# Análisis Estadístico

## Estimación

San Andrés

May 26, 2022

# Estimación

## Problema de estimación

Dada una muestra aleatoria $X_1, \ldots, X_n$ queremos estimar un parámetro $\theta$ asociado a la distribución de la muestra.

## Definition (Estimador)

Dadas variables aleatorias $X_1, \ldots, X_n$, para un parámetro cualquiera $\theta$, un **estimador** es una función de $X_1, \ldots, X_n$.

## Notación

Llamaremos $\widehat{\theta}_n$ a un estimador del parámetro $\theta$ basado en una muestra de tamanõ $n$.

# Estimación

## Estimación de la varianza

Sea $X_1, \dots X_n$ una muestra aleatoria.
¿Cómo podemos estimar $\sigma^2$ la varianza de $X$?.

$$\sigma^2 = E(X^2) - E^2(X)$$

**Idea:** Estimamos $E(X^2)$ y $E^2(X)$ y restamos los estimadores

# Estimación

## Estimación de la varianza

Sea $X_1, \ldots X_n$ una muestra aleatoria.
¿Cómo podemos estimar $\sigma^2$ la varianza de $X$?.

$$\sigma^2 = E(X^2) - E^2(X)$$

**Idea:** Estimamos $E(X^2)$ y $E^2(X)$ y restamos los estimadores

# Estimación

La clase pasada vimos que un estimador de la esperanza de una variable aleatoria es

$$\widehat{E(X)} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

¿Tiene sentido estimar $E^2(X)$ con $\overline{X}_n^2$?

¿Cómo estimamos $E(X^2)$?

$$\widehat{E(X^2)} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

# Estimación

## Estimador de la varianza

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2$$

¿Es el único estimador posible?

## Example

Sea $X_1, \ldots X_n$ una muestra aleatoria. Proponer otro estimador de la varianza en el caso de que

- $X \sim Ber(p)$
- $X \sim Poi(\lambda)$
- $X \sim Exp(\lambda)$

# Estimación

## Objetivo

Para distintos estimadores nos interesa responder:

- ¿Qué propiedades tiene la **distribución muestral** del estimador?
- ¿Cómo medimos si un estimador es **mejor** que otro? ¿Cúal es la manera **óptima** de estimar un parámetro?
- ¿Podemos garantizar que, con probabilidad alta, el estimador va a estar **cerca** del valor verdadero (fijo y desconocido) que queremos estimar?
- ¿Cuántas (y qué tipo de) observaciones necesitamos para garantizar cierto margen de error con probabilidad alta?

# Medir la calidad de un estimador

## Estimar el parámetro de una Bernoulli

Dada una muestra aleatoria $X_1, \ldots, X_n \sim Ber(p)$, llamemos

$$T = \sum_{i=1}^{n} X_i$$

Proponemos tres estimadores de $p$:

- La media muestral:
$$\widehat{p}_1 = T/n = \overline{X}_n\,.$$
- La media muestral agregando 2 fracasos y 2 éxitos “artificiales” a la muestra:
$$\widehat{p}_2 = \frac{T+2}{n+4}\,.$$
- La media muestral de solo dos observaciones:
$$\widehat{p}_3 = \frac{X_1 + X_n}{2}\,.$$

# Medir la calidad de un estimador

**Estimar el parámetro de una Bernoulli**

Intuitivamente podemos decartar al estimador $p_3$ pero...

¿cómo determinamos si es mejor $\widehat{p}_1$ o $\widehat{p}_2$?

# Error cuadrático medio

Para un estimador cualquiera $\widehat{\theta}_n$ de $\theta$, su **error de estimación**

$$\widehat{\theta}_n - \theta$$

es una variable aleatoria. Si queremos medir únicamente el **tamaño** del error de estimación podemos considerar

$$|\widehat{\theta}_n - \theta| \quad o \quad (\widehat{\theta}_n - \theta)^2,$$

que también son variables aleatorias. **Idealmente**, queremos que estas variables aleatorias sean **chicas**, que tengan su **centro de masa** en cero.

**Definition (Error cuadrático medio)**

Dado $\widehat{\theta}_n$, un estimador de $\theta$, definimos su **error cuadrático medio (ECM)** como

$$E\left[\left(\widehat{\theta}_n - \theta\right)^2\right]$$

# Error cuadrático medio

**Observación**

¿Por qué $E\left[\left(\widehat{\theta}_n - \theta\right)^2\right]$ y no $E\left(\left|\widehat{\theta}_n - \theta\right|\right)$?

Por conveniencia matemática.

# Error cuadrático medio

## Theorem (Descomposición del ECM)

$$E\left[\left(\widehat{\theta}_n-\theta\right)^2\right]=\underbrace{Var(\widehat{\theta}_n)}_{varianza}+\underbrace{\left(E(\widehat{\theta}_n)-\theta\right)^2}_{sesgo\ al\ cuadrado}.$$

## Proof.

Si llamamos $e=\widehat{\theta}_n-\theta$. Sabemos que

$$Var(e)=E\left(e^2\right)-\left[E(e)\right]^2.$$

Luego

$$E\left(e^2\right)=Var(e)+\left[E(e)\right]^2.$$

Además, como $\theta$ es constante, $Var(e)=Var(\widehat{\theta}_n)$. □

# Error cuadrático medio

## Estimar el parámetro de una Bernoulli

Volviendo al ejemplo, queremos comparar $\widehat{p}_1$ y $\widehat{p}_2$:

$\widehat{p}_1$ la media muestral:

$$\widehat{p}_1 = \overline{X}_n\,.$$

$\widehat{p}_2$ agregar 2 fracasos y 2 éxitos artificiales a la muestra y despues calcular la media muestral:

$$\widehat{p}_2 = \frac{T+2}{n+4}.$$

$\widehat{p}_3$ Promediar la primera y la ultima observación

$$\widehat{p}_3 = \frac{X_1 + X_n}{2}.$$

Calculemos el ECM de cada uno.

# Error cuadrático medio

## Estimar el parámetro de una Bernoulli

Sabemos que

- $$E\left(\widehat{p}_1\right) = p$$

- $$Var(\widehat{p}_1) = \frac{p(1-p)}{n}$$

Entonces el ECM de $\widehat{p}_1$ es

$$E\left[\left(\widehat{p}_1 - p\right)^2\right] = \frac{p(1-p)}{n}.$$

# Error cuadrático medio

## Estimar el parámetro de una Bernoulli

Calculemos ahora el sesgo y la varianza de $\widehat{p}_3$. Notemos que que

$$E\left(\frac{X_1+X_n}{2}\right)=\frac{1}{2}\left(E(X_1)+E(X_n)\right)=p$$

$$Var\left(\frac{X_1+X_n}{2}\right)=\frac{1}{2^2}\left(Var(X_1)+Var(X_n)\right)=\frac{1}{2}p(1-p).$$

Entonces

$$E\left((\widehat{p}_3-p)^2\right)=\frac{1}{2}p(1-p).$$

# Error cuadrático medio

## Estimar el parámetro de una Bernoulli

Calculemos ahora el sesgo y la varianza de $\widehat{p}_2$. Notemos que que $T \sim Bin(n, p)$.

