

Fundamentos de Inferencia Estadística

Francisco Javier Mercader Martínez

Índice

1 Muestreo y distribuciones muestrales	3
1.1 Introducción	3
1.2 Ejemplos	3
1.3 Surge una pregunta	3
1.4 Esbozo de respuesta: tasa de participación	3
1.5 Realización del experimento: conclusiones	4
1.6 En la práctica	5
1.7 Uso de la distribución muestral	5
1.8 Antes de extraer una muestra:	6
1.9 Otro ejemplo: valores muestrales de una distribución normal	6
1.10 Un resultado importante	6
1.11 Algunos términos	6
1.12 Ejemplos de estadísticos	7
1.13 La media muestral	7
1.13.1 Esperanza y varianza de la media muestral	7
1.14 Consecuencia práctica	8
1.14.1 Analogía con una diana	8
1.15 Varianza muestral	8
1.15.1 Dos apuntes	8
1.16 Esperanza de la varianza muestral	8
1.17 Distribuciones muestrales de \bar{X} y S^2	9
1.18 Distribución de \bar{X} y S^2 para una m.a.s. de una distribución normal	9
1.19 Recordatorio: distribución χ^2 con p grados de libertad	9
1.20 Distribución t-Student	10
1.20.1 ¿Cuál es la forma de la densidad de una t-Student?	11
1.21 Distribución F de Snedecor para el cociente de varianzas	11
1.21.1 ¿Cuál es la forma de la densidad de una F de Snedecor?	11
1.22 Si la distribución de X no es Normal	12
1.23 Distribución muestral de la proporción muestral	13
1.24 Simulación y método de Monte-Carlo	14
1.24.1 Muestreo de Monte-Carlo para aproximar esperanzas	14
1.24.1.1 Ejemplos	14
1.24.1.2 Aplicaciones	14
1.25 Movimiento Browniano	15
1.26 En finanzas, el modelo de Black-Scholes	15
¿Cuándo observamos $S(t) \geq 100\text{€}$?	15
1.27 Simulación y método de Monte-Carlo	16
Generación de valores de una distribución uniforme	17
Generación de valores de una distribución gamma	17
Transformación de una variable aleatoria	18
Función características	18
Desigualdades	18
2 Estimación	20
2.1 Introducción	20
2.2 Ejemplos de estimación paramétrica	20
2.3 Estimación paramétrica: estimación puntual	20
2.4 Métodos de construcción de estimadores	21
2.5 Método de los momentos	21
2.6 Método de máxima verosimilitud	21
2.7 Estimador de máxima verosimilitud	22

2.8	Métodos para evaluar un estimador	24
2.9	Sesgo	24
2.10	Error cuadrático medio	24
2.11	Balance entre sesgo y varianza	25
2.12	Comportamiento asintótico de un estimador	25
2.13	Consistencia	25
2.14	Normalidad asintótica	25
2.15	Estimación no paramétrica	26
2.16	Estimación tipo núcleo	27
2.16.1	Podemos variar el núcleo	27
2.16.2	Se pueden demostrar resultados asintóticos	28
2.17	Ancho de banda	29
2.17.1	Elección óptima del ancho de banda	29
2.17.1.1	Aproximación gaussiana para el cálculo de $R(f'')$	29
2.17.1.2	Aproximación no paramétrica para el cálculo de $R(f'')$, implementación en R	30
2.18	Introducción al Bootstrap	31
2.18.1	El error estándar Bootstrap	33
3	Estimación por intervalos	34
3.1	Introducción	34
3.2	Nivel de confianza	34
3.2.1	Nivel de cobertura	34
3.3	Un procedimiento general de construcción	34
3.4	Intervalo de confianza para la media μ de $X \sim \mathcal{N}(\mu, \sigma^2)$, σ conocida	35

1 Muestreo y distribuciones muestrales

1.1 Introducción

i El contexto

- Tenemos una pregunta acerca de un fenómeno aleatorio.
- Formulamos un modelo para la variable de interés X .
- Traducimos la pregunta de interés en términos de uno o varios parámetros del modelo.
- Repetimos el experimento varias veces, apuntamos los valores de X .
- ¿Cómo usar estos valores para extraer información sobre el parámetro?

1.2 Ejemplos

i ¿Está la moneda trucada?

- Experimento: tirar una moneda. $X =$ resultado obtenido:

$$P(X = +) = p, P(X = c) = 1 - p \\ \text{¿}p = \frac{1}{2}?$$

i Sondeo sobre intención de participación en unas elecciones

- Queremos estimar la tasa de participación antes de unas elecciones generales.
- Formulamos un modelo:
 - Experimento: “escoger una persona al azar en el censo”.
 - X : participación, variable dicotómica (“Sí” o “No”). $p = P(X = \text{Sí})$.
- ¿Cuánto vale p ?
- Censo: aprox. 37 000 000. Escogemos aprox. 3000 personas.

i Determinación de la concentración de un producto

- Quiero determinar la concentración de un producto.
- Formulo el modelo:
 - Experimento: “llevar a cabo una mediación”.
 - X : “valor proporcionado por el aparato”.
 - $X \sim \mathcal{N}(\mu, \sigma^2)$.
- ¿Qué vale μ ?

1.3 Surge una pregunta

En todas estas situaciones donde nos basamos en la repetición de un experimento simple... - ¿Cómo sabemos que nuestra situación es fiable? - ¿Qué confianza tenemos al extrapolar los resultados de una muestra de 3000 personas a una población de 37 millones de personas?

