

Análisis Estadístico

Estimación

San Andrés

May 24, 2022

Estimación

Problema de estimación

Dada una muestra aleatoria X_1, \dots, X_n queremos estimar un parámetro θ asociado a la distribución de la muestra.

Definition (Estimador)

Dadas variables aleatorias X_1, \dots, X_n , para un parámetro cualquiera θ , un **estimador** es una función de X_1, \dots, X_n .

Notación

Llamaremos $\hat{\theta}_n$ a un estimador del parámetro θ basado en una muestra de tamaño n .

Llamaremos estadístico a toda cuenta realizada sobre la muestra

Estimación de la varianza

Sea X_1, \dots, X_n una muestra aleatoria.

¿Cómo podemos estimar σ^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 de la varianza

Sea X_1, \dots, X_n una muestra aleatoria.

¿Cómo podemos estimar σ^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

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 $E(X^2)$?

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

Estimador de la varianza

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

¿Es el único estimador posible?

Example

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

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

Objetivo

Para distintos estimadores nos interesa responder:

- ¿Qué propiedades tiene la **distribución muestral** del estimador?
- ¿Cómo medimos si un estimador es **mejor** que otro? ¿Cuál 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?

¿Que tamaño de muestra necesito?

Medir la calidad de un estimador

Estimar el parámetro de una Bernoulli

Dada una muestra aleatoria $X_1, \dots, X_n \sim \text{Ber}(p)$, llamemos

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

Proponemos tres estimadores de p :

- La media muestral:

$$\hat{p}_1 = T/n = \bar{X}_n.$$

- La media muestral agregando 2 fracasos y 2 éxitos “artificiales” a la muestra:

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

- La media muestral de solo dos observaciones:

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

Medir la calidad de un estimador

Estimar el parámetro de una Bernoulli

Intuitivamente podemos descartar al estimador p_3 pero...

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

Error cuadrático medio

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

$$\hat{\theta}_n - \theta$$

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

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

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

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

Error cuadrático medio

Observación

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

Por conveniencia matemática.

Es mas sencillo calcular utilizando el cuadrado, que con la funcion modulo

Error cuadrático medio

Para que el error sea chico, ambos deben ser chicos

El SESGO es la Esperanza del estimador menos el valor que esta tomando el estimador

Theorem (Descomposición del ECM)

$$E \left[\left(\hat{\theta}_n - \theta \right)^2 \right] = \underbrace{\text{Var}(\hat{\theta}_n)}_{\substack{\text{varianza} \\ \text{Varianza del} \\ \text{estimador}}} + \underbrace{\left(E(\hat{\theta}_n) - \theta \right)^2}_{\text{sesgo al cuadrado}}.$$

Proof.

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

$$\text{Var}(e) = \underbrace{E(e^2)}_{\text{ECM}} - [E(e)]^2.$$

Luego

$$E(e^2) = \text{Var}(e) + [E(e)]^2.$$

Además, como θ es constante, $\text{Var}(e) = \text{Var}(\hat{\theta}_n)$.



Error cuadrático medio

Estimar el parámetro de una Bernoulli

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

\hat{p}_1 la media muestral:

$$\hat{p}_1 = \bar{X}_n.$$

\hat{p}_2 agregar 2 fracasos y 2 éxitos artificiales a la muestra y después calcular la media muestral:

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

\hat{p}_3 Promediar la primera y la última observación

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

Calculemos el ECM de cada uno.

Error cuadrático medio

Estimar el parámetro de una Bernoulli

Sabemos que



$$\text{Sesgo}(P) = E(p) - p = p - p = 0$$

$$E(\hat{p}_1) = p$$



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

Entonces el ECM de \hat{p}_1 es

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

Error cuadrático medio

Estimar el parámetro de una Bernoulli

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

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

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

Entonces

$$E((\hat{p}_3 - p)^2) = \frac{1}{2}p(1 - p). \quad \begin{matrix} \text{sesgo} \\ + (p-p)^2 \\ 0 \end{matrix}$$

