Tema 2 - Estimación Puntual

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

null

Definiciones básicas

Estadística inferencial

El problema usual de la **estadística inferencial** es:

- Queremos conocer el valor de una característica en una población
- No podemos medir esta característica en todos los individuos de la población
- Extraemos una muestra aleatoria de la población, medimos la característica en los individuos de esta muestra e **inferimos** el valor de la característica para la toda la población
  - ¿Cómo lo tenemos que hacer?
  - ¿Cómo tenemos que hacer la muestra?
  - ¿Qué información podemos inferir?

Idea intuitiva de muestra aleatoria simple

Muestra aleatoria simple (m.a.s.) de tamaño n: de una población de N individuos, repetimos n veces el proceso consistente en escoger **equiprobablemente** un individuo de la población; *los individuos escogidos se pueden repetir*

Ejemplo

Escogemos al azar n estudiantes de la Universidad de las Islas Baleares (UIB) (con reposición) para medirles la estatura

De esta manera, todas las muestras posibles de n individuos (posiblemente repetidos: *multiconjuntos*) tienen la misma probabilidad

Idea intuitiva de estadístico

Estadístico (*Estimador puntual*): una función que aplicada a una muestra nos permite *estimar* un valor que queramos conocer sobre toda la población.

Ejemplo

La media de las estaturas de una muestra de estudiantes de la UIB nos permite estimar la media de las alturas de todos los estudiantes de la UIB.

Definición formal de muestra aleatoria simple

Una m.a.s. de tamaño n (de una v.a. X) es

- un conjunto de n copias independientes de X, o
- un conjunto de n variables aleatorias independientes $X_1, \ldots, X_n$, todas con la distribución de X.

Ejemplo

Sea X la v.a. "escogemos un estudiante de la UIB y le medimos la altura". Una m.a.s. de X de tamaño n serán n copias independientes $X_1, \ldots, X_n$ de esta X.

Una realización de una m.a.s. son los n valores $x_1, \ldots, x_n$ que toman las v.a. $X_1, \ldots, X_n$.

Definición formal de estadístico

Un *estadístico* T es una función aplicada a la muestra $X_1, \ldots, X_n$:

$$T = f(X_1, \ldots, X_n)$$

Este estadístico se aplica a las realizaciones de la muestra

Definición: la **media muestral** de una m.a.s. $X_1, \ldots, X_n$ de tamaño n es

$$\overline{X} := \frac{X_1 + \cdots + X_n}{n}$$

y estima $E(X)$.

Ejemplo:

La **media muestral** de las alturas de una realización de una m.a.s. de las alturas de estudiantes estima la altura media de un estudiante de la UIB.

Definición formal de estadístico

Así pues, un **estadístico** es una (otra) variable aleatoria, con distribución, esperanza, etc.

La **distribución muestral** de T es la distribución de esta variable aleatoria.

Estudiando esta distribución muestral, podremos estimar propiedades de X a partir del comportamiento de una muestra.

Error estándar de T: desviación típica de T.

Convenio: LOS ESTADÍSTICOS, EN MAYÚSCULAS; las realizaciones, en minúsculas

- $X_1, \ldots, X_n$ una m.a.s. y
$$\overline{X} := \frac{X_1 + \cdots + X_n}{n},$$
el estadístico media muestral.
- $x_1, \ldots, x_n$ una realización de esta m.a.s. y
$$\overline{x} := \frac{x_1 + \cdots + x_n}{n},$$
la media (muestral) de esta realización.

En la vida real...

En la vida real, las muestras aleatorias se toman, casi siempre, sin reposición (es decir sin repetición del mismo individuo de la población).

No son muestras aleatorias simples. pero:

- Si N es mucho más grande que n, los resultados para una m.a.s. son (aproximadamente) los mismos, ya que las repeticiones son improbables y las variables aleatorias que forman la muestra son prácticamente independientes.
- En estos casos cometeremos el abuso de lenguaje de decir que es una m.a.s.
- Si n es relativamente grande, se suelen dar versiones corregidas de los estadísticos.

La media muestral

Definición de media muestral

Media muestral : sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X

La *media muestral* es:

$$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Proposición

En estas condiciones,

$$E(\overline{X}) = \mu_X, \quad \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

donde $\sigma_{\overline{X}}$ es el **error estándar** de $\overline{X}$.

Propiedades de la media muestral

Proposición

- Es un estimador puntual de μ_X
- $E(\overline{X}) = \mu_X$: el valor esperado de $\overline{X}$ es μ_X.
- Si tomamos muchas veces una m.a.s. y calculamos la media muestral, el valor medio de estas medias tiende con mucha probabilidad a ser μ_X.
- $\sigma_{\overline{X}} = \sigma_X/\sqrt{n}$: la variabilidad de los resultados de $\overline{X}$ tiende a 0 a medida que tomamos muestras más grandes.

Media muestral. Ejemplo del `iris`

Ejercicio

Consideremos la tabla de datos `iris` (que ya vimos en el tema de *Muestreo*). Vamos a comprobar las propiedades anteriores sobre la variable **longitud del pétalo** (`Petal.Length`).

1. Generaremos 10000 muestras de tamaño 40 con reposición de las longitudes del pétalo.
2. A continuación hallaremos los valores medios de cada muestra.
3. Consideraremos la media y la desviación típica de dichos valores medios y los compararemos con los valores exactos dados por las propiedades de la media muestral.