1.4 Esbozo de respuesta: tasa de participación

i Para convencernos, un experimento de simulación

- Voy a simular el proceso de extracción de una muestra de 3000 personas en una población de 37 millones de personas.
- Construyo a mi antojo los distintos componentes:
 - **La población:** defino en mi ordenador de 37 000 000 de ceros y unos (\Leftrightarrow el censo electoral).
 - * “1” \Leftrightarrow “la persona piensa ir a votar”.
 - * “0” \Leftrightarrow “la persona no piensa ir a votar”.
 - **La tasa de participación “real”:** Decido que en mi población el 70% piensa en ir a votar \rightarrow 25 900 000 “1”s.
 - **La extracción de una muestra:** construyo un pequeño programa que extrae al azar una muestra de 3000 números dentro del conjunto grande.

```
[1] 0.7056667
```

Queremos descartar que haya sido suerte. Vamos a repetir muchas veces (1000 veces por ejemplo), la extracción de una muestra de 3000 personas en la población.

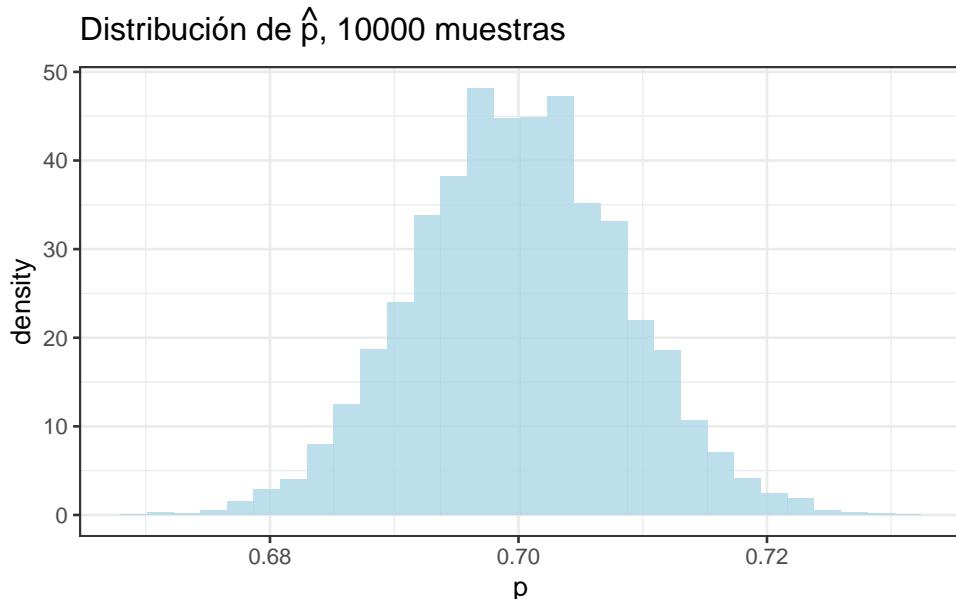
```
library(tidyverse)
lista_muestras <- replicate(
  10000,
  sample(poblacion, 3000, replace = FALSE),
  simplify = FALSE
)
p_muestras <- map_dbl(lista_muestras, mean)
head(p_muestras)
```

```
[1] 0.6970000 0.7030000 0.7036667 0.7023333 0.7013333 0.7226667
```

```
library(tidyverse)
p_muestras <- replicate(
  10000,
  sample(poblacion, 3000, replace = FALSE),
  simplify = FALSE
) |>
  map_dbl(mean)
head(p_muestras)
```

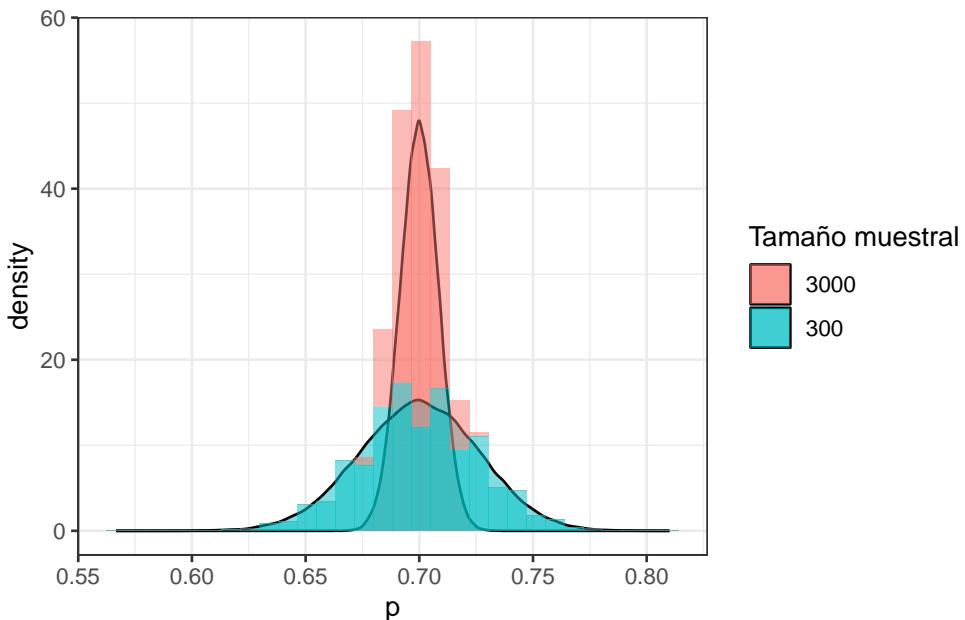
```
[1] 0.7090000 0.7080000 0.6946667 0.7216667 0.6906667 0.6990000
```

Recogemos los valores obtenidos en un histograma.



