimal suficiente.
- Bajo condiciones generales, en la familia exponencial:

$$X \sim f(x;\theta) = h(x)c(\theta)\exp\Big(w(\theta)t(x)\Big),$$

  $T_n = \sum_{i=1}^n t(X_i)$ es suficiente y minimal para θ.
- Un estadístico minimal suficiente no es necesariamente único.
  - Ej: $T_n = \sum_{i=1}^n X_i$ y $T_n' = \sum_{i=1}^n X_i/n$ con $X \sim \text{Bern}(\theta)$.

Agenda

1. Principio de Suficiencia
   - Definición y algunos ejemplos
   - Ancillaridad y Completitud (Teorema de Basu)
2. Principio de Verosimilitud
3. Principio de Invarianza
4. Apéndice

Estadístico Ancillar

- El concepto de *Ancillaridad* está contrapuesto al de Suficiencia:
  - S_n se dice ancillar para θ si con S_n no puedo inferir nada respecto de θ.
  - En otras palabras: S_n **NO contiene información sobre** θ.
- Paradójicamente, un estadístico ancillar combinado con otro suficiente para θ, facilitan y enriquecen los métodos de inferencia.
  - Ver ejemplo en § 19.

Definition (Estadístico ancillar)

Un estadístico $S_n(\underline{X})$ es ancillar para θ si la distribución de S_n **NO** depende de θ. En otras palabras, para cualquier conjunto $\mathcal{A}$ se cumple que:

$$P(S_n(\underline{X}) \in \mathcal{A}) \text{ es constante para todo } \theta \in \Theta.$$

Ejemplos

- Para un modelo normal $N(\mu, \sigma^2)$, S_n^2 es ancillar para μ.
  - Puesto que: $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}$.
- Dada $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x; \mu, \theta)$ donde:

$$f(x; \mu, \theta) = \frac{1}{2\theta}, \text{ si } \mu - \theta \leq x \leq \mu + \theta.$$

  - X_1 sigue un modelo de localización a partir de una uniforme en $[-\theta, \theta]$.
- Se puede demostrar que la densidad de $R_n \equiv X_{(n)} - X_{(1)}$ es:

$$R_n \sim f(r, \theta) = \frac{n(n-1)r^{n-2}}{(2\theta)^{n-1}} \left(1 - \frac{r}{2\theta}\right), \text{ si } 0 \leq r \leq 2\theta.$$

  - Luego R_n es ancillar para μ (Ejemplo 4.12 en KK).

Teorema de Basu

- Siendo la Suficiencia y la Ancillaridad conceptos contrapuestos, se tiende a pensar que dos estadísticos T_n y S_n, suficiente y ancillar para θ (en relación al modelo $f(x;\theta)$), deberían ser independientes:

$$P(T_n = t, S_n = s) = P(T_n = t)P(S_n = s) \text{ y } P(T_n = t \,|\, S_n = s) = P(T_n = t)$$

  - Y por lo tanto: $Cov(T_n, S_n) = 0$.

- Generalmente esto es cierto, sin embargo no siempre vale.
  - Un contraejemplo en CB (§ 6.1.11, pp 259).

- **Basu**: Un estadístico T_n minimal suficiente para θ es **independiente** de todos los estadísticos ancillares de θ cuando $f_T(t;\theta)$ es **completo**.
  1. ¿Porqué esto es importante? Necesitamos que el estadístico ancillar y suficiente sean independientes para diseñar herramientas de inferencia.
  2. ¿Cómo aseguramos la completitud?

Completitud

Definition (Completitud)

$T_n \sim f_T(t;\theta)$ es completo si para toda función (medible) g para la que $E_\theta(g(T_n)) = 0 \ \forall \theta \in \Theta$; se cumple que $P_\theta(g(T_n) = 0) = 1 \ \forall \theta \in \Theta$.

- La completitud es una propiedad del modelo estadístico.
- Si $\{X_1, \dots, X_n\} \overset{iid}{\sim} f(x;\theta)$, con $f(x;\theta) = h(x)c(\theta)\exp\Big(w(\theta)t(x)\Big)$.
  - $T_n = \sum_{i=1}^{n} t(X_i)$ es completo y minimal suficiente.