$$E\left(\frac{T+2}{n+4}\right) - p = \frac{E(T+2)}{n+4} - p = \frac{np+2}{n+4} - p = \frac{2/n - 4p/n}{1+4/n}.$$

El sesgo no es cero, salvo cuando $p = 0.5$. Sin embargo el sesgo converge a cero cuando $n \to \infty$.

$$Var\left(\frac{T+2}{n+4}\right) = \frac{Var(T+2)}{(n+4)^2} = \frac{Var(T)}{(n+4)^2} = \frac{np(1-p)}{(n+4)^2}.$$

Entonces

$$E\left((\widehat{p}_2 - p)^2\right) = \left(\frac{2/n - 4p/n}{1+4/n}\right)^2 + \frac{np(1-p)}{(n+4)^2}.$$

# Error cuadrático medio

## Estimar el parámetro de una Bernoulli

Tenemos

$$ECM(\widehat{p}_1) = E\left((\widehat{p}_1 - p)^2\right) = \frac{p(1-p)}{n}.$$

$$ECM(\widehat{p}_3) = E\left((\widehat{p}_3 - p)^2\right) = \frac{1}{2}p(1-p).$$

$$ECM(\widehat{p}_2) = E\left((\widehat{p}_2 - p)^2\right) = \left(\frac{2/n - 4p/n}{1 + 4/n}\right)^2 + \frac{np(1-p)}{(n+4)^2}.$$

¿Cuál es mejor?

Si uno de los dos ECM es más bajo para todos los valores de $n$ y de $p$, ese sería el estimador deseable.

En la siguiente figura veremos que eso no pasa.

# Error cuadrático medio

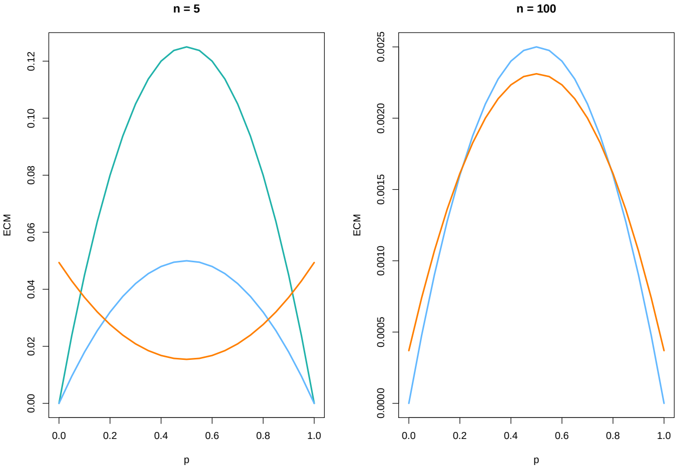



Figure: ECM para $\widehat{p}_1$ (celeste), $\widehat{p}_2$ (naranja) y $\widehat{p}_3$ (verde).

# Error cuadrático medio

```r
ecm_p13 = function(n, p){ p*(1-p)/n }
ecm_p2 = function(n, p){
     ((2/n-4*p/n)/(1+4/n))^2+n*p*(1-p)/((n+4)^2)
  }

p = seq(0,1, by = 0.05) #grilla

#grafico 1
plot(p, ecm_p13(2,p), col = "lightseagreen",
             type = "l", lwd = 2.5,
                ylab = "ECM", main = "n = 5")
lines(p, ecm_p13(5,p), col = "steelblue1", lwd = 2.5)
lines(p, ecm_p2(5,p), col = "darkorange1", lwd = 2.5)
```

# Error cuadrático medio

## Máximo de uniformes

Supongamos que el tiempo de reacción de una persona intoxicada a cierto estímulo es una variable aleatoria $X$, que tiene distribución uniforme en el intervalo $(0, \theta)$ para un valor de $\theta$ desconocido.

Tenemos una muestra aleatoria $X_1, \ldots, X_n$ y queremos estimar $\theta$. Un estimador posible para $\theta$ es

$$\widehat{\theta}_n = \max\{X_1, \ldots, X_n\}.$$

Por ejemplo, si $n = 3$ y la realización de la muestra es

$$x_1 = 4.2 \quad ; \quad x_2 = 1.7 \quad ; \quad x_3 = 2.4$$

el valor estimado será 4.2.

Calcular el sesgo del estimador y su error cuadrático medio.

# Propiedades de la convergencia en probabilidad

## Máximo de uniformes

Calculemos primero la función de distribución acumulada de $\widehat{\theta}_n$

$$\begin{aligned} P\left(\widehat{\theta}_n \leq x\right) &= P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \\ &= P(X_1 \leq x) \ldots P(X_n \leq x) = P(X \leq x)^n. \end{aligned}$$

Ahora

$$P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{\theta} & 0 \leq x \leq \theta \\ 1 & x \geq \theta \end{cases}$$

Luego

$$P\left(\widehat{\theta}_n \leq x\right) = \begin{cases} 0 & x \leq 0 \\ \left(\frac{x}{\theta}\right)^n & 0 \leq x \leq \theta \\ 1 & x \geq \theta. \end{cases}$$

# Error cuadrático medio

## Máximo de uniformes

Derivando, sacamos que la densidad de $\widehat{\theta}_n$ es

$$f_{\widehat{\theta}_n}(x) = nx^{n-1}\frac{1}{\theta^n} I_{(0,\theta)}(x)\,.$$

Entonces

$$E(\widehat{\theta}_n) = \int_0^{\theta} nx^n \frac{1}{\theta^n} dx = \frac{n}{n+1}\theta,$$

$$E(\widehat{\theta}_n^2) = \int_0^{\theta} nx^{n+1} \frac{1}{\theta^n} dx = \frac{n}{n+2}\theta^2$$

y

$$Var(\widehat{\theta}_n) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\,.$$

# Error cuadrático medio

## Máximo de uniformes

Observemos que el estimador es sesgado y que

$$\text{Sesgo}(\widehat{\theta}_n) = -\frac{\theta}{n+1} \underset{n\to\infty}{\longrightarrow} 0$$

y

$$Var(\widehat{\theta}_n) \underset{n\to\infty}{\longrightarrow} 0,$$

por lo que

$$ECM(\widehat{\theta}_n) = \underbrace{\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2}_{\text{Varianza}} + \left(\underbrace{-\frac{\theta}{n+1}}_{\text{Sesgo}}\right)^2 \underset{n\to\infty}{\longrightarrow} 0,$$

# Error cuadrático medio

- Buscar un estimador que tenga un ECM que es menor que el de cualquier otro estimador, para cualquier valor posible del parámetro a estimar, es demasiado ambicioso.
- Una manera de solucionar este dilema es restringir la clase de estimadores competidores. Una restricción usual es la de quedarse únicamente con los estimadores insesgados.

# Sesgo

**Definition (Estimador insesgado)**

Un estimador $\widehat{\theta}_n$ de $\theta$ se dice **insesgado** si

$$E(\widehat{\theta}_n) = \theta$$

para todos los valores posibles de $\theta$. En general, la diferencia

$$E(\widehat{\theta}_n) - \theta$$

se llama el **sesgo** del estimador.

