Inferencia Estadística

Introducción

Gabriel Martos Venturini
gmartos@utdt.edu

Matías Pérez
lic.matiasdperez@gmail.com

UTDT

UNIVERSIDAD TORCUATO DI TELLA

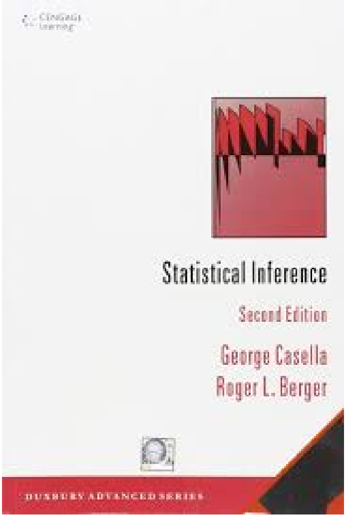
Objetivo y Programa

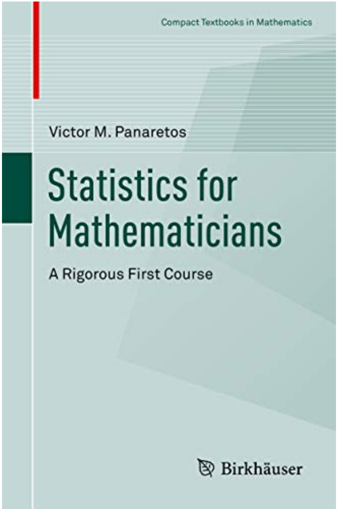
Objetivo: Formalizar conceptos y metodologías de inferencia clásica e introducir métodos modernos de inferencia Bayesiana y Noparamétrica.

- Muestreo y principios de reducción de datos.
- Estimación puntual.
- Nociones de riesgo.
- Propiedades asintóticas de los estimadores.
- Estimación por intervalos y test de hipótesis.
- Elementos de inferencia Bayesiana.
- Tópicos de inferencia no paramétrica.

@ Programa completo colgado en campus.

Bibliografía y evaluación





- Evaluación: Examen final (ver prototipo colgado en campus).

Agenda

Conceptos preliminares

- Inferir: *Deducir algo o sacarlo como conclusión de otra cosa* (RAE).
- Inferencia Estadística:
  - Formular juicios sobre cantidades desconocidas (*parámetros*) de una población con información parcial de la misma (muestra = datos).
    - Estimación puntual, por intervalos, test hipótesis, etc.
  - Cuantificar la incertidumbre en torno a dichos juicios.
    - Error standard de un estimador, amplitud del intervalo, etc.
- Para hacer inferencia estadística nos apoyamos en modelos estadísticos (en este curso, principalmente en modelos paramétricos).

"All models are wrong, but some are useful" (G. Box)

Definición (Modelo Estadístico)

Un modelo estadístico paramétrico $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$ es una *colección* de *distribuciones de probabilidad* definidas sobre un *espacio muestral* $\mathcal{X}$.

Example (El modelo exponencial)

Imaginemos que queremos estudiar la distribución del ingreso, v.a. denotada como X, en la población argentina. Para ello consideramos

- Modelo estadístico: $X \sim \text{Exp}(\theta)$, en otras palabras

$$\mathcal{F} : \{P_\theta(X \leq x) = \int_0^x \theta^{-1} e^{-\frac{t}{\theta}} \mathrm{d}t \,|\, x \geq 0, \theta > 0\}.$$

- Espacio muestral: Los reales no negativos ($\mathcal{X} := \{x \in \mathbb{R}^+ \cup 0\}$).
- Espacio de parámetros: Los reales positivos ($\Theta := \{\theta \in \mathbb{R}^+\}$).