1.5 Realizaccción del experimento: conclusiones

- La enorme mayoría de las muestras de 300 individuos proporcionan una tasa de participación muy próxima a la de la población.
 - **El riesgo** de cometer un error superior a ± 2 puntos, al coger **una** muestra de 3000 individuos es muy pequeño (y asumible...)
- Si nos limitamos a muestras de 300 individuos, ¿qué esperáis?



1.6 En la práctica

i Usamos las distribuciones muestrales

- Las empresas de sondeos no se basan en simulaciones sino en cálculos teóricos.
- Experimento aleatorio: escoger al azar una muestra de 3000 personas dentro de una población de 37 000, con una tasa de participación p .
- Llamamos a \hat{p} la variable aleatoria: proporción de “1”s en la muestra escogida.
- ¿Cuál es la distribución de valores de \hat{p} ?

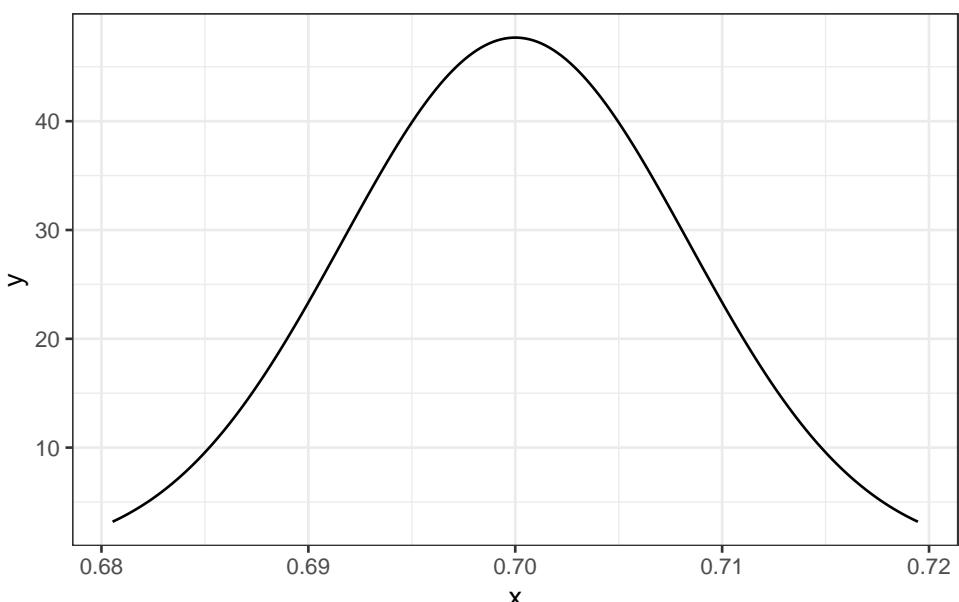
$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Es lo que llamamos la [distribución muestral](#) de \hat{p} .

1.7 Uso de la distribución muestral

i La distribución muestral de \hat{p} :

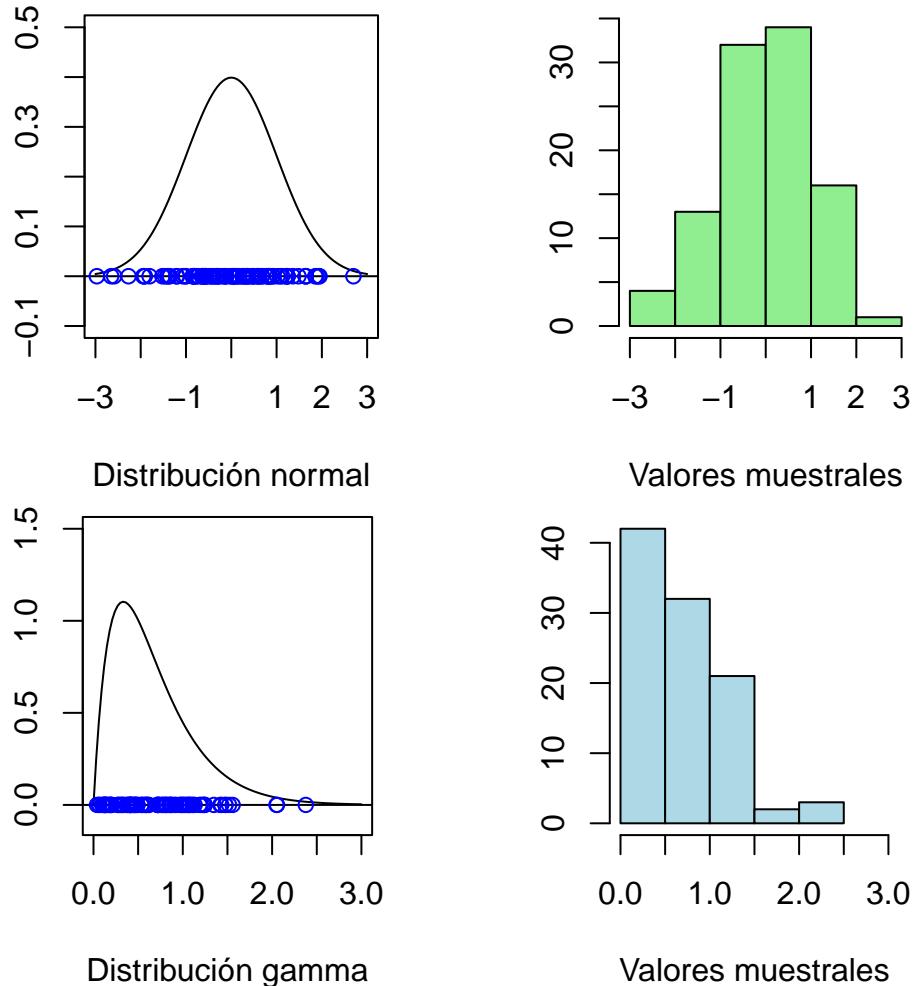
Es la distribución esperada de los valores de \hat{p} respecto a todas las muestras de ese tamaño que podría extraer