**Idea**

Un estimador es insesgado si, cualquiera sea el valor de $\theta$, la distribución muestral del estimador está centrada en $\theta$.

# Sesgo

**Ejemplo**

Si $X_1, \ldots, X_n$ es una muestra aleatoria con

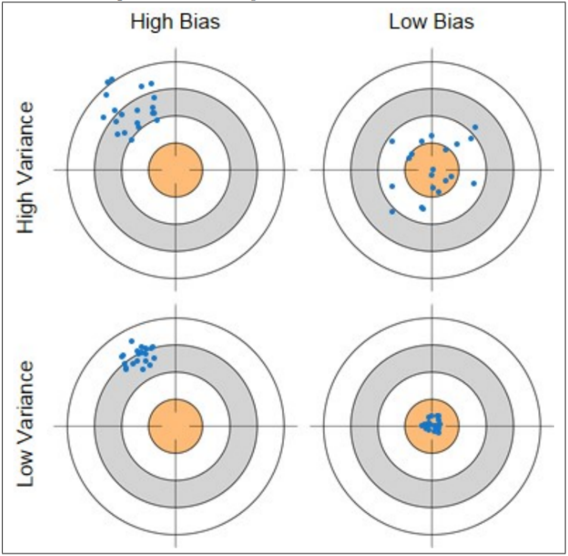
$$\mu = E(X_1) = \cdots = E(X_n)$$

entonces $\overline{X}_n$ es un estimador insesgado de $\mu$.

**Ejemplo**

En el ejemplo que hicimos antes, $\widehat{p}_2$ resultó ser un estimador sesgado de $\mu$, mientras que $\widehat{p}_1$ y $\widehat{p}_3$ son estimadores insesgados.

# Sesgo



# Sesgo

- Que un estimador sea insesgado no quiere decir que necesariamente este cerca del parámetro verdadero. El ECM en cambio sí mide en cierto sentido que un estimador esté cerca o lejos del parámetro que queremos estimar.
- ¡El sesgo no es necesariamente malo! Ya vimos un ejemplo de un estimador sesgado que puede ser mejor que otro insesgado, en el sentido de que tiene menor ECM.
- Toda una área de la estadística moderna estudia maneras inteligentes de balancear sesgo y varianza para tener un ECM bajo.

# Sesgo

## Ejemplo

Dada una muestra aleatoria $X_1, \ldots, X_n$ con esperanza $\mu$ y varianza $\sigma^2$ y queremos estimar $\sigma^2$. El estimador de la varianza

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$$

¿Será insesgado? Calculemos

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] \quad y \quad E\left[(\overline{X}_n)^2\right].$$

La primera es fácil:

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) = E(X_1^2) = \sigma^2 + \mu^2.$$

# Sesgo

## Ejemplo

Ahora

$$E\left[(\overline{X}_n)^2\right] = Var(\overline{X}_n) + (E(\overline{X}_n))^2 = \frac{\sigma^2}{n} + \mu^2.$$

Luego

$$E(\widehat{\sigma}^2) = \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \left(1 - \frac{1}{n}\right)\sigma^2$$

y

$$E(\widehat{\sigma}^2) - \sigma^2 = \frac{-\sigma^2}{n}.$$

Por lo tanto, el estimador es sesgado: subestima el valor real.

# Sesgo

## Ejemplo

Si consideramos ahora

$$S^2 = \frac{n}{n-1}\widehat{\sigma}^2.$$

es fácil ver que $S^2$ es insesgado para $\sigma^2$.

## Observación

El estimador $S^2$ muchas veces se encuentra definido como

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

La definición que vimos y esta son equivalentes.

# Sesgo

## Comentarios

- Que $S^2$ sea insesgado para $\sigma^2$ no implica que $S$ sea insesgado para $\sigma$, de hecho no lo es. En general no es cierto que si $\widehat{\theta}_n$ es insesgado para $\theta$ entonces $f(\widehat{\theta}_n)$ es insesgado para $f(\theta)$.
- Un estimador puede ser sesgado y ser buen estimador ($\widehat{\sigma}^2$). Un estimador puede ser insesgado y no ser bueno (¿por ejemplo?).
- ¿Cúal deberíamos usar? $S^2$ (insesgado) o $\widehat{\sigma}^2$ (sesgado). En la práctica no importa mucho...

# Ejemplos

**Ejemplo 1**

Tenemos una muestra aleatoria $X_1, \ldots, X_n$ con densidad

$$f(x) = \frac{1}{2}(1 + \theta x)\, I_{(-1,1)}(x)\,.$$

Muestre que $3\overline{X}_n$ es un estimador insesgado de $\theta$.

# Error estándar

Además de reportar la estimación debemos indicar su precisión.

**Definition (Error estándar)**

El **error estándar** de un estimador $\widehat{\theta}_n$ es otro nombre para el desvío estándar de su distribución: $\sqrt{Var(\widehat{\theta}_n)}$.

**Idea**

- El error estándar mide la variabilidad del estimador. Si repetimos el experimento, cambia la muestra y posiblemente cambie el valor estimado. Necesitamos una idea de qué tan estable es el estimador.
- En algunos casos podremos calcularlo exactamente y en otros, la mayoría, tendremos que estimarlo.

# Error estándar

## Ejemplo

El error estándar de $\overline{X}_n$ es

$$\sqrt{Var\left(\overline{X}_n\right)} = \frac{\sqrt{Var(X)}}{\sqrt{n}}.$$

Un estimador de $\sqrt{Var(X)}$ es

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}_n\right)^2}$$

otro es

$$S = \widehat{\sigma}\sqrt{\frac{n}{n-1}}.$$

Luego, podemos estimar el error estándar de $\overline{X}_n$ con $\frac{\widehat{\sigma}}{\sqrt{n}}$.

# Error estándar

## Ejemplo

Vimos además, usando el TCL y el Lema de Slutzky, que

$$\overline{X}_n - \mu \approx N\left(0, \frac{\widehat{\sigma}^2}{n}\right) \quad \text{y} \quad \sqrt{n}\,\frac{(\overline{X}_n - \mu)}{\widehat{\sigma}} \approx N\left(0, 1\right).$$

Por lo tanto, si $Z \sim N(0,1)$.

$$P\left(\left|\overline{X}_n - \mu\right| \leq \frac{2\widehat{\sigma}}{\sqrt{n}}\right) \approx P\left(|Z| \leq 2\right) \approx 0.95.$$

Es decir, si $n$ es grande y repetimos muchas veces el experimento, aproximadamente en el 95% de las repeticiones, el valor de $\overline{X}_n$ estará a una distancia menor o igual que dos desvíos estándar estimados del parámetro verdadero $\mu$.

**¡Esto habla sobre ninguna realización particular del experimento!**

# Error estándar

## Ejemplo

En el proceso de ensamblaje de celulares en una fábrica se cometen errores. El número de rayaduras en la pantalla de un celular elegido al azar de la línea de montaje es una variable aleatoria $X$ con distribución Poisson de parámetro $\lambda$. Se toma una muestra aleatoria de 150 celulares y se registran los siguientes resultados:

| número de rayaduras | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| ocurrencias en la muestra | 18 | 37 | 42 | 30 | 13 | 7 | 2 | 1 |

- Proponga un estimador insesgado para $\lambda$. Calcule el valor estimado obtenido para esta muestra.
- Proponga un estimador para la varianza de $X$. Proponga un estimador para el error estándar del estimador propuesto en el item anterior. ¿Cúal es el error estándar estimado para esta muestra?