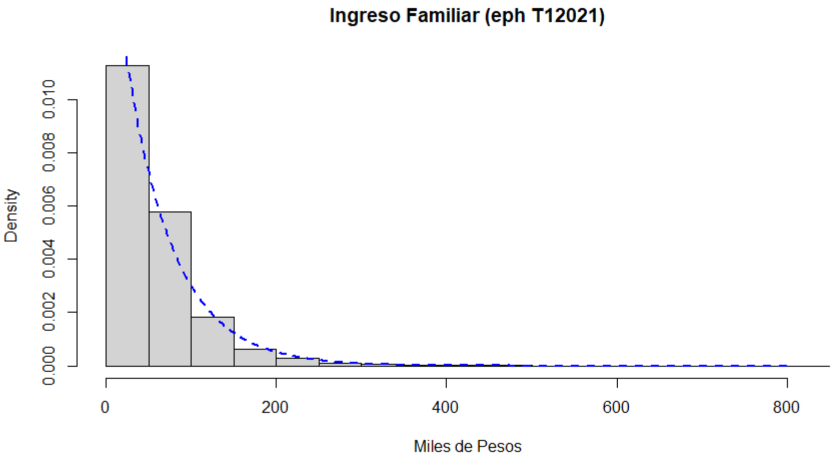


Figure: Ingresos familiares en la EPH (1T2021) en miles de pesos. En azul (- - -), la densidad de una distribución exponcial de parámetro $\lambda = 56.27$ (EMV).

Example (Regresión lineal)

Nos interesa modelar $E(Y|X_1, \ldots, X_p) \equiv \mu_{Y|X_1,\ldots,X_p}$ (la función de regresión). El modelo de regresión lineal propone:

$$Y = \theta_0 + \theta_1 X_1 + \cdots + \theta_p X_p + \varepsilon,$$

donde generalmente asumimos que $\varepsilon \sim N(0, \sigma^2)$.

- Modelo estadístico para $Y|X_1, \ldots, X_p$:

$$\mathcal{F} = \{N(\underbrace{\theta_0 + \theta_1 X_1 + \cdots + \theta_p X_p}_{\mu_{Y|X_1,\ldots,X_p}}, \sigma^2)\}$$

- Parámetros del modelo $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_p, \sigma^2) \in \mathbb{R}^{p+1} \times \mathbb{R}^+ = \Theta$.

- Objetivo: Disponemos de datos, que asumimos provienen del modelo especificado, y con ellos inferimos los parámetros desconocidos $\boldsymbol{\theta}$.

(S)Elección de modelo(s)

- A la hora de trabajar con datos en un problema concreto tendremos que elegir o definir un modelo estadístico para analizar los mismos.
- La (s)**elección** de un modelo estadístico *adecuado* con el que trabajar es una cuestión particularmente importante que resolvemos mediante:
  1. Criterios teóricos o científicos.
  2. Experimentación previa del fenómeno bajo estudio.
  3. Principios filosóficos (parsimonia).
  4. Kadane and Lazar: *Methods and Criteria for Model Selection* (JASA).

No abordaremos en profundidad estas cuestiones durante el curso.

Notación

- Parámetros con letras griegas: μ, σ^2, α, β...
- Variables aleatorias con mayúsculas: X, Y, Z...
- Realizaciones de las variables con minúsculas: x, y, z...
- Vector aleatorio (muestra aleatoria): $\underline{X} \equiv \{X_1, \ldots, X_n\}$.
- Realización de vector aleatorio (datos): $\underline{x} \equiv \{X_1 = x_1, \ldots, X_n = x_n\}$.
- Función de distribución $F(x;\theta)$ y de densidad $f(x;\theta)$:

$$F(x;\theta) \equiv P_\theta(X \leq x) = \int_{-\infty}^{x} f(t;\theta)\mathtt{d}t.$$

- Cuando sea necesario distinguiremos modelos continuos de discretos.
- Reservamos $\pi(\theta)$ y $\pi(\theta \,|\, \underline{x})$ para Inferencia Bayesiana.
- Las letras góticas $\mathcal{F}$, $\mathcal{C}$, $\mathcal{P}$ denotan en general conjuntos.

Agenda

Estadística Frecuentista vs Bayesiana

- Inferencia **clásica o frecuentista**:
  - θ es una cantidad fija y desconocida de la población.
  - Los métodos de inferencia tienen garantías de "*largo plazo*".
    - Ejemplo: Intervalos de Confianza.
- Inferencia **Bayesiana**:
  - Las hipótesis subjetivas (creencias) sobre θ se codifican en términos probabilíticos: Distribución a–priori (subjetiva) $\pi(\theta)$ para θ.
  - Con los datos (la verosimilitud) corregimos $\pi(\theta)$ de forma coherente.
  - Con la distribución a–posteri $\pi(\theta \,|\, \text{Datos})$ hacemos inferencia para θ.