1.8 Antes de extraer una muestra:

- ¿Es suficiente el tamaño para el riesgo asumible y la predicción requerida?
- Una vez la muestra:
 - ¿Puedo dar un margen de error?
 - ¿Puedo decidir si p poblacional es, por ejemplo, mayor que un valor dado?

1.9 Otro ejemplo: valores muestrales de una distribución normal



1.10 Un resultado importante

Ley (débil) de los grandes números

Sea X una variable aleatoria y $g(X)$ una variable aleatoria transformada de X , con esperanza y momento de orden 2 finitos. Supongamos $X_1, X_2, \dots, X_n, \dots$ una sucesión de variables (vv.aa) independientes con la distribución que X , entonces

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{\sum_{i=1}^n g(X_i)}{n} - E[g(x)] \right| < \varepsilon \right] = 1, \text{ para todo } \varepsilon > 0.$$

1.11 Algunos términos

- Definición 1.11.1**
- Sea una variable aleatoria X . Consideramos n variables aleatorias independientes e idénticamente distribuidas X_1, X_2, \dots, X_n , que se distribuyen como X . La variable aleatoria multidimensional (X_1, X_2, \dots, X_n) es una **muestra aleatoria simple** (m.a.s.) de X .
 - Cualquier cantidad calculada a partir de las observaciones de una muestra: **estadístico**.
 - **Experimento aleatorio**: extraer una muestra. Consideramos un estadístico como una variable aleatoria. Nos interesa conocer la distribución del estadístico: **distribución muestral**.

1.12 Ejemplos de estadísticos

- Proporción muestral: \hat{p} .
- Media muestral: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.
- Desviación típica muestral: $S_X = \sqrt{\frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2}$

1.13 La media muestral

i Contexto

Estudiamos una variable X cuantitativa. - Estamos interesados en μ , el centro de la distribución de X . - Extraemos una muestra de tamaño n :

$$x_1, x_2, \dots, x_n.$$

- Calculamos su media \bar{x} para aproximar μ .
- ¿Cuál es la distribución muestral de \bar{X} ?

Ejemplo:

- Quiero medir una cantidad. Hay variabilidad en las mediciones.
- Introduzco una variable aleatoria X = “valor proporcionado por el aparato”.
- μ representa el centro de los valores.
- Extraigo una muestra de tamaño 5 del valor de X

1.13.1 Esperanza y varianza de la media muestral

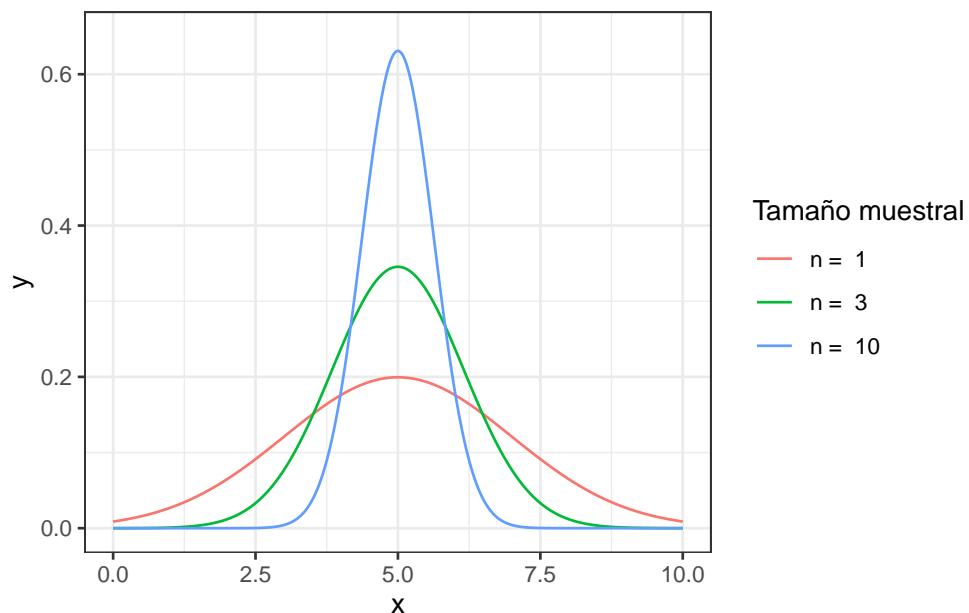
Llamamos $\mu = E[X]$ y $\sigma^2 = \text{Var}(X)$.

- Tenemos

$$E[\bar{X}] = \mu.$$

- Es decir que el centro de la distribución muestral de \bar{X} coincide con el centro de la distribución X .
- Tenemos $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$, es decir, la dispersión de la distribución muestral de \bar{X} es \sqrt{n} veces más pequeña que la dispersión inicial de X .