Media muestral. Ejemplo del `iris`

Para generar los valores medios de las longitudes del pétalo de las 10000 muestras usaremos la función `replicate` de `R`. Fijaos en su sintaxis:

- `replicate(n,expresión)` evalúa `n` veces la `expresión`, y organiza los resultados como las columnas de una matriz (o un vector, si el resultado de cada `expresión` es unidimensional).

```
set.seed(1001)
valores.medios.long.pétalo=replicate(10000,mean(sample(iris
                                                        repla
```

Los valores medios de las 10 primeras muestras anteriores serían

```
##  [1] 3.5975 3.5150 3.9400 3.2650 3.9125 3.9650 4.2825 3.
```

Media muestral. Ejemplo del `iris`

El valor medio de los valores medios de las muestras anteriores vale:

```
mean(valores.medios.long.pétalo)
```

```
## [1] 3.754478
```

Dicho valor tiene que estar cerca del valor medio de la variable longitud del pétalo:

```
mean(iris$Petal.Length)
```

```
## [1] 3.758
```

Fijaos que los dos valores están muy próximos.

Media muestral. Ejemplo del `iris`

La desviación típica de los valores medios de las muestras vale:

```
sd(valores.medios.long.pétalo)
```

```
## [1] 0.2796513
```

Dicho valor tiene que estar cerca de $\frac{\sigma_{lp}}{\sqrt{40}}$ (donde σ_{lp} es la desviación típica de la variable longitud del pétalo) tal como predice la propiedad de la media muestral referida a la desviación típica de la misma:

```
sd(iris$Petal.Length)/sqrt(40)
```

```
## [1] 0.2791182
```

Fijaos también en que los dos valores están muy próximos.

Poblaciones normales

Combinación lineal de distribuciones normales

Proposición La combinación lineal de distribuciones normales es normal. Es decir, si $Y_1, \ldots, Y_n$ son v.a. normales independientes, cada $Y_i \sim N(\mu_i, \sigma_i)$, y $a_1, \ldots, a_n, b \in \mathbb{R}$ entonces

$$Y = a_1 Y_1 + \cdots + a_n Y_n + b$$

es una v.a. $N(\mu, \sigma)$ con μ y σ las que correspondan:

- $E(Y) = a_1 \cdot \mu_1 + \cdots + a_n \cdot \mu_n + b$
- $\sigma(Y)^2 = a_1^2 \cdot \sigma_1^2 + \cdots + a_n^2 \cdot \sigma_n^2$

Distribución de la media muestral

Veamos cómo se distribuye la media muestral en el caso en que la población X sea normal.

Proposición

Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X de esperanza μ_X y desviación típica σ_X.

Si X es $N(\mu_X, \sigma_X)$, entonces

$$\overline{X} \text{ es } N\Big(\mu_X, \frac{\sigma_X}{\sqrt{n}}\Big)$$

y por lo tanto

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \text{ es } N(0, 1)$$

Z es la **expresión tipificada** de la media muestral.

Teorema Central del Límite

Teorema Central del Límite. Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X **cualquiera** de esperanza μ_X y desviación típica σ_X. Cuando $n \to \infty$,

$$\overline{X} \to N\Big(\mu_X, \frac{\sigma_X}{\sqrt{n}}\Big)$$

y por lo tanto

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \to N(0,1)$$

(estas convergencias se refieren a las distribuciones.)

Teorema Central del Límite

Caso n grande: Si n es grande ($n \geq 30$ **o 40**), $\overline{X}$ es aproximadamente normal, con esperanza μ_X y desviación típica $\dfrac{\sigma_X}{\sqrt{n}}$

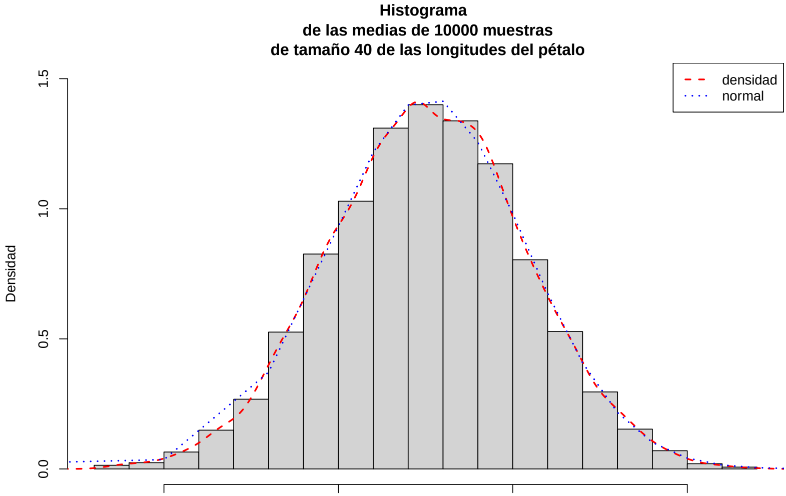
Ejemplo

Tenemos una v.a. X de media $\mu_X = 3$ y desviación típica. $\sigma_X = 0.2$. Tomamos muestras aleatorias simples de tamaño 50. La distribución de la media muestral $\overline{X}$ es aproximadamente

$$N\left(3, \frac{0.2}{\sqrt{50}}\right) = N(3, 0.0283).$$

Teorema Central del Límite

En el gráfico siguiente podemos observar el histograma de los valores medios de las longitudes del pétalo de las 10000 muestras junto con la distribución normal correspondiente:



Ejemplo

Ejercicio

El tamaño en megabytes (MB) de un tipo de imágenes comprimidas tiene un valor medio de 115 MB, con una desviación típica de 25. Tomamos una m.a.s. de 100 imágenes de este tipo.