Modelos Paramétricos y Noparamétricos

- Modelos **paramétricos**: Cada elemento de $\mathcal{F}$ se puede identificar mediante una cantidad finita de parámetros. En otras palabras, $\mathcal{F}$ esta indexado por $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m$ con $m \in \mathbb{N}$ y $m < \infty$. Escribimos:

$$\mathcal{F}_{\mathsf{p}} = \{f(\bullet; \boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta\}.$$

- Modelos **noparamétricos**: $\mathcal{F}$ contiene distribuciones que **no** se pueden representar mediante una cantidad finita de parámetros. Por ejemplo: $\mathcal{F}$ es el conjunto de todos los modelos de probabilidad cuyas funciones de densidad son $m \in \mathbb{N}$ veces diferenciables:

$$\mathcal{F}_{\mathsf{np}} = \{f(\bullet) \,|\, f \in C^m\}.$$

- Modelos **semi–paramétricos**:

$$\mathcal{F} = \{\alpha f_1(\bullet; \boldsymbol{\theta}) + (1-\alpha) f_2(\bullet) \,|\, f_1 \in \mathcal{F}_{\mathsf{p}},\ f_2 \in \mathcal{F}_{\mathsf{np}},\ \text{y } \alpha \in (0,1)\}.$$

- En la primera parte de este curso trabajaremos con modelos paramétricos (regulares e identificables), ASUMIENDO que:

$$f \in \mathcal{F} = \{f(\bullet; \boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m\},$$

  donde f es la "verdadera distribución" que genera datos en la pob.

- Existe un valor $\boldsymbol{\theta} \in \Theta$ (en general desconocido) que se corresponde con el verdadero valor del parámetro en la población: $f(x) \equiv f(x; \boldsymbol{\theta})$.

- En general no nos interesa $\boldsymbol{\theta}$ per–se, sino una función (o funcional) del mismo $\psi(\boldsymbol{\theta})$. A $\psi(\boldsymbol{\theta})$ lo llamaremos el parámetro de interés.
  - Ejemplo: La media, la varianza, la mediana, etc.

- Discutiremos métodos generales para estimar $\psi(\boldsymbol{\theta})$, testear hipótesis sobre $\psi(\boldsymbol{\theta})$ y construir regiones de confianza para $\psi(\boldsymbol{\theta})$.

- Para tener un marco de discusión amplio, será conveniente agrupar los modelos paramétricos en familias de distribuciones.

Agenda

Introducción

- Los modelos paramétricos que discutimos con los ejemplos–Normal, Exponencial, etc– comparten ciertas características y propiedades.
- Vamos a introducir un nivel de abstracción mayor agrupando dichos modelos en torno a una *familia* (conjunto amplio) de distribuciones.
- Marco de referencia más amplio para hacer inferencia.
  - Toda vez que podamos probar propiedades en una familia extensa de distribuciones, éstas serán válidas para todos los modelos de probabilidad que pertenecen dicha familia.
- Familias relevantes:
  1. Exponencial (ojo, no se trata de la distribución exponencial!).
  2. Localización y Escala.

Familia exponencial de 1 parámetro

- Consideremos $\mathcal{F} \equiv \{f(x;\theta) \,|\, \theta \in \Theta \subseteq \mathbb{R}\}$ donde:

$$f(x;\theta) = h(x)c(\theta)\exp\Big(w(\theta)t(x)\Big).$$

- $h(x) \geq 0$ y $t(x)$ **no** pueden depender de θ.
- $c(\theta) > 0$ y $w(\theta)$ **no** pueden depender de x.
  - Por lo tanto el *soporte* de $f(x;\theta)$ no depende de $\boldsymbol{\theta}$.

$$\text{Soporte}(f(x;\theta)) \equiv \{x \in \mathbb{R} \,|\, f(x;\theta) > 0\}$$

- Para **contrastar** que un modelo pertenece a la familia exponencial debes identificar las funciones h, c, w y t y verificar que se cumplen las condiciones arriba mencionada sobre cada una de estas funciones.
- Ejemplo: Modelo Bernoulli. Contraejemplo: Modelo Uniforme.

- Con $t(X)$ construimos un estadístico suficiente para θ.
  - $t(x)$ sintetiza toda la información relevante sobre θ.
- Nos van a interesar computar los momentos y la distribución (exacta o aproximada) del estadístico suficiente (para hacer inferencia).

$$E(t(X)) = \frac{d'(\theta)}{w'(\theta)} \text{ y } V(t(X)) = \frac{d''(\theta)w'(\theta) - d'(\theta)w''(\theta)}{[w'(\theta)]^3}$$

donde $d(\theta) = -\log(c(\theta))$.

  - Ejemplo: Modelo Bernoulli.
- La familia exponencial también es relevante en:
  - Inferencia con modelos lineales generalizados.
  - Modelos conjugados e inferencia Bayesiana.