Error cuadrático medio

Estimar el parámetro de una Bernoulli

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

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

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

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

Entonces

$$E((\hat{p}_2 - p)^2) = \overset{\text{Sesgo}^2}{\left(\frac{2/n - 4p/n}{1 + 4/n}\right)^2} + \overset{\text{varianza}}{\frac{np(1-p)}{(n+4)^2}}.$$

Error cuadrático medio

Estimar el parámetro de una Bernoulli

Tenemos

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

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

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

¿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

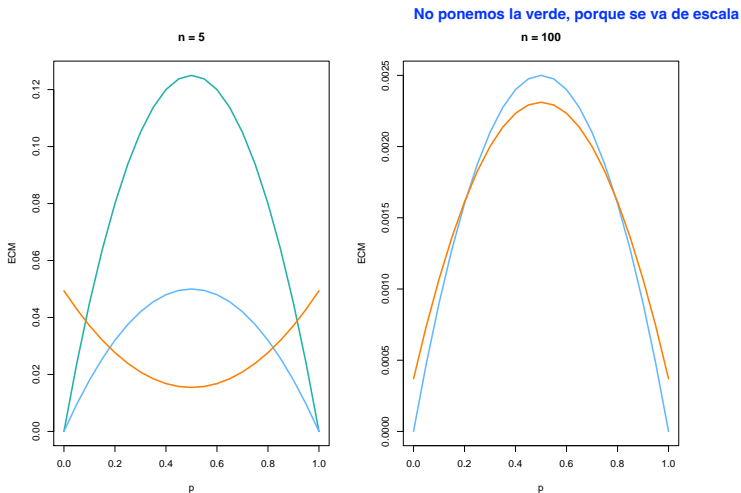


Figure: ECM para \hat{p}_1 (naranja) y \hat{p}_2 (celeste) y \hat{p}_3 (verde).