¿Cuál es la probabilidad de que la media muestral del tamaño de los ficheros sea ≤ 110 MB?

Sea X la variable aleatoria que nos da el tamaño en megabytes del tipo de imágenes comprimidas. La distribución de X será $X = N(\mu = 115, \sigma = 25)$

Sea $X_1, \ldots, X_{100}$ la m.a.s. La distribución aproximada de la media muestral $\overline{X}$ usando el **Teorema Central del Límite** será: $\overline{X} \approx N\left(\mu_{\overline{X}} = 115, \sigma_{\overline{X}} = \frac{25}{\sqrt{100}} = 2.5\right)$.

Nos piden la probabilidad siguiente: $P(\overline{X} \leq 110)$. Si estandarizamos:

Media muestral en muestras sin reposición

Sea $X_1, \ldots, X_n$ una m.a. **sin reposición** de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X.

Si n es pequeño en relación al tamaño N de la población, todo lo que hemos contado funciona (aproximadamente).

Si n es grande en relación a N, entonces

$$E(\overline{X}) = \mu_X, \quad \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

(**factor de población finita**)

El Teorema Central del Límite ya no funciona exactamente en este último caso.

Proporción muestral

Proporción muestral. Definición

Proporción muestral. Sea X una v.a. Bernoulli de parámetro p_X (1 éxito, 0 fracaso). Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de X.

$S = \sum_{i=1}^{n} X_i$ es el nombre de éxitos observados es $B(n, p)$.

La **proporción muestral** es

$$\widehat{p}_X = \frac{S}{n}$$

y es un estimador de p_X.

Notemos que $\widehat{p}_X$ es un caso particular de $\overline{X}$, por lo que todo lo que hemos dicho para medias muestrales es cierto para proporciones muestrales.

Proporción muestral. Propiedades

Proposición

- Valor esperado de la proporción muestral:

$$E(\widehat{p}_X) = p_X$$

- **Error estándar** de la proporción muestral:

$$\sigma_{\widehat{p}_X} = \sqrt{\frac{p_X(1-p_X)}{n}}$$

- Si la muestra es sin reposición y n es relativamente grande,

$$\sigma_{\widehat{p}_X} = \sqrt{\frac{p_X(1-p_X)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}.$$

Proporción muestral. Propiedades

Teorema: Si n es grande ($n \geq 30$ o 40) y la muestra es aleatoria simple, usando el Teorema Central del Límite,

$$\frac{\widehat{p}_X - p_X}{\sqrt{\frac{p_X(1-p_X)}{n}}} \approx N(0,1)$$

Proporción muestral. Ejemplo del `iris`

Ejercicio

Dada una muestra de 60 flores de la tabla de datos `iris`,

1. Estimar la proporción de flores de la especie `setosa`.
2. Estimar también la desviación estándar de dicha proporción.

Primero generamos la muestra de las 60 flores:

```
set.seed(1000)
flores.elegidas = sample(1:150,60,replace=TRUE)
muestra.flores = iris[flores.elegidas,]
```

Proporción muestral. Ejemplo del `iris`

A continuación miramos cuántas flores de la muestra son de la especie `setosa`:

```
table(muestra.flores$Species=="setosa")
```

```
##
## FALSE  TRUE
##    39    21
```

Tenemos entonces 21 flores de la especie `setosa`.

Proporción muestral. Ejemplo del `iris`

La estimación de la proporción de flores de especie `setosa` será:

```
(prop.setosa = table(muestra.flores$Species=="setosa")[2]/1
```

```
## TRUE
## 0.35
```

valor que no está muy lejos del valor poblacional de la proporción p_{setosa} que es $p_{setosa} = \frac{50}{150} = 0.3333$.

Para estimar la desviación estándar de la proporción muestral de flores de tamaño 60 de la especie `setosa`, repetiremos el experimento anterior 10000 veces y hallaremos la desviación estándar de las proporciones obtenidas. Al final, compararemos dicho valor con el valor exacto dado por la propiedad correspondiente.