Familia exponencial de k parámetros

- Consideremos $\mathcal{F} \equiv \{f(x;\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k\}$ donde:

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

- $h(x) \geq 0$ y $t_1(x), \ldots, t_k(x)$ **no** pueden depender de $\boldsymbol{\theta}$.
- $c(\boldsymbol{\theta}) > 0$ y $w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ **no** pueden depender de x.
- El soporte de $f(x;\boldsymbol{\theta})$ no depende de $\boldsymbol{\theta}$.
- Ejemplo: Modelos Normal, Beta, Gamma.

Si w y t son funciones vectoriales entonces: $w_i(\boldsymbol{\theta})t_i(x) \equiv \langle w_i(\boldsymbol{\theta}), t_i(x)\rangle$.

Reparametrización (BackUp)

- La densidad $f(x;\boldsymbol{\theta})$ se suele reparametrizar como:

$$f(x;\phi) = h(x)c(\phi)\exp\Big(\sum_{i=1}^{k}\phi_i t_i(x)\Big),$$

donde $\phi_i = w_i(\boldsymbol{\theta})$ para $i = 1,\ldots,k$ se llaman *parámetros naturales*.

- Notar que $h(x)$ y $t_1(x),\ldots,t_k(x)$ son las mismas funciones.
- El conjunto Φ definido como

$$\Phi : \Big\{(\phi_1,\ldots,\phi_k) \,|\, \int h(x)\exp\Big(\sum_{i=1}^{k}\phi_i t_i(x)\Big)dx < \infty\Big\}$$

es llamado *espacio paramétrico natural* y se cumple que

$$c(\phi) = \Big[\int h(x)\exp\big(\sum_{i=1}^{k}\phi_i t_i(x)\big)\mathrm{d}x\Big]^{-1}.$$

Familias de Localización y Escala

El objetivo es construir familias de modelos a partir de una densidad (o pmf en el caso discreto) que difieren solo en su *localización* y/o su *escala*.

- Si $f(x)$ es una densidad (o pmf en el caso discreto), entonces definimos la familia de **localización y escala** relativa a $f(x)$ como:

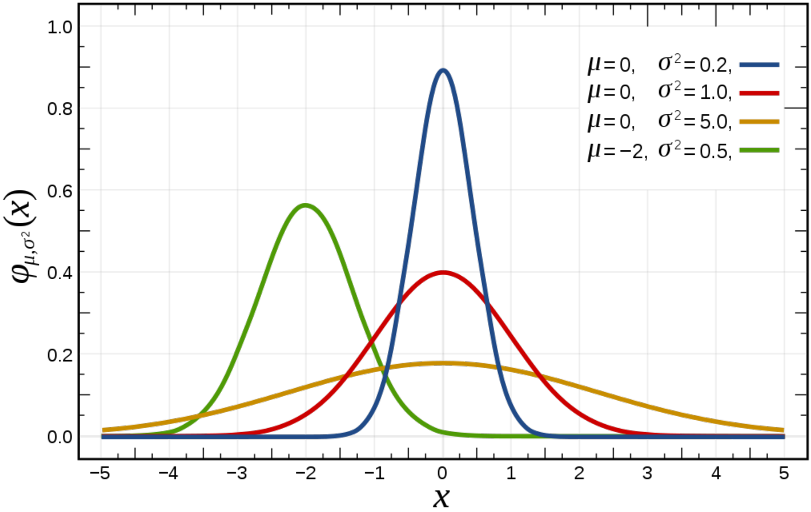
$$\mathcal{F} \equiv \{f(x;\mu,\sigma) = \frac{1}{\sigma} f\big(\frac{x-\mu}{\sigma}\big) \,|\, \mu \in \mathbb{R} \text{ y } \sigma > 0\}.$$

- **Location–family**: Si f es una pdf y μ una constante, entonces $f(x-\mu)$ es una familia de localización con parámetro μ.

- **Scale–family**: Si f es una pdf y $\sigma > 0$ una constante, entonces $\sigma^{-1} f(x/\sigma)$ es una familia de escala con parámetro σ.

- Ejemplo: Si $f(x) = e^{-x^2/2}/\sqrt{2\pi}$, la familia de localización y escala relativa a esta densidad es el conjunto $\{N(\mu,\sigma^2) \,|\, \mu \in \mathbb{R} \text{ y } \sigma^2 > 0\}$.