# Modelos paramétricos

**Notación**

A continuación usaremos la palabra densidad para referirnos tanto a la densidad de variables/vectores aleatorios continuos como a la función de probabilidad puntual de variables/vectores aleatorios discretos.

# Modelos paramétricos

## Definition (Modelo paramétrico)

Supongamos que tenemos una muestra aleatoria $X_1, \ldots, X_n$ con cierta función de densidad $f$. Un **modelo parámetrico** para la distribución de la muestra aleatoria es postular que $f$ pertenece un conjunto de funciones de densidad que depende de un parámetro desconocido $\theta$:

$$f \in \{f(x) = f(x; \theta) \text{ para cierto } \theta\},$$

donde la forma de las funciones $f(x; \theta)$ es completamente conocida, excepto por $\theta$.

**Modelo parametrico —> Modelo que se lo puede definir con un numero finito de parametros, ej-> Modelo normal, modelo exponencial**

## Idea

Suponer un modelo paramétrico para una muestra aleatoria es suponer que conocemos enteremante la distribución de los datos, **excepto por una cantidad finita de parámetros**.

# Modelos paramétricos

Ya vimos mucho modelos paramétricos (sin haberlos llamado así)

### Ejemplo 1

Tenemos una muestra aleatoria $X_1, \ldots, X_n$ con densidad

$$f(x) = 1/2(1 + \theta x) I_{(-1,1)}(x)$$

### Ejemplo 2

Supongamos que el tiempo de reacción de una persona intoxicada a cierto estímulo es una variable aleatoria $X$, que tiene distribución uniforme en el intervalo $(0, \theta)$ para un valor de $\theta$ desconocido. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ y queremos estimar $\theta$.

### Ejemplo 3

Suponer que los datos provienen de una normal con parámetros desconocidos, o de una Poisson con parámetro desconocido también son modelos paramétricos.

# Modelos paramétricos

Los modelos **no paramétricos** son aquellos que no ponen restricciones parámetricas a la distribución que suponemos tienen los datos.

El siguiente es un ejemplo de un modelo no-paramétrico, pero solo nos interesa estimar ciertos parámetros de la distribución.

**Ejemplo**

Nos interesa estudiar el estado de salud de la población Argentina, en particular su peso. Supongamos que el peso de una persona elegida al azar es una variable aleatoria con esperanza $\mu$ y varianza $\sigma^2$ desconocidos. A partir de una muestra aleatoria de pesos $X_1, \ldots, X_n$ queremos estimar $E(X)$ y $Var(X)$.

# Modelos paramétricos

Los siguientes ejemplos son modelos no paramétrico, nos interesa estimar una función entera que depende de la distribución más que una única característica numérica.

**Ejemplo**

- Estimar **toda** la función de distribución acumulada del ingreso anual de una población. A partir de la estimación podemos obtener estimaciones de diferentes cuantiles.

# Modelos paramétricos

## Modelos paramétricos

Los modelos parámetricos hacen suposiciones muy fuertes. Suponen que conocemos prácticamente todo sobre cómo se generan los datos.

## ¿Cúando es razonable un modelo parámetrico?

Depende. . .

- En algunos problemas físicos o de ingeniería, hay experimentos duros que validan el modelado de ciertas cantidades aleatorias de manera paramétrica. Un ejemplo: como emite radiación un átomo inestable sigue una distribución Poisson.
- En problemas socioeconómicos, es más díficil creerse que valga exactamente un modelo paramétrico.

# Modelos paramétricos

George P. Box: "All models are wrong, but some are useful."

# Modelos paramétricos

Aunque un modelo paramétrico no valga exactamente, si vale "aproximadamente" puede ser útil para **aprender** de los datos.

¿Cómo sabemos si un modelo parámetrico es razonable?

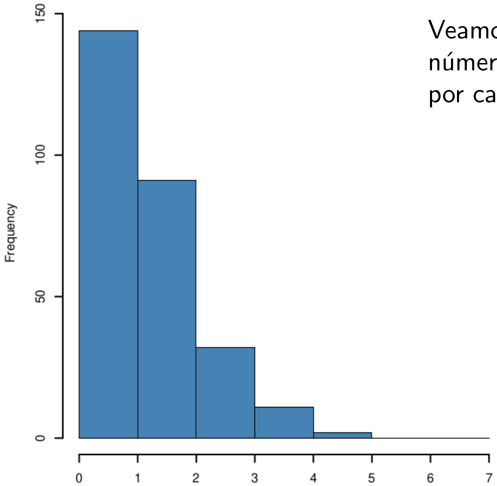
Esto escapa al alcance de esta materia, pero veamos algunas ideas en un ejemplo simple:

Tenemos información sobre el número de soldados Prusianos muertos por recibir una patada de caballo. Los datos están separados por grupo de caballería y por año.

# Ejemplo: Containers perdidos

| Año | C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C14 | C15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1875 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 1876 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1877 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 0 |
| 1878 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 1879 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 0 | 1 | 0 | 0 | 2 | 1 | 0 |
| 1880 | 0 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 2 | 1 | 4 | 3 | 0 |
| 1881 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1882 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 4 | 1 |
| 1883 | 0 | 0 | 1 | 2 | 0 | 1 | 2 | 1 | 0 | 1 | 0 | 3 | 0 | 0 |
| 1884 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 1 |
| 1885 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 1 |
| 1886 | 2 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 3 | 0 |
| 1887 | 1 | 1 | 2 | 1 | 0 | 0 | 3 | 2 | 1 | 1 | 0 | 1 | 2 | 0 |
| 1888 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 1889 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 2 | 0 | 2 |
| 1890 | 1 | 2 | 0 | 2 | 0 | 1 | 1 | 2 | 0 | 2 | 1 | 1 | 2 | 2 |
| 1891 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 3 | 3 | 1 | 0 |
| 1892 | 1 | 3 | 2 | 0 | 1 | 1 | 3 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 1893 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 1 | 3 | 0 | 0 |
| 1894 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 |

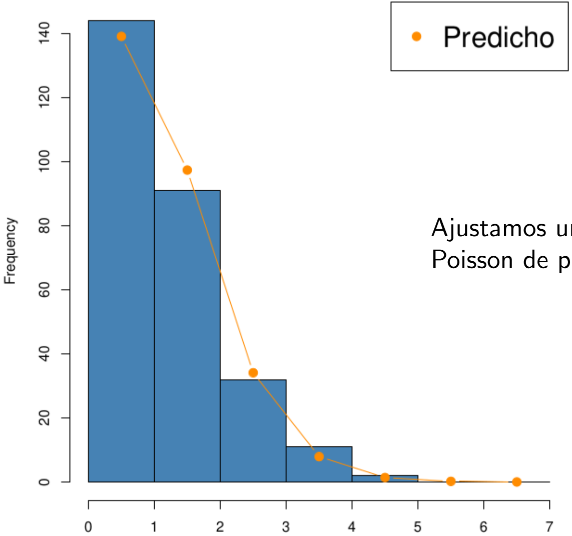
# Ejemplo: Patadas de caballo




Veamos la distribución del número de soldados fallecidos por caballería y por año.

| | |
|---|---|
| 0 | 144 |
| 1 | 91 |
| 2 | 32 |
| 3 | 11 |
| 4 | 2 |
| 5 | 0 |
| 6 | 0 |
| 7 | 0 |

# Ejemplo: Containers perdidos




Ajustamos una distribución Poisson de parámetro $\lambda = 0.7$.