Proporción muestral. Ejemplo del `iris`

Para generar las proporciones de las 10000 muestras usaremos la función `replicate` de `R`:

```
set.seed(1002)
props.muestrales = replicate(10000,table(sample(iris$Specie
                                  replace=TRUE)=="setosa")[2]/6
```

La desviación típica de las proporciones muestrales anteriores vale:

```
sd(props.muestrales)
```

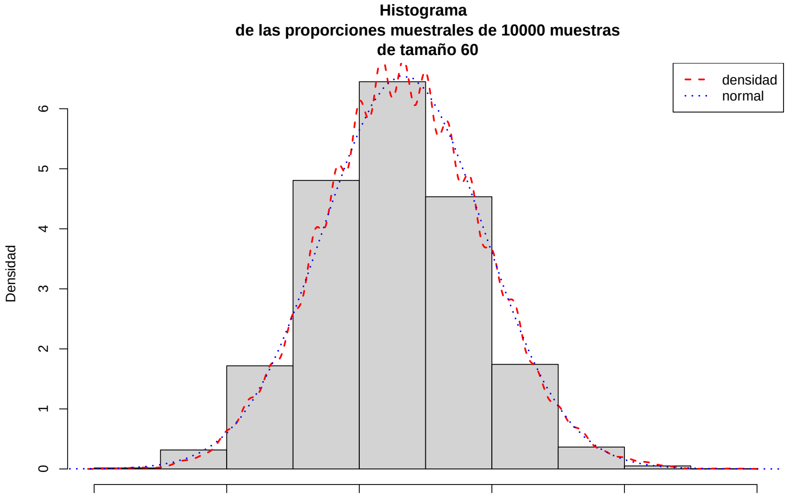
```
## [1] 0.06021098
```

valor muy próximo al valor real que vale:

$$\sigma_{\widehat{p}_X} = \sqrt{\frac{p_X(1-p_X)}{n}} = \sqrt{\frac{\frac{50}{150}\cdot\left(1-\frac{50}{150}\right)}{60}} = 0.0609.$$

Proporción muestral. Ejemplo del `iris`

En el gráfico siguiente podemos observar el histograma de las proporciones muestrales de las 10000 muestras junto con la distribución normal correspondiente:



Varianza muestral y desviación típica muestral

Varianza muestral y desviación típica muestral. Definición

Varianza muestral y desviación típica muestral. Sea $X_1, \ldots, X_n$ una m.a.s. de tamaño n de una v.a. X de esperanza μ_X y desviación típica σ_X.

La **varianza muestral** es

$$\widetilde{S}_X^2 = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

La **desviación típica muestral** es

$$\widetilde{S}_X = +\sqrt{\widetilde{S}_X^2}$$

Varianza muestral y desviación típica muestral. Definición

Además, escribiremos

$$S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} = \frac{(n-1)}{n}\widetilde{S}_X^2 \quad \text{y} \quad S_X = +\sqrt{S_X^2}$$

Varianza muestral y desviación típica muestral. Propiedades

Propiedades

$$S_X^2 = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n} = \left(\frac{\sum\limits_{i=1}^{n} X_i^2}{n} - \overline{X}^2 \right)$$

$$\widetilde{S}_X^2 = \frac{n}{n-1} \left(\frac{\sum\limits_{i=1}^{n} X_i^2}{n} - \overline{X}^2 \right)$$

Varianza muestral y desviación típica muestral. Propiedades

Teorema. Si la v.a. X es normal, entonces $E(\widetilde{S}_X^2) = \sigma_X^2$ y la v.a.

$$\frac{(n-1)\widetilde{S}_X^2}{\sigma_X^2}$$

tiene distribución χ_{n-1}^2.

La distribución χ^2_{n-1}

Distribución χ^2_n

La distribución χ^2_n (χ: en catalán, **khi**; en castellano, **ji**; en inglés, **chi**), donde n es un parámetro llamado **grados de libertad**:

es la de

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

donde $Z_1, Z_2, \ldots, Z_n$ son v.a. independientes $N(0,1)$.

La distribución χ^2_{n-1}. Propiedades

Propiedades χ^2_n

- Su función de densidad es:

$$f_{\chi^2_n}(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}, \quad \text{si } x \geq 0$$

  donde $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$, si $x > 0$.

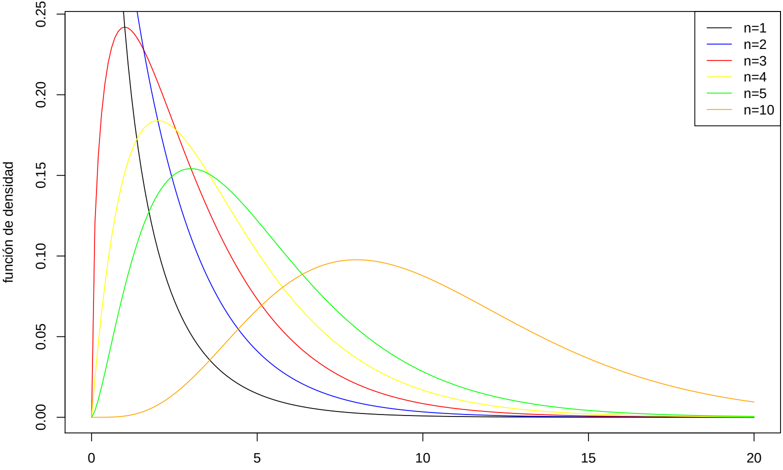
- Si $X_{\chi^2_n}$ es una v.a. con distribución χ^2_n,

$$E\left(X_{\chi^2_n}\right) = n, \quad Var\left(X_{\chi^2_n}\right) = 2n$$

- χ^2_n se aproxima a una distribución normal $N\left(n, \sqrt{2n}\right)$ para n grande ($n > 40$ o 50).