Familias de Localización y Escala



Propiedades

- Si $Z \sim f(z)$ y $X \sim \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ (es decir que X sigue un modelo de localización y escala a partir de $f(z)$), luego se cumple que:

$$E(X) = \sigma E(Z) + \mu \text{ y } V(X) = \sigma^2 V(Z).$$

  - Notar que $E(X) = \mu$ sólo cuando $E(Z) = 0$ y que $V(X) = \sigma^2$ sólo cuando $V(Z) = 1$ (esto ocurre en el ejemplo dado en § 21).

- Los cuantiles de X y Z también están relacionados de manera lineal:
  - $x_p = z_p \sigma + \mu$ para todo $p \in (0, 1)$, donde

$$x_p = \inf\{x \in \mathbb{R} : F_X(x) \geq p\} \text{ y } z_p = \inf\{z \in \mathbb{R} : F_Z(z) \geq p\}.$$

- Ejemplo: $Z \sim f(z) = 2z$ con $z \in [0, 1]$ (más ejemplos en CB § 3.4).

Agenda

Agenda

Muestra aleatoria

- La colección de variables aleatorias (va) $\underline{X} \equiv \{X_1, \ldots, X_n\}$ se dice una **muestra aleatoria** de tamaño n de una población donde X tiene una densidad $f(x;\theta)$ si se cumplen las dos condiciones siguientes:

  1- El conjunto de va $\{X_1, \ldots, X_n\}$ son simultáneamente independientes[1].

  2- El conjunto de va $\{X_1, \ldots, X_n\}$ sigue la misma distribución que X.

  - Escribimos indistintamente: $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ o $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f$.

- Bajo el supuesto iid, la pdf de la muestra aleatoria $\{X_1, \ldots, X_n\}$ es:

$$f(x_1, \ldots, x_n; \theta) = \Pi_{i=1}^{n} f(x_i; \theta).$$

  - Responder preguntas probabilísticas respecto de la muestra aleatoria.

- Ejemplo: Modelo Exponencial.

- Poblaciones finitas y pequeñas: Muestreo con reposición.

[1] Si $X \sim f_X(x) \perp\!\!\!\perp Y \sim f_Y(y)$ entonces $f_{X|Y=y}(x) = f_X(x)$ y $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

Estadísticos

- **Definición**: Un *estadístico* $T(X_1, \ldots, X_n)$ es una función (posiblemente vectorial) de los elementos de una muestra aleatoria.
- Por lo tanto un estadístico es una variable (vector) aleatoria(o).
- La definición de estadístico es muy amplia, sólo le pedimos a T que **no dependa** de forma explícita de θ, el parámetro del modelo.
- Estadístico y Estimador (terminología).
- La distribución del estadístico T se denomina distribución *muestral*, puesto que T se origina a partir de una muestra aleatoria.
- Veamos algunos ejemplos de estadísticos (estimadores) importantes.

Dos estadísticos importantes

Para $\{X_1, \ldots, X_n\} \stackrel{iid}{\sim} X$ definimos:

- Media muestral

$$\overline{X}_n = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{T(X_1,\ldots,X_n)} \; .$$

- La (cuasi)varianza muestral

$$S_n^2 = \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}_{T(X_1,\ldots,X_n)} .$$

- Denotaremos con letras pequeñas $(\overline{x}_n, s_n^2)$ a las realizaciones de estas variables aleatorias (dada una muestra en concreto de la población).

Algunas propiedades importantes de $\overline{X}_n$ y S_n^2

No importa como se distribuya X en la población se tiene que:

1. $E(\overline{X}_n) = E(X)$.
2. $Var(\overline{X}_n) = Var(X)/n$.
3. $E(S_n^2) = Var(X)$.

Por la LGN se verifica que (refresh en apéndice):

- $\overline{X}_n \to_P E(X)$.
- $S_n^2 \to_P Var(X)$ (porque $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_P E(X^2)$ y $\overline{X}_n^2 \to_P E(X)^2$).
  - Con n grande, es altamente probable que $\overline{X}_n$ y S_n^2 tomen valores cercanos a $E(X)$ y $Var(X)$ (cantidades en general desconocidas).

Si conocemos (o asumimos) la distribución de X en la población, podremos caracterizar mejor la distribución muestral de ambos estadísticos...

Muestreo aleatorio en poblaciones normales

- Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} N(\mu, \sigma^2)$, es decir que:

$$P(X \leq x) = \int_{-\infty}^{x} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}}_{f(X=t;\boldsymbol{\theta}=(\mu,\sigma^2))} \mathrm{d}t.$$

- Entonces las variables aleatorias $\overline{X}_n$ y S_n^2 verifican que:
  1. $\overline{X}_n$ y S_n^2 son variables aleatorias independientes (T. Basu).
  2. $\overline{X}_n \sim N(\mu, \sigma^2/n)$.
  3. $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}$.