# Métodos de estimación

## Métodos de estimación

Hasta ahora hemos obtenido estimadores a través de argumentos intuitivos. A continuación, estudiaremos **métodos** para construir estimadores en problemas generales: dada una muestra $X_1, \ldots, X_n$ y un parámetro $\theta$ asociado a la distribución de la muestra, ¿cómo podemos estimar $\theta$?

- método de los momentos
- método de máxima verosimilitud

# Método de los momentos

**Definition (Momento poblacional)** Son valores fijos

Para $k \in \mathbb{N}$, se define el $k$-ésimo **momento poblacional** de la variable aleatoria $X$ como $E\left(X^k\right)$.

**Definition (Momento muestral)** Son variables aleatorias

Dada una muestra aleatoria con la misma distribución que $X$. Se define el $k$-ésimo **momento muestral** de $X_1, \ldots, X_n$ como $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.

- El primer momento poblacional es $E(X)$.
- El primer momento muestral es $\overline{X}_n$, la media muestral.
- El segundo momento poblacional es $E(X^2)$.
- El segundo momento muestral es $\frac{1}{n}\sum_{i=1}^{n} X_i^2$.

Notar que los momentos poblacionales son números fijos, y los muestrales son variables aleatorias.

# Método de los momentos

**Definition (Método de los momentos)**

Sea $X_1, \ldots, X_n$ una muestra aleatoria con densidad $f(x;\theta)$ donde $\theta = (\theta_1, \ldots, \theta_m)$ son parámetros desconocidos. Supongamos que los primeros $m$ momentos poblacionales son funciones de $\theta_1, \ldots, \theta_m$, es decir que para $k = 1, \ldots, m$

$$E\left(X^k\right) = g_k(\theta_1, \ldots, \theta_m).$$

Los estimadores del método de los momentos $\widehat{\theta}^{(1)}_{MM}, \ldots, \widehat{\theta}^{(m)}_{MM}$ cumplen

**E(X) = promedio(X)**

$$\begin{cases} g_1\left(\widehat{\theta}^{(1)}_{MM}, \ldots, \widehat{\theta}^{(m)}_{MM}\right) & = \frac{1}{n}\sum_{i=1}^{n} X_i \\ g_2\left(\widehat{\theta}^{(1)}_{MM}, \ldots, \widehat{\theta}^{(m)}_{MM}\right) & = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ & \vdots \\ g_m\left(\widehat{\theta}^{(1)}_{MM}, \ldots, \widehat{\theta}^{(m)}_{MM}\right) & = \frac{1}{n}\sum_{i=1}^{n} X_i^m \end{cases}$$

**Momentos poblacionales** **Momentos muestrales**

# Método de los momentos

**Notación**

En $\widehat{\theta}_{MM}$ omitimos el subíndice $n$ que indica el tamaño de muestra para no sobrecargar la notación.

# Método de los momentos

## Estimador de momentos para la Bernoulli

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $Ber(p)$. El estimador de momentos de $p$ se obtiene de la siguiente manera. El primer momento de una $Ber(p)$ (su esperanza) es igual a $p$. Luego el estimador es

$$\widehat{p}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Es decir, el estimador es la media muestral.

**E(X) = p —>Primer Momento poblacional**
**Ahora, ¿Cómo hallo el ESTIMADOR de momentos de “p”?. Donde veas “MOMENTO POBLACIONAL”, cambialo por “MOMENTO MUESTRAL”, en este caso, donde ves E(X), pone Promedio(X), y en la ecuacion donde veas el parametro a estimar, ponele ^ (notacion de estimador), QUEDARIA...**

$\bar{X} = \hat{p}$

# Método de los momentos

## Estimador de momentos para la normal

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $N(\mu, \sigma^2)$.

Calculemos el estimador de momentos de $\theta = (\mu, \sigma^2)$. Tenemos que resolver

$$E(X_1) = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$E\left(X_1^2\right) = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Sabemos que $E(X_1) = \mu$ y $E(X_1^2) = \sigma^2 + \mu^2$.

# Método de los momentos

## Estimador de momentos para la normal

Luego

$$\widehat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\sigma}^2_{MM} + \widehat{\mu}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

y por lo tanto el estimador de momentos de $\mu$, $\widehat{\mu}_{MM}$, es igual a $\overline{X}_n$ y el estimador de momentos de $\sigma_2$ es

$$\widehat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2,$$

que es el estimador $\widehat{\sigma}^2$ que ya habíamos estudiado antes.

# Método de los momentos

## Ejemplo

Sea $X$ el tiempo que dedica un alumno a estudiar para un examen. Supongamos que la densidad de $X$ es

$$f(x;\theta) = (\theta + 1)x^{\theta} I_{(0,1)}(x)$$

y que sabemos que $\theta > -1$.

Calcule el estimador de momentos de $\theta$ y dar el valor estimado para la siguiente muestra:

$$x_1 = 0.92, x_2 = 0.79, x_3 = 0.90, x_4 = 0.65, x_5 = 0.86,$$
$$x_6 = 0.47, x_7 = 0.73, x_8 = 0.97, x_9 = 0.94, x_{10} = 0.77$$

Estimar la probabilidad de que un alumno dedique mas de 0.5 a estudiar.

# Variable aleatoria exponencial

**X1,,xn tiene dist exp —> E(x) = 1/lambda —> X(raya) = 1/lambda(sombrero) —> lambda (sombrero) = 1 / x(raya)**

## Ejemplo

Supongamos que el tiempo, en minutos, que pasa entre que llegan pedidos a un centro de distribución de una compañía alimenticia es una variable $X \sim Exp(\lambda)$.

1. Dada una muestra aleatoria $X_1, \ldots, X_n$, calcular el estimador de momentos de $\lambda$.
2. Dada la siguiente muestra:

   2.63 1.96 0.13 1.83 0.44 0.23 0.11 0.06 0.31 0.44

   Obtenga el valor estimado para esta muestra. Para ese valor estimado, ¿cúal es la probabilidad de que pase más de un minuto entre dos pedidos?

# Método de los momentos

## Ventajas

- Es una idea intuitiva.
- Los estimadores son fáciles de calcular.

## Desventajas

- No está guiado por ningún principio teórico (es un poco ad-hoc).
- La teoría asintótica muestra que no en general no es **óptimo**.

Por estas razones nos enfocaremos más en el **método de máxima verosimilitud** que estudiaremos con más profundidad.