La distribución χ^2_{n-1}. Gráficos

El gráfico de la función de densidad de distintas distribuciones χ^2_n para $n = 1, 2, 3, 4, 5, 10$ se puede observar en el gráfico siguiente:



La distribución χ^2_{n-1}. Ejemplo

Ejercicio

Supongamos que el aumento diario del la ocupación de una granja de discos duros medido en Gigas sigue una distribución normal con desviación típica 1.7. Se toma una muestras de 12 discos. Supongamos que esta muestra es pequeña respecto del total de la población de la granja de discos.

¿Cual es la probabilidad de que la desviación típica muestral sea ≤ 2.5?

Sea X= aumento diario en Gigas de un disco duro elegido al azar.

Sabemos que $\sigma_X^2 = (1.7)^2 = 2.89$.

Como que X es normal y $n = 12$, tenemos que

$$\frac{11 \cdot \widetilde{S}_X^2}{2.89} = \frac{(n-1)\widetilde{S}_X^2}{\sigma_X^2} \sim \chi^2_{11}$$

La distribución χ^2_{n-1}. Ejemplo

Nos piden: $P(\widetilde{S}_X < 2.5) = P\left(\widetilde{S}_X^2 < 2.5^2\right)$:

$$P\left(\widetilde{S}_X^2 < 2.5^2\right) = P\left(\frac{11 \cdot \widetilde{S}_X^2}{2.89} < \frac{11 \cdot 2.5^2}{2.89}\right) = P(\chi^2_{11} < 2.89) = 0.0079.$$

Propiedades de los estimadores

Estimadores insesgados

¿Cuándo un estimador es bueno?

Estimadores insesgados Un estimador puntual $\widehat{\theta}$ de un parámetro poblacional θ es **insesgado, no sesgado o sin sesgo** cuando su valor esperado es precisamente el valor del parámetro:

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

Entonces se dice que el estimador puntual es **no sesgado**.

El **sesgo** de $\widehat{\theta}$ es la diferencia

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

Estimadores insesgados. Ejemplos

Proposición

- $\overline{X}$ es estimador no sesgado de μ_X: $E(\overline{X}) = \mu_X$.
- $\widehat{p}_X$ es estimador no sesgado de p_X: $E(\widehat{p}_X) = p_X$.
- Si X es normal: $\widetilde{S}_X^2$ es estimador no sesgado de σ_X^2: $E(\widetilde{S}_X^2) = \sigma_X^2$
- Si X es normal: $E(S_X^2) = \dfrac{n-1}{n}\sigma_X^2$. Por lo tanto S_X^2, es sesgado, con sesgo

$$E(S_X^2) - \sigma_X^2 = \frac{n-1}{n}\sigma_X^2 - \sigma_X^2 = -\frac{\sigma_X^2}{n} \text{ que tiende a 0.}$$

Estimadores eficientes

¿Cuando un estimador es **bueno**?

Cuando es no segado y tiene poca variabilidad (así es más probable que aplicado a una m.a.s. dé un valor más cercano al valor esperado)

Error estándar de un estimador $\widehat{\theta}$: es su desviación típica

$$\sigma_{\widehat{\theta}} = \sqrt{Var(\widehat{\theta})}$$

Estimadores eficientes

Eficiencia de un estimador Dados dos estimadores $\widehat{\theta}_1$, $\widehat{\theta}_2$ no sesgados (o con sesgo que tiende a 0) del mismo parámetro θ, diremos que $\widehat{\theta}_1$ es **más eficiente** que $\widehat{\theta}_2$ cuando

$$\sigma_{\widehat{\theta}_1} < \sigma_{\widehat{\theta}_2},$$

es decir, cuando

$$Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2).$$

Estimadores eficientes. Ejemplo

Sea X una v.a. con media μ_X y desviación típica σ_X

Consideremos la mediana $Me = Q_{0.5}$ de la realización de una m.a.s. de X como estimador puntual de μ_X

Si X es normal,

$$\begin{aligned}
& E(Me) = \mu_X, \\
& Var(Me) \approx \frac{\pi}{2}\frac{\sigma_X^2}{n} \approx \frac{1.57\sigma_X^2}{n} = 1.57 \cdot Var(\overline{X}) > Var(\overline{X}).
\end{aligned}$$

Por lo tanto, Me es un estimador no sesgado de μ_X, pero menos eficiente que $\overline{X}$.

Estimadores eficientes

Proposición

- Si la población es normal, la **media muestral** $\bar{X}$ es el estimador no sesgado más eficiente de la **media poblacional** μ_X.
- Si la población es Bernoulli, la **proporción muestral** $\hat{p}_X$ es el estimador no sesgado más eficiente de la **proporción poblacional** p_X.
- Si la población es normal, la **varianza muestral** $\tilde{S}_X^2$ es el estimador no sesgado más eficiente de la **varianza poblacional** σ_x^2.

Estimadores eficientes. ¿S_X^2 o $\widetilde{S}_X^2$?

Como hemos visto, si la población es normal, la varianza muestral es el estimador no sesgado más eficiente de la varianza poblacional

El estimador *varianza*

$$S_X^2 = \frac{(n-1)}{n}\widetilde{S}_X^2$$

aunque sea más eficiente, tiene sesgo que tiende a 0.

Si n es pequeño (≤ 30 o 40), es mejor utilizar la varianza muestral $\widetilde{S}_X^2$ para estimar la varianza, ya que el sesgo influye, pero si n es grande, el sesgo ya no es tan importante y se puede utilizar S_X^2.