Example

Si $\{X_1, \dots, X_n\} \overset{iid}{\sim} N(\mu, \sigma^2)$ (familia exponencial), $\overline{X}_n$ es completo y (minimal) suficiente respecto de μ. Por otro lado S_n^2 es ancillar para μ. Por el teorema de Basu: $\overline{X}_n$ y S_n^2 resultan variables aleatorias independientes.

Utilidad de los resultados expuestos

- Si $X \sim N(\mu, \sigma^2)$, por T. Basu $\overline{X}_n$ y S_n^2 son independientes, luego*

$$\text{Como } \overline{X}_n \perp S_n^2 \Rightarrow \frac{\frac{(\overline{X}_n - \mu)}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_n^2}{(n-1)\sigma^2}}} = \frac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}} = \frac{(\overline{X}_n - \mu)}{S_n/\sqrt{n}} \sim t_{n-1}.$$

- Nos permite construir intervalos de confianza para μ en poblaciones normales cuando σ^2 es desconocido:

$$P(\overline{X}_n - t_{\alpha/2,n-1}\, S_n/\sqrt{n} \leq \mu \leq \overline{X}_n + t_{\alpha/2,n-1}\, S_n/\sqrt{n}) = 1 - \alpha.$$

*Si $Z \sim N(0,1)$ y $V \sim \chi_n$ son v.a. independientes: $Z/\sqrt{V/n} \sim t_n$.

Recapitulemos:

- T_n suficiente: Contiene toda la información sobre θ en la muestra.
- S_n ancillar: No contiene información sobre θ en la muestra.
- Los estadísticos en familias exponenciales de k parámetros:

$$\mathbf{T}_n(\underline{X}) = \left(T_{n,1}(\underline{X}) = \sum_{j=1}^{n} t_1(X_j), \ldots, T_{n,k}(\underline{X}) = \sum_{j=1}^{n} t_k(X_j) \right),$$

  son minimal suficientes y completos para $\boldsymbol{\theta}$ (en general).
- Basu: Si T_n es (minimal) suficiente y completo $\Rightarrow$ T_n es independiente respecto de cualquier otro estadístico ancillar de θ.
- Si T_j es ancillar respecto del parámetro que estimamos con T_i $(i \neq j)$
  - $P(T_i = t_i | T_j = t_j) = P(T_i = t_i) \Rightarrow \mathrm{Cov}(T_i, T_j) = 0.$

Agenda

Función de Verosimilitud

- Sea $\underline{X} = \underline{x}$ (datos, una realización de $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x; \boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Theta$) definimos la función de verosimilitud como:

$$L(\boldsymbol{\theta}) \equiv L_n(\boldsymbol{\theta}|\underline{x}) = \Pi_{i=1}^{n} f(x_i; \boldsymbol{\theta}).$$

- En el apéndice se discute como construir $L(\boldsymbol{\theta})$ en muestreos NO iid.
- $L(\boldsymbol{\theta})$ debe entenderse como una función de $\boldsymbol{\theta}$, las $\underline{X} = \underline{x}$ están fijas.
- En el caso discreto: $L(\boldsymbol{\theta}) = P_{\theta}(\underline{X} = \underline{x}) = P_{\theta}(\text{Datos} \,|\, \text{Modelo})$.
- $L(\boldsymbol{\theta})$ NO indica probabilidades de $\boldsymbol{\theta}$ (que está fijo en la población!).
- Ejemplo: $X \sim \text{Bern}(\theta)$ y $x_1 = 1, x_2 = 1, x_3 = 0$.
- En la familia exponencial en general se cumple que $L(\boldsymbol{\theta}|\underline{x})$ depende de los datos (fijos) a través del valor que toma el estadístico suficiente.