- Momentos de las variables aleatorias $\overline{X}_n$ y S_n^2:
  1. $E(\overline{X}_n) = \mu$ y $E(S_n^2) = \sigma^2$.
  2. $V(\overline{X}_n) = \sigma^2/n$ y $V(S_n^2) = 2\sigma^4/(n-1)$.

Distribución en el muestreo en familias exponenciales

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

$$f(x;\theta) = h(x)c(\theta)\exp\Big(w(\theta)t(x)\Big).$$

- El estadístico (suficiente y completo para θ):

$$T(X_1, \ldots, X_n) = \sum_{i=1}^{n} t(X_i).$$

  bajo condiciones generales[2], sigue una distribución que también pertenece a la familia exponencial:

$$T \sim f_T(u;\theta) = H(u)c(\theta)^n \exp\Big(w(\theta)u\Big).$$

- No necesitas integrar para computar los momentos de T (ver § 18).

[2] CB pp 212, Th 5.2.5.

Agenda

Definiciones relevantes

- Los estadísticos de orden relativos a $\underline{X} = \{X_1, \ldots, X_n\} \overset{iid}{\sim} F$ son simplemente los valores ordenados de dichas variables aleatorias que denotamos como $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.
- Algunos ejemplos de estadísticos de orden:
  - Mínimo: $X_{(1)} = \min\{X_1, \ldots, X_n\}$.
  - Máximo: $X_{(n)} = \max\{X_1, \ldots, X_n\}$.
  - Mediana: $M = X_{(n+1)/2}$ si n es impar y $M = \frac{X_{(n/2)}+X_{(n/2+1)}}{2}$ si n es par.
  - Rango: $R = X_{(n)} - X_{(1)}$.
- Lo que resulta realmente interesante, es que la distribución de estos estadísticos se puede escribir de manera bastante general en términos de $f(x;\theta)$ y $F(x;\theta)$ (para demostraciones formales ver CB § 5.5).

Distribuciones marginales

- Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ siendo X una variable aleatoria continua (discreta) con densidad (p.m.f.) $f(x;\theta)$ y distribución $F(x;\theta)$.
- La función de densidad de $X_{(j)}$ (para $j = 1, \ldots, n$) se escribe como:

$$f_{X_{(j)}}(x;\theta) = \frac{n!}{(j-1)!(n-j)!} f(x;\theta)[F(x;\theta)]^{j-1}[1-F(x;\theta)]^{n-j}.$$

- Notar que para el máximo y el mínimo, $X_{(1)}$ y $X_{(n)}$ respectivamente, la funciones de densidad de estos estadísticos se reducen a:

$$f_{X_{(1)}}(x;\theta) = nf(x;\theta)[1-F(x;\theta)]^{n-1}, \text{y } f_{X_{(n)}}(x;\theta) = nf(x;\theta)[F(x;\theta)]^{n-1}$$

- Ejemplo I: Si $X \sim \text{Exp}(\lambda)$ entonces $X_{(1)} \sim \text{Exp}(\lambda/n)$.
- Ejemplo II: Modelo uniforme (próxima slide).

- Si $X \sim \text{Unif}(0,1)$, luego $f(x) = 1$ y $F(x) = x$, para $x \in (0,1)$.
- Luego reemplazando en

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f(x)[F(x)]^{j-1}[1 - F(x)]^{n-j},$$

- Distribución del mínimo:

$$f_{X_{(1)}}(x) = n(1-x)^{n-1}, \text{ para } x \in (0,1).$$

- Distribución del máximo:

$$f_{X_{(n)}}(x) = nx^{n-1}, \text{ para } x \in (0,1).$$

- Se puede demostrar que para $X \sim \text{Unif}(0,1)$:

$$f_{X_{(j)}}(x) = \text{Beta}(j, n-j+1), \text{ para } x \in (0,1).$$

Distribución conjuntas

- Llamemos $f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n;\theta)$ a la densidad de del vector aleatorio ordenado $(X_{(1)},\ldots,X_{(n)})$, se puede demostrar que:

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n;\theta) = \begin{cases} n!\Pi_{i=1}^{n} f(x_i;\theta), & \text{si } -\infty < x_1 < \cdots < x_n < \infty, \\ 0 \text{ en otro caso.} \end{cases}$$