Estimadores eficientes. Ejemplo

Tenemos una población numerada $1, 2, \ldots, N$

Tomamos una m.a.s. $x_1, \ldots, x_n$; sea $m = \max\{x_1, \ldots, x_n\}$.

Teorema. El estimador no segado más eficiente del tamaño de la población N es

$$\widehat{N} = m + \frac{m - n}{n}.$$

O sea, la manera más eficiente de estimar el número de elementos de la población a partir de una muestra es usar la fórmula anterior.

Ejemplo

Sentados en una terraza de un bar del Paseo Marítimo de Palma hemos anotado el número de licencia de los 40 primeros taxis que hemos visto pasar:

```
taxis=c(1217,600,883,1026,150,715,297,137,508,134,38,961,53
        314,1121,823,158,940,99,977,286,1006,1207,264,1183,
        498,606,566,1239,860,114,701,381,836,561,494,858,18
```

Supondremos que estas observaciones son una m.a.s. de los taxis de Palma. Vamos a estimar el número total de taxis.

Entonces, estimamos que el número de taxis de Palma es

```
(N=max(taxis)+(max(taxis)-length(taxis))/length(taxis))
```

```
## [1] 1268.975
```

En realidad, hay 1246.

Estimadores máximo verosímiles

¿Cómo encontramos buenos estimadores?

Antes de explicar la metodología, necesitamos una definición previa:

Función de verosimilitud de la muestra. Sea X una v.a. **discreta** con función de probabilidad $f_X(x;\theta)$ que depende de un parámetro desconocido θ.

Sea $X_1, \ldots X_n$ una m.a.s. de X, y sea $x_1, x_2, \ldots, x_n$ una realización de esta muestra.

La **función de verosimilitud** de la muestra es la probabilidad condicionada siguiente:

$$\begin{aligned} L(\theta|x_1, x_2, \ldots, x_n) \quad &:= P(x_1, x_2, \ldots, x_n|\theta) = P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= f_X(x_1;\theta) \cdots f_X(x_n;\theta). \end{aligned}$$

Estimadores máximo verosímiles

Dada la función de verosimilitud $L(\theta|x_1, \ldots, x_n)$ de la muestra, indicaremos por $\hat{\theta}(x_1, \ldots, x_n)$ el valor del parámetro θ en el que se alcanza el máximo de $L(\theta|x_1, \ldots, x_n)$. Será una función de $x_1, \ldots, x_n$.

Estimador máximo verosímil. Un estimador $\hat{\theta}$ de un parámetro θ es **máximo verosímil** (**MV**) cuando, para cada m.a.s, la probabilidad de observarlo es máxima, cuando el parámetro toma el valor del estimador aplicado a la muestra, es decir, si la función de verosimilitud

$$L(\theta|x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_n|\theta)$$

alcanza su máximo.

Estimadores máximo verosímiles. Ejemplo

Supongamos que tenemos una v.a. Bernoulli X de probabilidad de éxito p desconocida.

Para cada m.a.s. $x_1, \ldots, x_n$ de X, sean $\widehat{p}_x$ su proporción muestral y $P(x_1, \ldots, x_n \mid p)$ la probabilidad de obtenerla cuando el verdadero valor del parámetro es p.

Teorema. El valor de p para el que $P(x_1, \ldots, x_n \mid p)$ es máximo es $\widehat{p}_x$.

Dicho en otras palabras, la proporción muestral es un estimador MV de p.

Ejercicio

Demostrar el teorema anterior.

Estimadores máximo verosímiles. Ejemplo

En general, al ser ln una función creciente, en lugar de maximizar $L(\theta|x_1,\ldots,x_n)$, maximizamos

$$\ln(L(\theta|x_1,\ldots,x_n))$$

que suele ser más simple (ya que transforma los productos en sumas, y es más fácil derivar estas últimas).

Estimadores máximo verosímiles. Ejemplo

Sea $X_1, \ldots X_n$ una m.a.s. de una v.a. Bernoulli X de parámetro p (desconocido). Denotemos $q = 1 - p$

$$f_X(1; p) = P(X = 1) = p, \quad f_X(0; p) = P(X = 0) = q$$

es a decir, para $x \in \{0, 1\}$, resulta que
$f_X(x; p) = P(X = x) = p^x q^{1-x}$.

La función de verosimilitud es:

$$L(p|x_1, \ldots, x_n) = f_X(x_1; p) \cdots f_X(x_n; p) = p^{x_1} q^{1-x_1} \cdots p^{x_n} q^{1-x_n}$$
$$= p^{\sum_{i=1}^n x_i} q^{\sum_{i=1}^n (1-x_i)} = p^{\sum_{i=1}^n x_i} q^{n-\sum_{i=1}^n x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}$$

Estimadores máximo verosímiles. Ejemplo

La función de verosimilitud es

$$L(p|x_1,\ldots,x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = p^{n\overline{x}}(1-p)^{n-n\overline{x}},$$

donde $\overline{x} = \dfrac{\sum_{i=1}^n x_i}{n}$.