Ilustración: X inicial, \bar{X} con $n = 3$, \bar{X} con $n = 10$

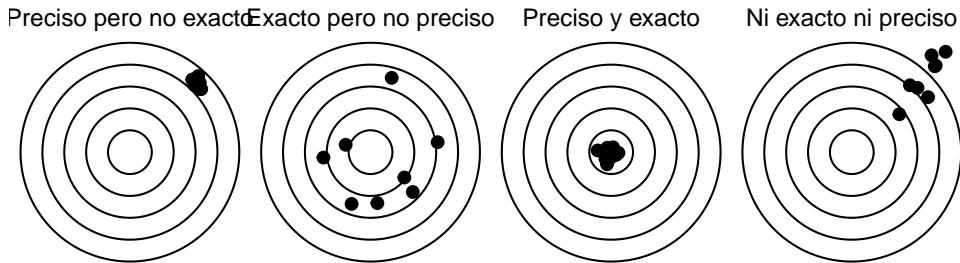


1.14 Consecuencia práctica

i Aparato de medición

- Experimento: llevar a cabo una medición con un aparato.
- Variable aleatoria X : “valor proporcionado por el aparato”.
- $E[X]$: centro de la distribución de los valores proporcionados por el aparato.
 - Lo deseable: $E[X] =$ valor exacto de la cantidad que buscamos medir.
 - En este caso, decimos: el aparato es **exacto**.
- σ_X : dispersión de la distribución de los valores proporcionados por el aparato.
 - Lo deseable: σ_X pequeño.
 - En este caso, decimos: el aparato es **preciso**.

1.14.1 Analogía con una diana



1.15 Varianza muestral

Si (X_1, X_2, \dots, X_n) es una muestra aleatoria simple de X , definimos la **varianza muestral** S_n^2 como

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

i Fórmula alternativa para S_n^2 :

$$S_n^2 = \frac{n}{n-1} (\bar{X}^2 - (\bar{X}_n)^2),$$

donde $\bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$.

1.15.1 Dos apuntes

i En algunos textos en castellano

Se suele llamar S_n^2 **cuasi-varianza muestral**, reservando el término varianza muestral para la cantidad $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

i En estas fórmulas

Omitimos, si no hay confusión posible, el subíndice n , escribiendo S^2 , $\bar{X} = \sum_{i=1}^n X_i$ y $\bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$.

1.16 Esperanza de la varianza muestral

Proposición 1.16.1 Si (X_1, X_2, \dots, X_n) es una muestra aleatoria simple de X con varianza σ_X^2 ,

$$E[S_n^2] = \sigma_X^2.$$

1.17 Distribuciones muestrales de \bar{X} y S^2

⚠️ Tened en cuenta

- Los resultados anteriores sobre $E[\bar{X}]$ y $\sigma_{\bar{X}}$ son válidos sea cual sea el modelo escogido para la distribución de X .
- Si queremos decir algo más preciso sobre la distribución de \bar{X} (densidad, etc.) necesitamos especificar la distribución de X .
- En el caso en que la variable X siga una distribución normal, el **teorema de Fisher** analiza cómo se comportan los estadísticos anteriores y nos permiten establecer una serie de consecuencias que serán utilizadas posteriormente en los temas de intervalos de confianza y de contraste de hipótesis.

1.18 Distribución de \bar{X} y S^2 para una m.a.s. de una distribución normal

💡 Teorema de Fisher

Consideramos una muestra aleatoria simple de una variable aleatoria X con distribución normal $\mathcal{N}(\mu, \sigma^2)$, entonces se verifica: 1) \bar{X}_n y S_n^2 son dos variables aleatorias independientes. 2) $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$. 3) $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$.

1.19 Recordatorio: distribución χ^2 con p grados de libertad

💡 La distribución χ^2 .

Para $p \in \mathbb{N}^+$, la función de densidad de la distribución χ^2 es igual a

$$\frac{1}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}} \cdot x^{\frac{p}{2}-1} e^{-\frac{x}{2}}, \text{ si } x > 0,$$

donde Γ denota la función Gamma (Nota: para cualquier real $\alpha > 0$, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$).

💡 Caracterización de la χ^2

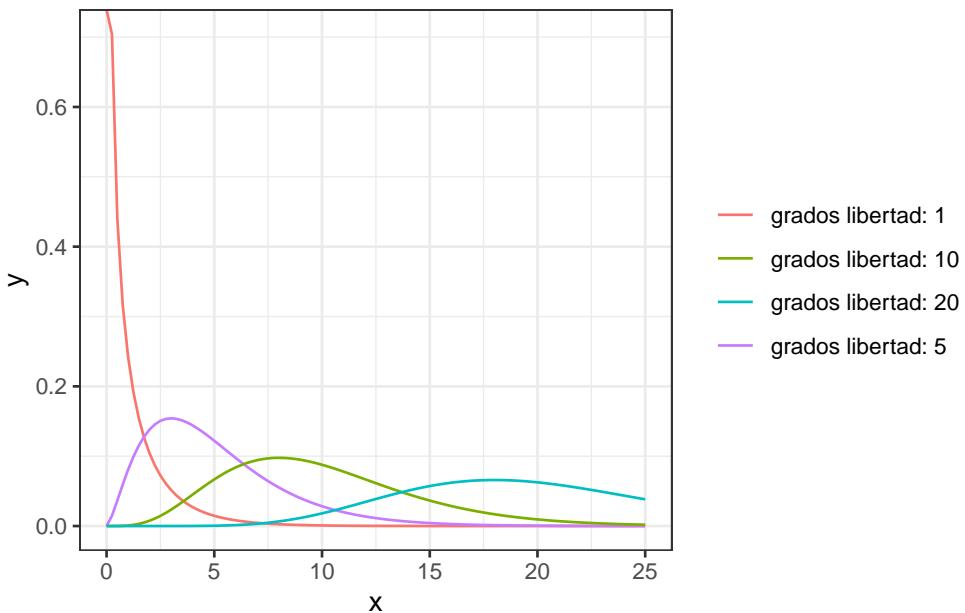
Si Z_1, \dots, Z_p son p variables aleatorias independientes, con $Z_i \sim \mathcal{N}(0, 1)$, entonces la variable aleatoria X definida como

$$X = Z_1^2 + \dots + Z_p^2 = \sum_{i=1}^p Z_i^2$$

tiene una distribución χ^2 con p grados de libertad.