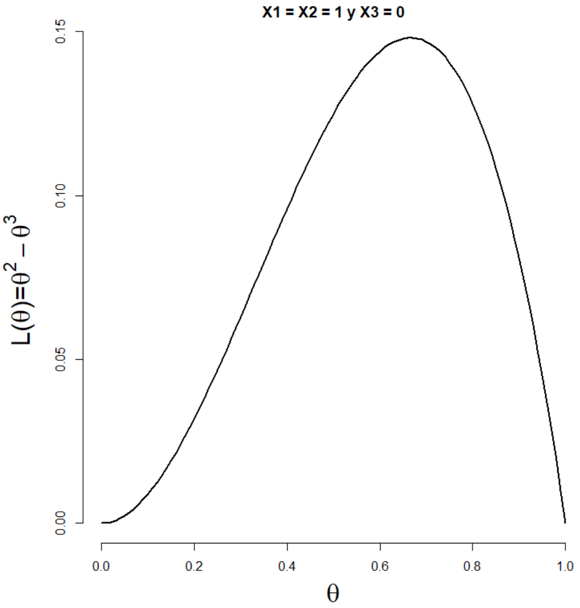


Figure: Función de verosimilitud del ejemplo anterior.

Ratio de verosimilitud

- La función de verosimilitud nos permite juzgar la plausibilidad de observar $\underline{X} = \underline{x}$ entre diferentes valores posibles de $\boldsymbol{\theta} \in \Theta$.

- Comparamos la verosimilitud de $\boldsymbol{\theta}_1$ contra $\boldsymbol{\theta}_2$ a partir del ratio:

$$\frac{L(\boldsymbol{\theta}_1|\underline{x})}{L(\boldsymbol{\theta}_2|\underline{x})} = \frac{P\Big(\text{Observar } \underline{x} \,|\, X \sim f(x;\boldsymbol{\theta}_1)\Big)}{P\Big(\text{Observar } \underline{x} \,|\, X \sim f(x;\boldsymbol{\theta}_2)\Big)}.$$

- ¿Cómo interpretamos el ratio?
  - Si $L(\boldsymbol{\theta}_1|\underline{x})/L(\boldsymbol{\theta}_2|\underline{x}) > 1 \Rightarrow$ Es más factible observar $\underline{X} = \underline{x}$ (datos de la muestra) si en la población $X \sim f(x;\boldsymbol{\theta}_1)$ que si $X \sim f(x;\boldsymbol{\theta}_2)$.
    - $\boldsymbol{\theta}_1$ es más verosímil que $\boldsymbol{\theta}_2$ dada la evidencia $\underline{x}$[1].

- Ej: $X \sim \text{Bern}(\theta)$ con $X_1 = 1, X_2 = 1, X_3 = 0$ y $\theta_1 = 0.5$ vs $\theta_2 = 0.75$.

[1] Y asumiendo que en la población $X \sim f(x;\boldsymbol{\theta})$.

Otros ejemplos con $X \sim \text{Bern}(\theta)$

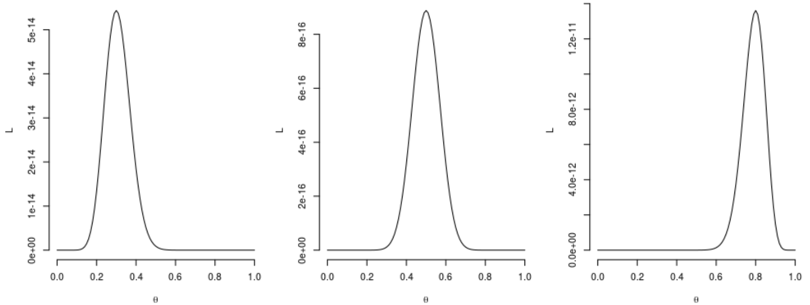


Figure: Izquierda $\sum_{i=1}^{100} X_i = 15$, centro $\sum_{i=1}^{100} X_i = 25$, y derecha $\sum_{i=1}^{100} X_i = 40$.

Interpretación de la gráfica de la <u>izquierda</u>:

- Es más factible que los datos de la muestra se hayan generado de una población $\text{Bern}(\theta \approx 0.3)$ que de una población $\text{Bern}(\theta \approx 0.9)$.
- La inferencia del punto anterior no se modifica si observamos dos muestras distintas $\underline{x}$ y $\underline{x}'$ tales que $T_n(\underline{x}) = T_n(\underline{x}') = \sum_{i=1}^{100} X_i = 15$.