Queremos encontrar el valor de p en el que se alcanza el máximo de esta función (donde $\overline{x}$ es un parámetro y la variable es p)

Maximizaremos su logaritmo:

$$\ln(L(p|x_1,\ldots,x_n)) = n\overline{x}\ln(p) + n(1-\overline{x})\ln(1-p).$$

Estimadores máximo verosímiles. Ejemplo

Derivamos respecto de p:

$$\begin{aligned}\ln(L(p|x_1,\ldots,x_n))' &= n\overline{x}\tfrac{1}{p} - n(1-\overline{x})\tfrac{1}{1-p} \\ &= \tfrac{1}{p(1-p)}\Big((1-p)n\overline{x} - pn(1-\overline{x})\Big) = \tfrac{1}{p(1-p)}(n\overline{x} - pn) \\ &= \tfrac{n}{p(1-p)}(\overline{x} - p)\end{aligned}$$

Estudiamos el signo:

$$\ln(L(p|x_1,\ldots,x_n))' \geq 0 \Leftrightarrow \overline{x} - p \geq 0 \Leftrightarrow p \leq \overline{x}$$

Estimadores máximo verosímiles. Ejemplo

Por lo tanto

$$\ln(L(p|x_1,\ldots,x_n)) \begin{cases} \text{creciente hasta } \overline{x} \\ \text{decreciente a partir de } \overline{x} \\ \text{tiene un máximo en } \overline{x} \end{cases}$$

El resultado queda demostrado. $L(\widehat{p}_X|x_1,\ldots,x_n) \geq L(p|x_1,\ldots,x_n)$ para cualquier p.

Algunos estimadores **MV**

Proposición

- $\widehat{p}_x$ es el estimador MV del parámetro p de una v.a. Bernoulli.
- $\overline{X}$ es el estimador MV del parámetro θ de una v.a. Poisson.
- $\overline{X}$ es el estimador MV del parámetro μ de una v.a. normal.
- S_X^2 (**no** $\widetilde{S}_X^2$) es el estimador MV del parámetro σ^2 de una v.a. normal.
- El máximo (**no** $\widehat{N}$) es el estimador MV de la N en el problema de los taxis.

Estimadores máximo verosímiles. Ejemplo: captura-recaptura

En una población hay N individuos, capturamos K, los marcamos y los volvemos a soltar.

Ahora volvemos a capturar n, de los que k están marcados. A partir de estos datos, queremos estimar N.

Supongamos que N y K no han cambiado de la primera a la segunda captura.

La variable aleatoria $X =$ *Un individuo esté marcado* es $Be(p)$ con $p = \dfrac{K}{N}$.

Si $X_1, \ldots, X_n$ es la muestra capturada la segunda vez, entonces $\widehat{p}_X = \frac{k}{n}$.

Estimadores máximo verosímiles. Ejemplo: captura-recaptura

$\widehat{p}_X$ es un estimador máximo verosímil p. Por tanto, estimamos que:

$$\frac{K}{N} = \frac{k}{n} \Rightarrow N = \frac{n \cdot K}{k}$$

Por lo tanto, el estimador

$$\widehat{N} = \frac{n \cdot K}{k}$$

maximiza la probabilidad de la observación k *marcados de* n *capturados*, por lo que $\hat{N}$ es el **estimador máximo verosímil** de N.

Estimadores máximo verosímiles. Ejemplo: captura-recaptura

Ejercicio

Supongamos que hemos marcado 15 peces del lago, y que en una captura, de 10 peces, hay 4 marcados. ¿Cuántos peces estimamos que contiene el lago?

El número de peces estimados del lago será:

$$\widehat{N} = \frac{15 \cdot 10}{4} = 37.5$$

Estimamos que habrá entre 37 y 38 peces en el lago.

Estimadores máximo verosímiles. Ejemplo: captura-recaptura

El estimador
$$\widehat{N} = \frac{n \cdot K}{k}$$
es sesgado, con sesgo que tiende a 0.

El **estimador de Chapman**
$$\widehat{N} = \frac{(n+1) \cdot (K+1)}{k+1} - 1,$$
es menos segado para muestras pequeñas, y no sesgado si $K + n \geq N$ (pero no máximo verosímil).

Estimación puntual con R

La función `fitdistr`

Para obtener estimaciones puntuales con `R` hay que usar la función `fitdistr` del paquete **MASS**:

```
fitdistr(x, densfun=..., start=...)
```

donde

- `x` es la muestra, un vector numérico.
- El valor de `densfun` ha de ser el nombre de la familia de distribuciones:`"chi-squared"`, `"exponential"`, `"f"`, `"geometric"`, `"lognormal"`, `"normal"` y `"poisson"`.

La función `fitdistr`

- Si `fitdistr` no dispone de una fórmula cerrada para el estimador máximo verosímil de algún parámetro, usa un algoritmo numérico para aproximarlo que requiere de un valor inicial para arrancar. Este valor (o valores) se puede especificar igualando el parámetro `start` a una `list` con cada parámetro a estimar igualado a un valor inicial.

Ejemplos de uso de `fitdistr`. Estimación del parámetro λ de una variable de Poisson

Ejemplo

Consideramos la muestra siguiente de tamaño 50 de una variable de Poisson de parámetro $\lambda = 5$:

```
set.seed(98)
muestra.poisson = rpois(50,lambda=5)
muestra.poisson
```

```
##  [1]  5  4  4  5  3  4  1  4  6  3  7  7  3  5  4  8  4
## [26]  7  5  2  8  3  5  4  1  5  6  4  7  7  3  4  6 10
```

Vamos a estimar el valor del parámetro λ a partir de la muestra anterior.