❓ ¿Cómo es su función de densidad?

Depende de los grados de libertad.



1.20 Distribución t-Student

i Hemos visto, si X es Normal:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

So queremos centrarnos en μ natural sustituir en ella σ por S_n .

Proposición 1.20.1 Sea (X_1, \dots, X_n) una muestra aleatoria simple de una población $\mathcal{N}(\mu, \sigma^2)$,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene por densidad

$$f_{n-1}(t) \propto \frac{1}{\left(\frac{1+t^2}{n-1}\right)^{(n-1)/2}}, \quad -\infty < t < \infty, \quad (1)$$

La distribución que admite esta densidad se llama **distribución t-Student** con $n-1$ grados de libertad. Escribimos $T \sim t_{n-1}$.

! Su densidad

La función de densidad de una t de Student con k grados de libertad:

$$f_k(t) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+t^2/k)^{\frac{k+1}{2}}}, \quad -\infty < t < \infty,$$

donde Γ denota la función Gamma.

i Caracterización de la t de Student como cociente

Si Z e Y son dos variables aleatorias independientes, con $Z \sim \mathcal{N}(0, 1)$ e $Y \sim \chi_p^2$, el cociente

$$T = \frac{Z}{\sqrt{Y/p}} \sim t_p,$$

donde t_p denota la t de Student con p grados de libertad.

1.20.1 ¿Cuál es la forma de la densidad de una t-Student?

💡 Tiene colas más pesadas que una normal

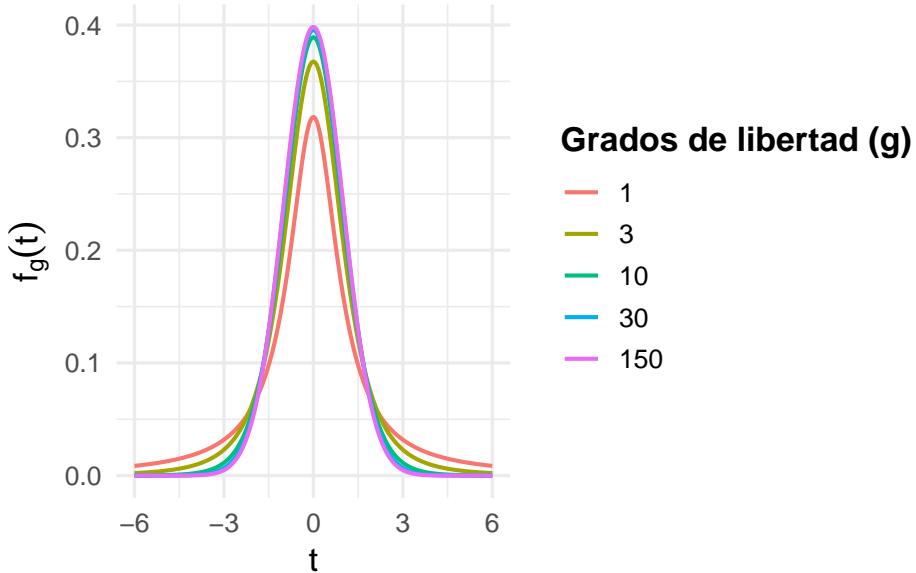


Figure 1: Densidad de la distribución t de Student para varios valores de los grados de libertad

1.21 Distribución F de Snedecor para el cociente de varianzas

Proposición 1.21.1 Consideremos U_1 y U_2 dos variables aleatorias independientes con distribución χ^2 con p_1 y p_2 grados de libertad, respectivamente.

El cociente $F = \frac{U_1/p_1}{U_2/p_2}$ admite la densidad

$$f_F(x) = \frac{\Gamma(\frac{p_1+p_2}{2})}{\Gamma(p_1)\Gamma(p_2)} \left(\frac{p_1}{p_2}\right)^{p_1} \frac{x^{p_1/2-1}}{\left(1 + \frac{p_1}{p_2}x\right)^{\frac{p_1+p_2}{2}}}.$$

Esta distribución se llama F de Snedecor con p_1 y p_2 grados de libertad y escribimos $F \sim F_{p_1, p_2}$.

💡 Consecuencia

Consideremos X e Y vv.aa. normales independientes con varianzas σ_X^2 y σ_Y^2 , así como X_1, \dots, X_{n_X} y Y_1, \dots, Y_{n_Y} dos m.a.s. de X y Y , respectivamente. Deducimos que

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n_X-1, n_Y-1}.$$

1.21.1 ¿Cuál es la forma de la densidad de una F de Snedecor?