Interpretando la verosimilitud en el caso continuo

- Para todo $\varepsilon > 0$ y dada $\underline{X} = \underline{x}$, $B(\underline{x}, \varepsilon)$ representa el conjunto de todas las muestras "ε–parecidas a $\underline{x}$". Luego se tiene que:

$$P_{\boldsymbol{\theta}}\big(\underline{X} \in B(\underline{x}, \varepsilon)\big) \approx c(\varepsilon, \underline{x}) f_{\underline{X}}(\underline{x}; \boldsymbol{\theta}) = c(\varepsilon, \underline{x}) L(\boldsymbol{\theta} | \underline{x}).$$

- Al igual que en el caso discreto, podemos comparar la plausibilidad de que los datos se generen de un modelo con $\boldsymbol{\theta}_1$ o $\boldsymbol{\theta}_2$ (en un *entorno* de la evidencia $\underline{x}$) utilizando el ratio de verosimilitudes, ya que:

$$\frac{P_{\boldsymbol{\theta}_1}\big(\underline{X} \in B(\underline{x}, \varepsilon)\big)}{P_{\boldsymbol{\theta}_2}\big(\underline{X} \in B(\underline{x}, \varepsilon)\big)} \approx \frac{c(\varepsilon, \underline{x}) L(\boldsymbol{\theta}_1 | \underline{x})}{c(\varepsilon, \underline{x}) L(\boldsymbol{\theta}_2 | \underline{x})} = \frac{L(\boldsymbol{\theta}_1 | \underline{x})}{L(\boldsymbol{\theta}_2 | \underline{x})}.$$

  Por tanto decimos que $\boldsymbol{\theta}_1$ es más verosímil que $\boldsymbol{\theta}_2$, dada la evidencia $\underline{X} = \underline{x}$, si el ratio de verosimilitud es mayor que uno y viceversa.

- Ej: $X \sim \text{Exp}(\theta)$ con $X_1 = 1, X_2 = 3, X_3 = 2$ y $\theta_1 = 2$ vs $\theta_2 = 0.5$.

El principio de verosimilitud

- Sean $\underline{x}$ e $\underline{y}$ dos realizaciones de la muestra tales que

$$L(\boldsymbol{\theta}|\underline{x}) = c(\underline{x}, \underline{y})L(\boldsymbol{\theta}|\underline{y}) \text{ para todo } \boldsymbol{\theta} \in \Theta,$$

  donde $\underline{c(\underline{x}, \underline{y}) \text{ no depende de } \boldsymbol{\theta}}$.

- La inferencia que hacemos respecto de $\boldsymbol{\theta}$ con $\underline{x}$ tiene que ser la misma que hacemos con $\underline{y}$, ya que para cualquier par $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ se cumple:

$$\text{Si} \quad \frac{L(\boldsymbol{\theta}_1|\underline{x})}{L(\boldsymbol{\theta}_2|\underline{x})} > 1 \Rightarrow \frac{L(\boldsymbol{\theta}_1|\underline{x})}{L(\boldsymbol{\theta}_2|\underline{x})} = \frac{c(\underline{x}, \underline{y})L(\boldsymbol{\theta}_1|\underline{y})}{c(\underline{x}, \underline{y})L(\boldsymbol{\theta}_2|\underline{y})} = \frac{L(\boldsymbol{\theta}_1|\underline{y})}{L(\boldsymbol{\theta}_2|\underline{y})} > 1.$$

- Las conclusiones que extraemos de $\underline{x}$ e $\underline{y}$ son las mismas:

$$L(\boldsymbol{\theta}_1|\underline{x}) > L(\boldsymbol{\theta}_2|\underline{x}) \Leftrightarrow L(\boldsymbol{\theta}_1|\underline{y}) > L(\boldsymbol{\theta}_2|\underline{y}).$$

- El principio de verosimilitud sugiere entonces que $L(\boldsymbol{\theta})$ contiene toda la información relevante para hacer inferencias respecto de $\boldsymbol{\theta}$.