- A partir de esta expresión podemos obtener las distribuciones condicionales de los estadísticos de orden. Ej. con $1 \leq i \leq j \leq n$:

$$f_{X_{(i)},X_{(j)}}(x_i,x_j;\theta) = n!\frac{f(x_i;\theta)f(x_j;\theta)[F(x_i;\theta)]^{i-1}[F(x_j;\theta)-F(x_i;\theta)]^{j-1-i}[1-F(x_j;\theta)]^{n-j}}{(i-1)!(j-1-i)!(n-j)!}$$

- Para una discusión completa respecto de la distribución de los estadísticos de orden consultar CB § 5.5.

Agenda

Motivación

- En las secciones anteriores discutimos algunos estadísticos muestrales habituales (ej: media y varianza) y su distribución en el muestreo.
- Cuando nos interesan ciertas *funciones de los estadísticos*–una nueva variable aleatoria $g(T)$–, cuya distribución no conocemos, entonces podemos utilizar el método delta para aproximar dicha distribución.
- No vale para cualquier estadístico y/o cualquier transformación g.
  - T (estandarizado) tiene que converger en distribución a una normal.
  - g tiene que ser diferenciable respecto de θ.

Example (Odds ratio)

$\{X_1, \ldots, X_n\} \overset{iid}{\sim}$ Bernulli(θ): Si el parámetro de interés es $g(\theta) = \theta/(1-\theta)$; el estadístico con el que vamos a inferir $g(\theta)$ será $g(\overline{X}) = \overline{X}/(1-\overline{X})$.

Método delta

- Piensa en X_n como un estadístico que depende de n (media muestral).

Theorem (First order Delta method)

Sea $\{X_1, \ldots, X_n\}$ *una secuencia de v.a. donde* $\sqrt{n}(X_n - \theta) \to_F N(0, \sigma^2)$. *Para una función* g *diferenciable en* θ *y con* $g'(\theta) \neq 0$, *se tiene que:*

$$\sqrt{n}(g(X_n) - g(\theta)) \to_F N(0, \sigma^2[g'(\theta)]^2).$$

Theorem (Second order Delta method)

Sea $\{X_1, \ldots, X_n\}$ *una secuencia de v.a. donde* $\sqrt{n}(X_n - \theta) \to_F N(0, \sigma^2)$. *Para* g *2 veces diferenciable en* θ *con* $g'(\theta) = 0$ *y* $g''(\theta) \neq 0$, *se cumple:*

$$\sqrt{n}(g(X_n) - g(\theta)) \to_F \sigma^2 \frac{g''(\theta)}{2} \chi_1^2.$$

- Ejemplo del odds ratio (§ 38).

Agenda

Refresh I: Convergencia en probabilidad

Definición (Convergencia en probabilidad)

Una sucesión de variables aleatorias $\{X_1, \ldots, X_n\}$ converge en probabilidad a la variable aleatoria X si, para todo $\varepsilon > 0$, se verifica que:

$$\lim_{n\to\infty} P\big(|X_n - X| \geq \varepsilon\big) = 0, \text{ o equivalentemente } \lim_{n\to\infty} P\big(|X_n - X| < \varepsilon\big) = 1.$$

- Típicamente en inferencia asumimos condiciones de tipo *iid* y nos interesan las convergencias a constantes. Ej: Ley grandes números.

Theorem (Ley débil de los grandes números)

Si $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$, *y* $E(X^2) < \infty$, *entonces:*

$$\lim_{n\to\infty} P\big(|\overline{X}_n - E(X)| \geq \varepsilon\big) = 0,$$

donde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Denotamos esta convergencia con* $\overline{X}_n \to_P E(X)$.

Álgebra de convergencia en probabilidad

Sea a_n es una sucesión de números reales tal que $a_n \to a$. Sean X_n y Y_n dos variables aleatorias tales que $X_n \to_P X$ y $Y_n \to_p Y$. Entonces:

1. $a_n + X_n \to_P a + X$.
2. $a_n X_n \to_P aX$.
3. $X_n + Y_n \to_P X + Y$.
4. $X_n Y_n \to_P XY$.
5. Si $P(Y = 0) = 0$, $X_n / Y_n \to_P X/Y$.
6. Si $g : \mathbb{R} \to \mathbb{R}$ es continua luego $g(X_n) \to_P g(X)$.
7. Si $g : \mathbb{R}^2 \to \mathbb{R}$ es continua luego $g(X_n, Y_n) \to_P g(X, Y)$.