💡 Depende mucho de los grados de libertad

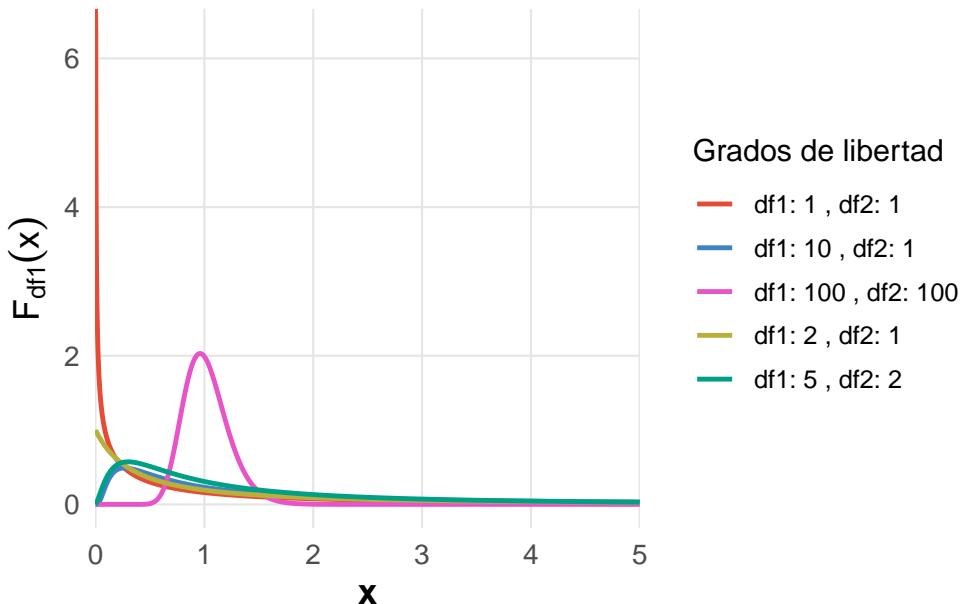


Figure 2: Densidad de la distribución F de Snedecor para varios valores de los grados de libertad

1.22 Si la distribución de X no es Normal

No podemos decir nada en general, [excepto](#) si n es grande...

i Teorema Central del Límite

Si n es “suficientemente” grande, se puede aproximar la distribución de \bar{X} por una Normal con media μ y varianza $\frac{\sigma^2}{n}$:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \text{ aproximadamente.}$$

i Formulación matemática

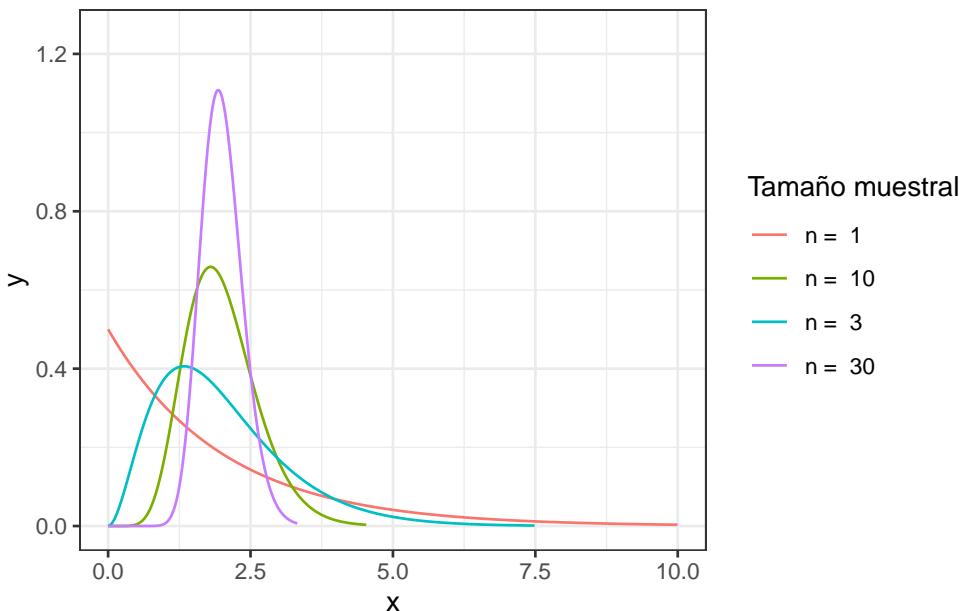
El resultado anterior se traduce por una convergencia de la sucesión de las variables aleatorias $(\bar{X}_n)_n$ en distribución cuando $n \rightarrow \infty$.

i Teorema Centro del Límite

¿Cuándo considerar que n es grande?

- Depende de la forma de la distribución de X :
 - Si X casi Normal: n pequeño es suficiente.
 - Si X muy asimétrico: n mucho más grande necesario.
- En general, se suele considerar $n \geq 30$ suficiente...

Ilustración, X inicial $\sim \text{Exp}(\lambda = 0.5)$, \bar{X} con $n = 3, 10$ y $n = 30$



1.23 Distribución muestral de la proporción muestral

i Contexto

- Hay situaciones donde X toma el valor 0 o 1, con probabilidades $1 - p$ y p , respectivamente.
- Por ejemplo, el siguiente experimento: escoger una pieza en la producción. $X = 1$ si es defectuosa, $X = 0$ si es correcta.
- Repetimos n veces el experimento. Obtenemos

$$x_1 = 1, x_2 = 0, x_3 = 0, x_4 = 1, x_5 = 0, \dots, x_n$$

- La proporción muestral es:

$$\hat{p} = \frac{N}{n}.$$

i Distribución exacta de \hat{p}

¿Cuál es la distribución de N ?