# Máxima verosimilitud

A continuación presentaremos otro método de estimación, llamado el método de **máxima verosimilitud**. Este método tiene muchas propiedades interesantes:

- Está basado en una idea muy intuitiva (más que el método de los momentos).
- Es invariante: si $\widehat{\theta}_{MV}$ es el estimador de máxima verosimilitud de $\theta$, $g(\widehat{\theta}_{MV})$ es el estimador de máxima verosimilitud de $g(\theta)$ (momentos no tiene esta propiedad).

# Máxima verosimilitud

- Bajo condiciones generales, el estimador de máxima verosimilitud es asíntoticamente normal, es decir, para tamaños de muestra grandes su distribución muestral es aproximadamente normal (momentos tiene la misma propiedad).
- Más aún la normal está centrada en el parámetro verdadero (momentos tiene la misma propiedad).
- Bajo condiciones generales, la **varianza asintótica** del estimador de máxima verosimilitud es la mínima posible entre los estimadores que tienen sesgo que es asíntoticamente cero (**momentos no tiene esta propiedad**).

# La función de verosimilitud

Supongamos que tenemos una muestra aleatoria $X_1, \ldots, X_n$. El vector aleatorio $(X_1, \ldots, X_n)$ tiene cierta función de densidad conjunta.

Supongamos que esta función depende de un $\theta$ (posiblemente un vector) desconocido. Digamos que $f(x_1, \ldots, x_n; \theta)$ es la función de densidad de $(X_1, \ldots, X_n)$ y que conocemos $f$, excepto por $\theta$. Es decir, tenemos un modelo paramétrico para la distribución de los datos.

**Definition (Función de verosimilitud)**

Sea $f(x_1, \ldots, x_n; \theta)$ la función de densidad de una muestra aleatoria $X_1, \ldots, X_n$, que depende de un parámetro desconocido $\theta$. La función

$$L_n(\theta) = f(x_1, \ldots, x_n; \theta),$$

pensada cómo función de $\theta$ para $x_1, \ldots, x_n$ fijos, se llama la función de verosimilitud.

# La función de verosimilitud

## Proposición

*Sea $L_n(\theta)$ la función de verosimilitud de una muestra aleatoria $X_1, \ldots, X_n$. Llamemos $f(x;\theta)$ a la función de densidad que tienen en común las v.a. de la muestra. Entonces*

$$L_n(\theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta)$$

## Proof.

Como $X_1, \ldots, X_n$ son independientes, su función de densidad se factoriza. Luego

$$L_n(\theta) = f(x_1, \ldots, x_n;\theta) = f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta).$$

$\square$

# La función de verosimilitud

## Notación

Usaremos la notación

$$\mathcal{L}_n(\theta) = f(X_1;\theta)f(X_2;\theta)\dots f(X_n;\theta)$$

es decir, $\mathcal{L}_n(\theta)$ es la 'versión aleatoria' de la función de verosimilitud.

# La función de verosimilitud

Veamos que tipo de información que está codificada en la función de verosimilitud.

**Example**

Consideremos una muestra aleatorias de 10 bicicletas del nuevo sistema de bicicleta públicas en CABA.

Sea $p = P(\text{bicicleta vandalizada})$ y definamos $X_i = 1$ si la $i-$ésima bicicleta fue vandalizada y $X_i = 0$ en otro caso. Luego $X_i \sim Ber(p)$. Llamemos $\theta = p$, que es el parámetro que nos interesa estimar.

Supongamos que en una muestra dada se ve que la primera, la tercera y la décima fueron vandalizadas, mientras que las otras están bien. Es decir, las $x_i$ son

$$1, 0, 1, 0, 0, 0, 0, 0, 0, 1$$

y por lo tanto

$$L_{10}(\theta) = \theta^{\sum_i x_i}(1-\theta)^{10-\sum_i x_i} = \theta^3(1-\theta)^7$$

**La maxima verosimilitud me dice la probabilidad de ver esta muestra en particular entre todas las que pueden existir**

# La función de verosimilitud

### Example

Para cada $\theta$, $L_{10}(\theta)$ mide cúan probable es haber observado la muestra que observamos. Por ejemplo

$$L_{10}(0.5) = 0.5^3 0.5^7 = 0.001 \quad \text{y} \quad L_{10}(0.\overset{1}{\cancel{3}}) = 0.1^3 0.9^7 = 0.002.$$

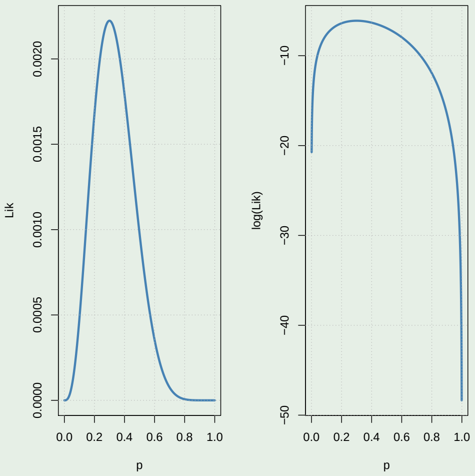
Podemos definir un estimador como el valor de $\theta$ que maximiza $L_{10}(\theta)$, es decir, que hace más probable haber observado los datos que observamos. Esta es la idea fundamental detrás del método de estimación de máxima verosimilitud.

**Para cada p de vandalizmo, hay un valor distinto de verosimilitud.**
**Es mas verosimil el valor de p=0.1, que un valor de p=0.5**

# La función de verosimilitud

**El estimador de maximaverosimilitud, es aquel valor donde se MAXIMIZA la verosimilitud**

## Example




**Es lo mismo querer maximizar la verosimilitud que el log(verosimilitud), ya que es una funcion creciente**

# La función de verosimilitud

**Example**

Tenemos entonces que resolver el problema de encontrar el valor de $\theta$ que maximiza $L_{10}(\theta)$. Este es un problema de optimización como los que ya vieron en análisis.

En este caso (y en casi todos los casos), es más fácil encontar el $\theta$ que maximiza $\log(L_{10}(\theta))$. Notar que las soluciones de estos dos problemas de maximización son iguales ya que, como el logaritmo es creciente,

$$L_{10}(\widehat{\theta}) \geq L_{10}(\theta) \quad \forall\theta \quad \Leftrightarrow \quad \log(L_{10}(\widehat{\theta})) \geq \log(L_{10}(\theta)) \quad \forall\theta.$$

# La función de verosimilitud

**Example**

Tenemos entonces que maximizar $\log(L_{10}(\theta))$. Como se trata de una función suave, para encontrar su máximo, nos alcanza con

1. Derivar.
2. Igualar a cero y despejar.
3. Chequear que la solución encontrada sea un máximo, por ejemplo viendo que la segunda derivada es negativa en la solución.