Definición (Convergencia multivariante en probabilidad)

La sucesión de vectores aleatorios $\{X_1, \ldots, X_n\}$ d–dimensionales converge en probabilidad al vector aleatorio X si, para todo $\varepsilon > 0$:

$$\lim_{n\to\infty} P\big(\|X_n - X\|_\infty \geq \varepsilon\big) = 0,$$

donde $\|(y_1, \ldots, y_d)\|_\infty \equiv \max_{i=1,\ldots,d}\{y_1, \ldots, y_d\}$.

Utilidad de la convergencia en proba

- Dada $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$, recordemos que:

$$Var(X) = E(X^2) - (E(X))^2 \,.$$

- Sabemos que $\overline{X}_n$ es un estimador consistente de $E(X)$ porque (LGN) $\overline{X}_n \to_P E(X)$. Por lo tanto, utilizando la propiedad (6) con $f(x) = x^2$:

$$f(\overline{X}_n) = (\overline{X}_n)^2 \to_P f(E(X)) = E(X)^2 \,.$$

- Por la ley de los grandes números sabemos además que:

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_P E(X^2) = Var(X) + E(X)^2 \,.$$

- De los dos resultados anteriores más la propiedad (3):

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \to_P Var(X).$$

Refresh II: Convergencia en distribución

Definición (Convergencia en distribución)

Una sucesión de variables aleatorias $\{X_1, \ldots, X_n\}$ converge en distribución a la variable aleatoria X si se cumple que:

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

en todo punto de continuidad de F_X. Escribimos $X_n \to_F X$.

- Cuando $n \gg 0$ la CDF de X_n esta cerca de la CDF de X.
  - Esto no implica que X_n este cerca de X.
- Relaciones entre convergencias:
  - Si $X_n \to_p X \Rightarrow X_n \to_F X$ (y viceversa si $X = c$).
  - Si $X_n \to_F X$ y $Y_n \to_p c \Rightarrow$ (a) $Y_n X_n \to_F cX$ y (b) $X_n + Y_n \to_F X + c$.

Convergencia en distribución (caso multivariante)

Lo presentamos para el caso en que $d = 2$, la extensión al más caso general es directa a partir de la definición que sigue.

Definición (Convergencia en distribución multivariante)

Una sucesión de variables aleatorias bi-variantes $\{X_1, \ldots, X_n\}$ converge en distribución a la variable aleatoria bivariante X con función de distribución $F_X(x_1, x_2)$ si se cumple que:

$$\lim_{n \to \infty} F_{X_n}(x_1, x_2) = F_X(x_1, x_2)$$

en todo punto de continuidad (x_1, x_2) de F_X. Escribimos $X_n \to_F X$.

Refresh III: Teorema del Límite Central

Definición (CLT (strong))

Si $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ con $0 < V(X) < \infty$; llamemos G_n a la CDF de la variable aleatoria $Z_n \equiv \sqrt{n}(\overline{X}_n - \mu)/\sigma$, luego:

$$\lim_{n\to\infty} G_n(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

Es decir que $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ converge en distribución a una $N(0, 1)$.

- Supuestos: *iid* y varianza finita.
- No importa como se distribuya X en la población.
- El resultado es válido en términos asintóticos, pero no nos dice cuan buena es esta "aproximación" para un n en general.

Polinomio de Taylor

Si g una función que admite r derivadas ($g \in C^r$). Para una constante x_0, el polinomio de Taylor *de orden* r en torno a el punto x_0 se define como:

$$P_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(x_0)}{i!}(x - x_0)^i + R_r(x),$$

- $P_r(x)$ nos permite aproximar g en un entorno de x_0.
- $R_r(x) = P_r(x) - g(x)$ tiende a cero a una tasa más rápida que r.
- Nos interesan los momentos (y la distribución) aproximados de $g(T)$.
- La aproximación la hacemos en torno a θ (parámetro de interés).

El método delta (demostración simplificada)

- Hacemos una aproximación de primer orden:

$$g(T) \approx g(\theta) + g'(\theta)(T - \theta)$$

- Si $\sqrt{n}(T - \theta) \rightarrow_F N(0, \sigma^2)$, luego se tiene que:

$$\sqrt{n}(g(T) - g(\theta)) = g'(\theta)\sqrt{n}(T - \theta) + \underbrace{O_P(1)}_{\text{Resto}} \rightarrow_F N(0, \sigma^2[g'(\theta)]^2)$$