Error cuadrático medio

```
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 θ desconocido.

Tenemos una muestra aleatoria X_1, \dots, X_n y queremos estimar θ . Un estimador posible para θ es

$$\hat{\theta}_n = \max \{X_1, \dots, X_n\}.$$

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 $\hat{\theta}_n$

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

ya que son i.i.d \rightarrow

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

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

Funcion de distribucion
acumulada del maximo
(nuestro estimador)

Error cuadrático medio

Máximo de uniformes

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

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

Para sacar la esperanza,
multiplico por "x"

Entonces

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

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

y

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

SI EL SESGO DEL ESTIMADOR ES "0", EL ESTIMADOR SE DICE QUE ES "INSESGADO"

Error cuadrático medio

Máximo de uniformes

Observemos que el estimador es sesgado y que

$$\text{Sesgo}(\hat{\theta}_n) = -\frac{\theta}{n+1} \xrightarrow{n \rightarrow \infty} 0$$

y

$$\text{Var}(\hat{\theta}_n) \xrightarrow{n \rightarrow \infty} 0,$$

por lo que

$$ECM(\hat{\theta}_n) = \underbrace{\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2}_{\text{Varianza}} + \left(\underbrace{-\frac{\theta}{n+1}}_{\text{Sesgo}}\right)^2 \xrightarrow{n \rightarrow \infty} 0,$$

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.

Definition (Estimador insesgado)

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

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

para todos los valores posibles de θ . En general, la diferencia

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

se llama el **sesgo** del estimador.

Idea

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

Ejemplo

Si X_1, \dots, X_n es una muestra aleatoria con

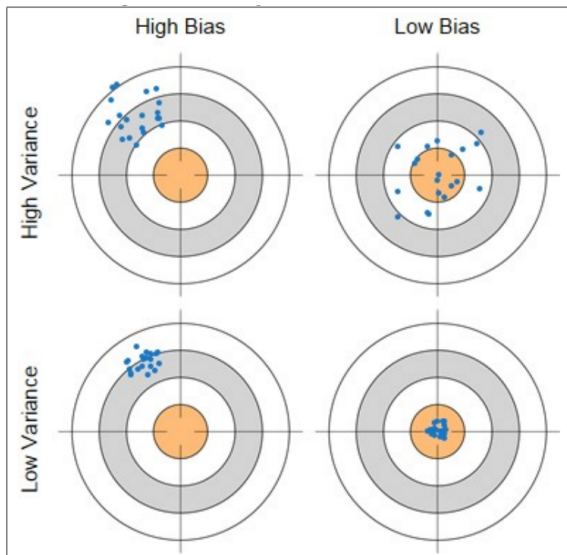
$$\mu = E(X_1) = \dots = E(X_n)$$

entonces \bar{X}_n es un estimador insesgado de μ .

Ejemplo

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

Sesgo



sesgo bajo y varianza alta, significa que apunta bien, pero tiene alta dispersion.

sesgo alto y varianza baja, apunta mal, pero tiene poca dispersion el estimador

- 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.

Ejemplo

Dada una muestra aleatoria X_1, \dots, X_n con esperanza μ y varianza σ^2 y queremos estimar σ^2 . El estimador de la varianza

$$\hat{\sigma}^2 = \underbrace{\frac{1}{n} \sum_{i=1}^n X_i^2}_{\text{}} - (\bar{X}_n)^2$$

¿Será insesgado? Calculemos

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

La primera es fácil:

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

Ejemplo

Ahora

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

Luego

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

y

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

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

Ejemplo

Si consideramos ahora

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

es fácil ver que S^2 es insesgado para σ^2 .

Observación

El estimador S^2 muchas veces se encuentra definido como

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

La definición que vimos y esta son equivalentes.

Comentarios

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

Ejemplos

Ejemplo 1

Tenemos una muestra aleatoria X_1, \dots, X_n con densidad

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

Muestre que $3\bar{X}_n$ es un estimador insesgado de θ .

$$E(3\bar{X}) = 3E(\bar{X}) = 3E(X)$$

$$E(X) = \int_{-1}^1 (x^3/2)(1 + \theta x) dx = \theta/3 \rightarrow \text{Concluyo que el estimador es insesgado}$$

Error estándar

Error estandar → Cuanto varia el estimador que me dio como resultado

Además de reportar la estimación debemos indicar su precisión.

Definition (Error estándar)

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

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 \bar{X}_n es

$$\sqrt{\text{Var}(\bar{X}_n)} = \frac{\sqrt{\text{Var}(X)}}{\sqrt{n}}.$$

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

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2}$$

otro es

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

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

Error estándar

Ejemplo

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

error de estimacion

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

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

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

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

¡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 λ . 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 consistente e insesgado para λ . Calcule el valor estimado obtenido para esta muestra.