# La función de verosimilitud

## Example

Ahora,

$$\log(L_{10}(\theta)) = \sum_i x_i \log(\theta) + (10 - \sum_i x_i) \log(1-\theta).$$

$$\frac{d}{d\theta} \log(L_{10}(\theta)) = \frac{\sum_i x_i}{\theta} - \frac{10 - \sum_i x_i}{1-\theta}.$$

$$\frac{\sum_i x_i}{\theta} - \frac{10 - \sum_i x_i}{1-\theta} = 0 \quad \Leftrightarrow \theta = \frac{\sum_{i=1}^{10} x_i}{10}.$$

Además para todo $\theta$,

$$\frac{d^2}{d\theta^2} \log(L_{10}(\theta)) = -\frac{\sum_i x_i}{\theta^2} - \frac{10 - \sum_i x_i}{(1-\theta)^2} < 0.$$

# La función de verosimilitud

**Example**

El valor estimado que obtenemos con este método entonces es

$$\frac{\sum_{i=1}^{10} x_i}{10} = 0.3.$$

El estimador en general, para una muestra de tamaño 10, se obtiene de la misma manera, pero ahora reemplazando $L_{10}(\theta)$ por $\mathcal{L}_{10}(\theta)$. Obtenemos que el estimador es

$$\frac{\sum_{i=1}^{10} X_i}{10} = \overline{X}_{10},$$

la media muestral.

# Máxima verosimilitud

## Definition (Máxima verosimilitud)

Sea $X_1, \ldots, X_n$ una muestra aleatoria que tiene función de densidad en común $f(x;\theta)$, para un $\theta$ (posiblemente vectorial) desconocido. El estimador de máxima verosimilitud se define como el $\widehat{\theta}_{MV}$ que maximiza la verosimilitud (aleatoria), es decir que cumple que para todo $\widetilde{\theta}$

$$\mathcal{L}_n(\widehat{\theta}_{MV}) = f(X_1;\widehat{\theta}_{MV}) \ldots f(X_n;\widehat{\theta}_{MV}) \geq \mathcal{L}_n(\widetilde{\theta}) = f(X_1;\widetilde{\theta}) \ldots f(X_n;\widetilde{\theta})$$

Para una realización de la muestra $x_1, \ldots, x_n$, el valor estimado se obtiene reemplazando $X_i$ por $x_i$, luego, el valor estimado es el que maximiza la verosimilitud (no aleatoria) $L_n(\theta)$.

## Consejo

Casi siempre es más fácil maximizar $\log(\mathcal{L}_n(\theta))$ que $\mathcal{L}_n(\theta)$. Esto es porque el logartimo convierte productos en sumas, y las sumas son más fáciles de derivar que los productos.

# Máxima verosimilitud: ejemplos

## Example (Poisson)

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución Poisson de parámetro $\lambda$. La verosimilitud es

$$\mathcal{L}_n(\lambda) = \exp(-\lambda n) \frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}.$$

Luego

$$\log\left(\mathcal{L}_n(\lambda)\right) = -\lambda n + \sum_{i=1}^n X_i \log(\lambda) - \log(\prod_{i=1}^n X_i!).$$

Es fácil ver entonces que el estimador de máxima verosimilitud es

$$\widehat{\lambda}_{MV} = \frac{1}{n} \sum_{i=1}^n X_i.$$

# Máxima verosimilitud: ejemplos

**Example (Poisson)**

Como el estimador de máxima verosimilitud es la media muestral, ya sabemos muchos sobre el.

$\widehat{\lambda}_{MV}$ es insesgado ya que

$$E\left(\widehat{\lambda}_{MV}\right) = E(\overline{X}_n) = E(X_1) = \lambda$$

# Máxima verosimilitud: ejemplos

## Example (Poisson)

Sabemos además estimar su error standard (recordar clases pasadas...)

$$\frac{\sqrt{\hat{\lambda}}}{\sqrt{n}}$$

y hasta sabemos que para tamaños de muestra grande, por el TCL y el Lemma de Slutzky

$$\widehat{\lambda}_n - \lambda \approx N\left(0, \frac{\hat{\lambda}}{n}\right).$$

# Máxima verosimilitud: ejemplos

## Example (Exponencial)

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución exponencial de parámetro $\lambda$. La verosimilitud es

$$\mathcal{L}_n(\lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} X_i\right) \prod_{i=1}^{n} I_{(0,+\infty)}(X_i).$$

Luego

n* log(lambda)

$$\log\left(\mathcal{L}_n(\lambda)\right) = \cancel{\lambda n} - \lambda \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} \log(I_{(0,+\infty)}(X_i)).$$

Es fácil ver entonces que el estimador de máxima verosimilitud es

$$\widehat{\lambda}_{MV} = 1/(\overline{X}_n).$$

# Máxima verosimilitud: ejemplos

**Example (Exponencial)**

¡El estimador es sesgado!

$$E\left(\frac{1}{\overline{X}_n}\right) \neq \frac{1}{E\left[\overline{X}_n\right]} = \frac{1}{E(X_1)} = \lambda.$$

# Invarianza

El estimador de máxima verosimilitud tiene la siguiente propiedad útil.

**Proposición (Invarianza)**

*Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $f(x; \theta)$. Sea $g : \mathbb{R} \to \mathbb{R}$ una función. Entonces el estimador de máxima verosimilitud de $g(\theta)$ es igual $g(\widehat{\theta}_{MV})$.*

# Invarianza

## Example

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución $Ber(p)$. Calcular el estimador de máxima verosimilitud de las odds

$$\mathbf{q} = \frac{p}{1-p}\,.$$

# Máxima verosimilitud

**Example (Uniforme)**

Sea $X_1, \ldots, X_n$ una muestra aleatoria con distribución uniforme en el intervalo $(0, \theta]$. Calculemos el estimador de máxima verosimilitud de $\theta$. La verosimilitud es

$$\mathcal{L}_n(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta]}(X_i),$$

que no es derivable como función de $\theta$. La estrategia que veniamos usando no sirve.

# Máxima verosimilitud

## Example (Uniforme)

Notemos que

$$\mathcal{L}_n(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{(0,\theta]}(X_i),$$

es no-negativa y que es positiva siempre que $\theta \geq X_i$ para todo $i = 1, \ldots, n$.

Pero entonces $\mathcal{L}_n(\theta) > 0$ siempre que $\theta \geq \max(X_1, \ldots, X_n)$. Además, si $\theta \geq \max(X_1, \ldots, X_n)$, $\mathcal{L}_n(\theta) > 0$ es decreciente en $\theta$. Concluimos que la verosimilitud se maximiza en $\max(X_1, \ldots, X_n)$ y por lo tanto el estimador de máxima verosimilitud es

$$\widehat{\theta}_{MV} = \max(X_1, \ldots, X_n).$$

# Máxima verosimilitud

## Example (Uniforme)

En este contexto ya habíamos estudiado a

$$\max\left(X_1, \ldots, X_n\right)$$

como estimador de $\theta$, pero sin saber que era el de máxima verosimilitud. Habíamos visto que

$$E\left(\widehat{\theta}_{MV}\right) = \frac{n}{n+1}\theta$$

y

$$Var\left(\widehat{\theta}_{MV}\right) \underset{n\to\infty}{\to} 0.$$

Con lo cual $\widehat{\theta}_{MV}$ es sesgado (pero el sesgo tiende a cero).