Ejemplos de uso de `fitdistr`

Para estimar λ usamos la función `fitdistr`:

```
library(MASS)
fitdistr(muestra.poisson, densfun = "poisson")
```

```
##      lambda
##   4.760000
##  (0.308545)
```

La función `fitdistr` nos ha dado el siguiente valor de λ: 4.76, valor que se aproxima al valor real de $\lambda = 5$, con un error típico de 0.308545.

Ejemplos de uso de `fitdistr`

Recordemos que el estimador máximo verosímil de λ es $\overline{X}$ con error típico $\frac{\sqrt{\lambda}}{\sqrt{n}}$. Veamos si la función `fitdistr` nos ha mentido:

```
(estimación.lambda = mean(muestra.poisson))
```

```
## [1] 4.76
```

```
(estimación.error.típico= sqrt(estimación.lambda/50))
```

```
## [1] 0.308545
```

Comprobamos que los valores anteriores coinciden con los dados por la función.

Ejemplos de uso de fitdistr

¿Qué estimaciones hubiésemos obtenido de la media μ y la desviación típica σ si suponemos que la muestra anterior es normal?

```
fitdistr(muestra.poisson,densfun = "normal")
```

```
##       mean          sd    
##   4.7600000   2.1868699 
##  (0.3092701) (0.2186870)
```

Dichos valores coinciden con la media muestral $\overline{X}$ y la desviación típica "verdadera" de la muestra considerada:

```
sd(muestra.poisson)*sqrt(49/50)
```

```
## [1] 2.18687
```

Guía rápida

- `fitdistr` del paquete **MASS**, sirve para calcular los estimadores máximo verosímiles de los parámetros de una distribución a partir de una muestra. Parámetros principales:
  - `densfun`: el nombre de la familia de distribuciones, entre comillas.
  - `start`: permite fijar el valor inicial del algoritmo numérico para calcular el estimador, si la función lo requiere.

Estimación puntual en Python

Primero haremos la carga de paquetes

```
from scipy.stats import norm
from numpy import linspace
import matplotlib.pyplot as plt
```

Elegimos una m.a.s. de tamaño 150 de una variable aleatoria $N(0, 1)$.

```
sample = norm.rvs(loc = 0, scale = 1, size = 150)
print(sample[1:10])
```

```
## [-0.74758959 -0.54061326  0.6310327   0.35470952 -0.1186
##   0.28610282 -1.2387943  -0.05551449]
```

Estimación puntual en Python

Para estimar los parámetros, usamos la distribución adecuada (en nuestro caso una normal) e invocamos el método `fit`

```
params = norm.fit(sample)
print("Media = {mu}".format(mu=params[0]))
```

```
## Media = -0.07441383975513464
```

```
print("Desviacion tipica = {sd}".format(sd=params[1]))
```

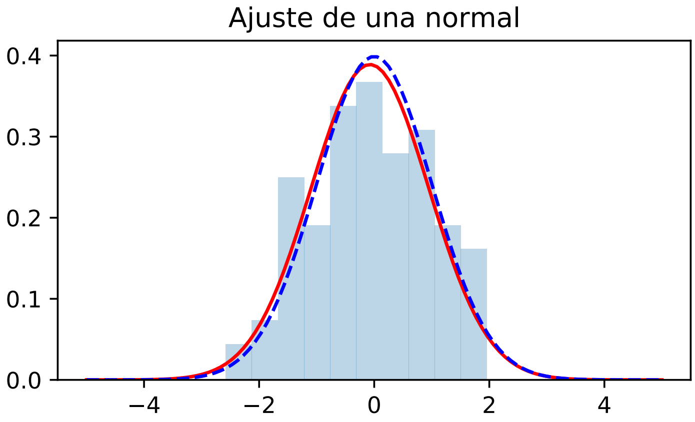
```
## Desviacion tipica = 1.0255054462844888
```

Estimación puntual en Python

Vamos a comparar la distribución con los parámetros estimados vs nuestra muestra.

```
x = linspace(-5,5,100)
pdf_fitted = norm.pdf(x, loc=params[0], scale=params[1])
pdf_original = norm.pdf(x, loc=0, scale=1)
```

Estimación puntual en `Python`



Ejercicio de distribución Rayleigh

$$f(x) = \frac{(x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma^2}$$

Importar las librerías para crear la m.a.s., generar la muestra y obtener los parámetros

```
from scipy.stats import rayleigh
sample = rayleigh.rvs(loc=5, scale=2, size=150)
params = rayleigh.fit(sample)
print("Media = {mu}".format(mu=params[0]))
```

```
## Media = 5.22435091127449
```

```
print("Desviacion tipica = {sd}".format(sd=params[1]))
```

```
## Desviacion tipica = 1.8624809484819602
```

Ejercicio de distribución Rayleigh

Generamos la distribución con los parámetros estimados

```
x = linspace(5, 15, 100)
pdf_fitted = rayleigh.pdf(x, loc=params[0], scale=params[1
pdf_original = rayleigh.pdf(x, loc=5, scale=2)
```

Ejercicio de distribución Rayleigh