Agenda

- Invarianza de Escala: Si el modelo $f(x;\theta)$ y la evidencia empírica $\underline{x}$ están fijos, los cambios de escala en $\underline{x}$ solo deben producir cambios de escala en las estimaciones que hacemos de θ. Discutamos un ejemplo:
  - Dos investigadores utilizan el mismo modelo $f(x;\theta)$ y la misma muestra $\underline{x}$. Al primero se le reporta $\underline{x}$ en escala de metros, al segundo en escala de centímetros. La inferencia de ambos debe ser compatible.
- Si hacemos inferencia sobre θ utilizando la función de verosimilitud, entonces el procedimiento/método será invariante (ver apéndice).
- Equivarianza: Imaginemos 2 investigadores resolviendo problemas de inferencia diferentes pero utilizando el mismo modelo estadístico $f(x;\theta)$. Si la evidencia de ambos coincide: $\underline{x}_1 = \underline{x}_2$, la inferencia que hacen ambos investigadores sobre θ debería ser la misma.
  - Si hacemos inferencia sobre θ utilizando la función de verosimilitud, entonces el procedimiento/método será equivariante.

Agenda

Verosimilitud y datos dependientes

- Sea $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{id}{\sim} f(x; \boldsymbol{\theta})$ (v.a. dependientes) con $\boldsymbol{\theta} \in \Theta$. Dada $\underline{X} = \underline{x}$, podemos definir la función de verosimilitud como:

$$L(\boldsymbol{\theta}|\underline{x}) = f(x_1; \boldsymbol{\theta})\Pi_{i=2}^{n} f(x_i; \boldsymbol{\theta}|x_1, \ldots, x_{i-1}).$$

  - La anterior expresión surge de la “regla del producto”.

- En el contexto de datos dependientes, solemos hacer algunos supuestos adicionales sobre la estructura probabilítistica subyacente. Por ejemplo, asumiendo que la secuencia $\{X_1, \ldots, X_n\}$ se observa de forma ordenada (por ejemplo en el tiempo) y que el modelo tiene estructura Markoviana, luego:

$$L(\boldsymbol{\theta}|\underline{x}) = f(x_1; \boldsymbol{\theta})\Pi_{i=2}^{n} f(x_i; \boldsymbol{\theta}|x_{i-1}).$$

- El caso no *id* es menos relevante en la práctica.

Verosimilitud e Invarianza de Escala

- Sea $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x; \boldsymbol{\theta})$ con $\boldsymbol{\theta} \in \Theta$. Dada $\underline{X} = \underline{x}$, consideremos la transformación biyectiva $Z = g(X)$[2].

- Como $f(z; \boldsymbol{\theta}) = f(x; \boldsymbol{\theta})|J|_x$, donde $J_x = dX/dZ$ es el inverso del Jacobiano de la transformación g (que NO depende de $\boldsymbol{\theta}$), luego:

$$L(\boldsymbol{\theta}|\underline{z}) = L(\boldsymbol{\theta}|\underline{x})c(\underline{x}).$$

  - Donde $c(\underline{x}) = \Pi_{i=1}^{n}|J_{x_i}| > 0$ es una constante que no depende de $\boldsymbol{\theta}$.

- Como la verosimilitud $L(\boldsymbol{\theta}|\underline{z})$ es proporcional a $L(\boldsymbol{\theta}|\underline{x})$, la inferencia que hacemos con $\underline{z} = g(\underline{x})$ en el contexto del modelo $f(z; \boldsymbol{\theta})$ es equivalente a la que hacemos con $\underline{x}$ en el contexto del modelo $f(x; \boldsymbol{\theta})$.

$$L(\boldsymbol{\theta}_1|\underline{x}) > L(\boldsymbol{\theta}_2|\underline{x}) \Leftrightarrow L(\boldsymbol{\theta}_1|\underline{z}) > L(\boldsymbol{\theta}_2|\underline{z}).$$

[2] Consideremos la muestra aleatoria: $\underline{Z} \equiv \{Z_1 = g(X_1), \ldots, Z_n = g(X_n)\}$.