# Máxima verosimilitud

**Example (Uniforme)**

Es fácil ver que $\widehat{\theta}_{MM} = 2\overline{X}_n$. Se puede probar que $\widehat{\theta}_{MV}$ tiene un error cuadrático medio menor que $\widehat{\theta}_{MM}$, con lo cual en este caso la estrategia de máxima verosimilitud da un mejor resultado.

# Máxima verosimilitud

## Example (Laplace)

Decimos que una variable aleatoria tiene distribución Laplace de parámetro $\theta$ si su función de densidad es

$$f_X(x;\theta) = (1/2)\exp(-|x-\theta|).$$

Supongamos que tenemos una muestra aleatoria $X_1, \ldots, X_n$ de una distribución Laplace de parámetro $\theta$. ¿Cuál es el estimador de máxima verosimilitud de $\theta$?

# Máxima verosimilitud

## Example (Laplace)

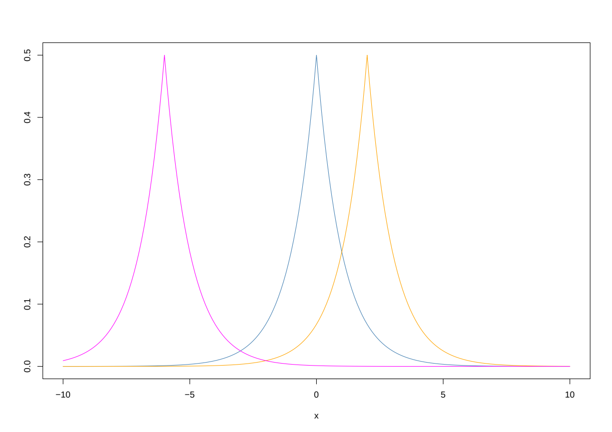



Figure: Densidad de laplace con: $\theta = -6$ (magenta), $\theta = 0$ (celeste) y $\theta = 2$ (naranja).

# Máxima verosimilitud

## Example (Laplace)

Tenemos que

$$\mathcal{L}_n(\theta) = (1/2)\exp(-|X_1 - \theta|)\dots(1/2)\exp(-|X_n - \theta|)$$
$$= (1/2^n)\exp\left(-\sum_{i=1}^{n}|X_i - \theta|\right).$$

Luego

$$\log(\mathcal{L}_n(\theta)) = n\log(1/2) - \sum_{i=1}^{n}|X_i - \theta|,$$

que no es derivable, ya que el valor absoluto no es derivable en 0.

## Example (Laplace)

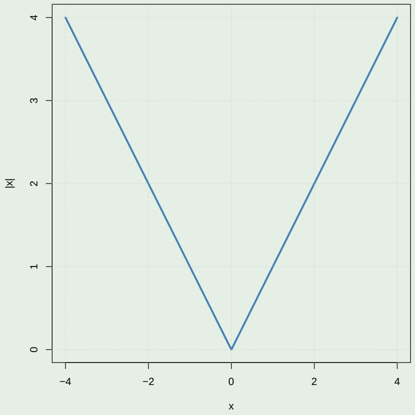



Figure: Valor absoluto.

# Máxima verosimilitud

**Definition (Estadísticos de orden)**

Dada una muestra aleatoria $X_1, \dots, X_n$, el k-ésimo estadístico de orden $X_{(k)}$ es el k-ésimo valor más chico en la muestra.

**Definition (Mediana muestral)**

Dada una muestra aleatoria $X_1, \dots, X_n$, se define la mediana muestral como

$$\text{median}(X_1, \dots, X_n) = X_{((n+1)/2)} \text{ si n es impar}$$

y

$$\text{median}(X_1, \dots, X_n) = \frac{X_{(n/2)} + X_{(n/2+1)}}{2} \text{ si n es par}$$

# Máxima verosimilitud

## Example (Laplace)

Utilizando otras tecnicas de maximización resulta

$$\widehat{\theta}_{MV} = \text{mediana}(X_1, \ldots, X_n).$$

Es fácil ver que el estimador de momentos es

$$\widehat{\theta}_{mom} = \overline{X}_n$$

# Máxima verosimilitud

## Example (Laplace)

Recordemos que la mediana poblacional de una variable aleatoria continua se define como el único valor $q$ tal que $P(X \leq q) = 1/2$.

En la práctica veran que si $X$ es continua y simétrica respecto de $\theta$, entonces la mediana poblacional de $X$ es igual a la esperanza de $X$, que a su vez es igual a $\theta$.

Tenemos entonces que si para la distribución Laplace, el estimador de máxima verosimilitud de su mediana poblacional es la mediana muestral.

**Si una funcion, como la funcion modulo o la funcion normal,**
**es continua y simetrica, la mediana coincide con la esperanza, por lo que**
**ahora calculamos la mediana**

# Máxima verosimilitud

**Example (Laplace)**

Se puede mostrar que $\widehat{\theta}_{MV}$ es insesgado.

# Máxima verosimilitud

**Example**

Supongamos que tenemos una muestra aleatoria $X_1, \ldots, X_n$ tal que $X_i$ tiene densidad

$$f_X(x|\alpha) = \frac{1 + \alpha x}{2} I_{[-1,1]}(x)$$

para cierto $\alpha \in [-1, 1]$. Nos interesa estimar $\alpha$. La esperanza de $X_i$ es $\frac{\alpha}{3}$. El estimador de momentos de $\alpha$ es entonces $\widehat{\alpha}_{MM} = 3\overline{X}_n$. Es fácil ver que $\widehat{\alpha}_{MM}$ es insesgado.

# Máxima verosimilitud

## Example

Calculemos ahora el estimador de máxima verosimilitud. Tenemos que

$$\mathcal{L}_n(\alpha) = \prod_{i=1}^{n} \left( \frac{1 + \alpha X_i}{2} \right) I_{[-1,1]}(X_i).$$

Luego

$$\log(\mathcal{L}_n(\alpha)) = \sum_{i=1}^{n} \log\left(1 + \alpha X_i\right) - n\log(2) + \sum_{i=1}^{n} \log\left(I_{[-1,1]}(X_i)\right).$$

# Máxima verosimilitud

## Example

Tenemos que resolver entonces

$$\frac{d}{d\alpha}\log(\mathcal{L}_n(\alpha)) = \sum_{i=1}^{n} \frac{1}{1+\alpha X_i} = 0,$$

que es una ecuación no–lineal que no tiene solución cerrada. El estimador de máxima verosimilitud está bien definido como la solución de esta ecuación, pero no podemos dar una fórmula cerrada para calcularlo. Si quisieramos calcularlo, tendríamos que recurrir a algún algoritmo numérico (Newton-Rhapson) para resolver aproximadamente la ecuación anterior.

## Observación

Es lo usual, para problemas complejos, que el estimador de máxima verosimilitud no tenga fórmual cerrada.

# Máxima verosimilitud

**Example**

Por qué querríamos hacer máxima verosimilitud, si momentos sale tan fácil?

Los estimadores de máxima verosimilitud tiene mejores propiedades cuando el tamaño de la muestra es grande.

**El ECM de MVS, es mas pequeño que el de momento, por lo que voy a buscar calcular los parametros a partir de este metodo. Ya que cuando el n de la muestra es grande, el ECM, es siempre menor en MVS**