- Proponga un estimador consistente para la varianza de X .
Proponga un estimador para el error estándar del estimador propuesto en el ítem anterior. ¿Cuál es el error estándar estimado

Media muestral = $\bar{x} = (0 \cdot 18 + 1 \cdot 37 + 2 \cdot 42 + 3 \cdot 30 + 4 \cdot 13 + 5 \cdot 7 + 6 \cdot 2 + 7 \cdot 1) / 150$

desvio = $\sqrt{((0 \cdot 18 + 1 \cdot 37 + 2 \cdot 42 + 3 \cdot 30 + 4 \cdot 13 + 5 \cdot 7 + 6 \cdot 2 + 7 \cdot 1) / 150)}$

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, \dots, X_n con cierta función de densidad f . Un **modelo paramétrico** para la distribución de la muestra aleatoria es postular que f pertenece un conjunto de funciones de densidad que depende de un parámetro desconocido θ :

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

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

Idea

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

Modelos paramétricos

Ya vimos muchos modelos paramétricos (sin haberlos llamado así)

Ejemplo 1

Tenemos una muestra aleatoria X_1, \dots, 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 θ desconocido. Tenemos una muestra aleatoria X_1, \dots, X_n y queremos estimar θ .

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 paramétricas 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 μ y varianza σ^2 desconocidos. A partir de una muestra aleatoria de pesos X_1, \dots, 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 paramétricos hacen suposiciones muy fuertes. Suponen que conocemos prácticamente todo sobre cómo se generan los datos.

¿Cuándo es razonable un modelo paramétrico?

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 difícil 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 paramétrico es razonable?

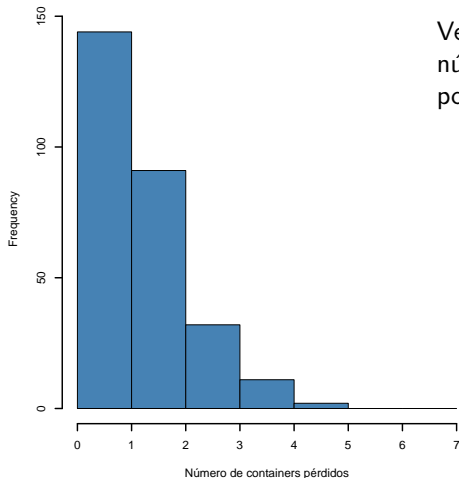
Esto escapa al alcance de esta materia, pero veamos algunas ideas en un ejemplo simple:

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

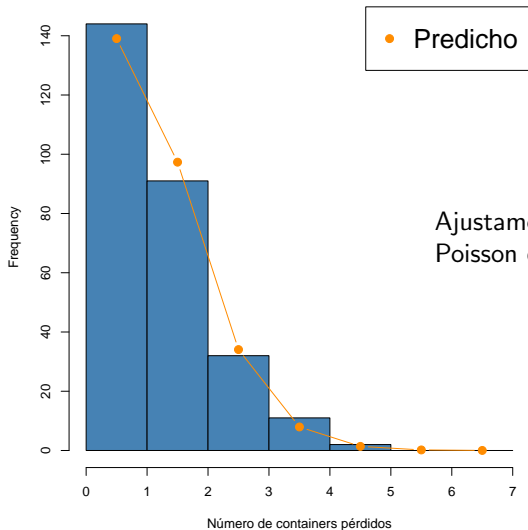
Ejemplo: Patadas de caballo



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



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, \dots, X_n y un parámetro θ asociado a la distribución de la muestra, ¿cómo podemos estimar θ ?

- método de los momentos
- método de máxima verosimilitud

Método de los momentos

Definition (Momento poblacional)

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

Definition (Momento muestral)

Dada una muestra aleatoria con la misma distribución que X . Se define el k -ésimo **momento muestral** de X_1, \dots, 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 \bar{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, \dots, X_n una muestra aleatoria con densidad $f(x; \theta)$ donde $\theta = (\theta_1, \dots, \theta_m)$ son parámetros desconocidos. Supongamos que los primeros m momentos poblacionales son funciones de $\theta_1, \dots, \theta_m$, es decir que para $k = 1, \dots, m$

$$E(X^k) = g_k(\theta_1, \dots, \theta_m).$$

Los estimadores del método de los momentos $\hat{\theta}_{MM}^{(1)}, \dots, \hat{\theta}_{MM}^{(m)}$ son los que se obtienen de igualar los primeros m momentos muestrales a los poblacionales y despejar $\theta_1, \dots, \theta_m$. Es decir, de resolver conjuntamente para $k = 1, \dots, m$.

$$g_k(\hat{\theta}_{MM}^{(1)}, \dots, \hat{\theta}_{MM}^{(m)}) = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Método de los momentos

Notación

En $\hat{\theta}_{MM}$ 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, \dots, 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

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

Es decir, el estimador es la media muestral.

Método de los momentos

Estimador de momentos para la normal

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

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

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

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

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

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

y por lo tanto el estimador de momentos de μ , $\hat{\mu}_{MM}$, es igual a \bar{X}_n y el estimador de momentos de σ^2 es

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

que es el estimador $\hat{\sigma}^2$ 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 θ y dar el valor estimado para la siguiente muestra:

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

Variable aleatoria exponencial

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 $X \sim \text{Exp}(\lambda)$.

- 1 Dada una muestra aleatoria X_1, \dots, X_n , calcular el estimador de momentos de λ .
- 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, ¿cuál 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.