- Experimento simple con situación dicotómica, repetimos n veces...

$$N = \mathcal{B}(n, p).$$

- Podemos usar esta distribución para hacer cálculos exactos...

i Distribución aproximada de \hat{p}

Tenemos que $N = X_1 + X_2 + \dots + X_n$, por lo tanto

$$\hat{p} = \frac{N}{n} = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}.$$

Por el Teorema Central del Límite:

$$\hat{p} = \mathcal{N} \left(p, \frac{p(1-p)}{n} \right) \text{ aproximadamente.}$$

Podemos usar esta distribución para hacer cálculos aproximados.

1.24 Simulación y método de Monte-Carlo

i Motivación

En muchas situaciones, la capacidad de simular valores de las distribuciones de interés puede resultar útil calcular o estimar cantidades relevantes para la inferencia sobre la distribución de X_1, X_2, \dots, X_n .

i ¿Qué es ser capaz de simular valores de una distribución f ?

- Se refiere a la posibilidad de producir, para cualquier tamaño k , conjuntos de valores u_1, u_2, \dots, u_k , cuyo comportamiento imita el de k realizaciones aleatorias independientes de la distribución f .
- Quiere decir que las propiedades del conjunto generado lo hace indistinguible, si le aplicamos tests de independencia o de bondad de ajuste, de k realizaciones independientes de f .

i Simulación y método de Monte-Carlo

- Como hemos visto en los primeros ejemplos, gráficos como el histograma de frecuencias se comportan como la función de densidad de la variable de la que provienen las observaciones. También se pueden utilizar gráficos como la función de distribución empírica que veremos más adelante.
- Como consecuencia, dado un estadístico, si podemos obtener un número grande de observaciones del mismo podemos a través de algunos gráficos obtener información sobre su distribución.
- Podemos generar esas observaciones a través de lo que se conoce como simulaciones, observaciones generadas mediante algún algoritmo.
- Esta metodología que se puede aplicar en muchas otras situaciones se conoce como el [método de Monte-Carlo](#).

1.24.1 Muestreo de Monte-Carlo para aproximar esperanzas

i Ley de los grandes números

Consideremos una muestra aleatoria simple X_1, X_2, \dots, X_n de una distribución X . Para cualquier función g que cumple $E[g^2(X)] < +\infty$, tenemos que, con probabilidad 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n g(X_i) = E[g(X)].$$

1.24.1.1 Ejemplos

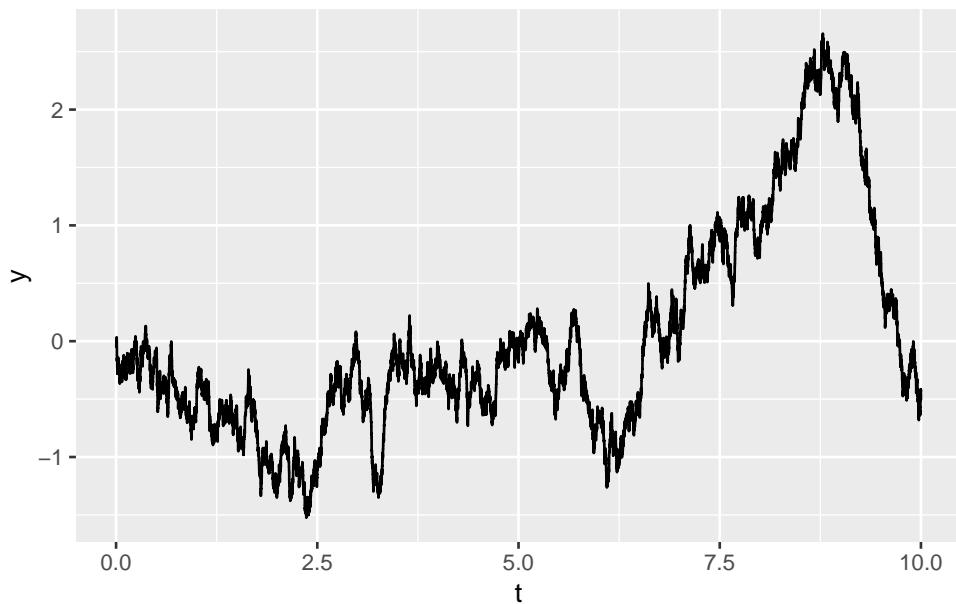
- $g(x) = x : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]$, es decir $\bar{X}_n \rightarrow E[X] = \mu_X$.
- $g(x) = (x - \mu_X)^2 : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)^2 = E[(X - \mu_X)^2] = \text{Var}(X)$.
- Si combinamos las dos convergencias anteriores: $\lim_{n \rightarrow \infty} \frac{1}{n} (X_i - \bar{X}_n)^2 = \text{Var}(X)$, es decir, $S_n^2 \rightarrow \text{Var}(X)$.
- $g(x) = 1_{x \leq q} : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq q} = P(X \leq q)$, es decir, que las frecuencias acumuladas relativas convergen hacia la probabilidad asociada.

1.24.1.2 Aplicaciones

i Ejemplo: el movimiento Browniano

- Es proceso muy usado en la predicción de precios (opciones) en matemáticas financieras.
- Una caracterización simplificada: es la suma infinitesimal de pequeñas contribuciones normales independientes.
- Para cualquier t , $W_t \sim \sum_{i=1}^{\frac{t}{h}} \sqrt{h} \cdot Z_i$, donde h es el paso infinitesimal y Z_i son normales estándares independientes e idénticamente distribuidos.

1.25 Movimiento Browniano



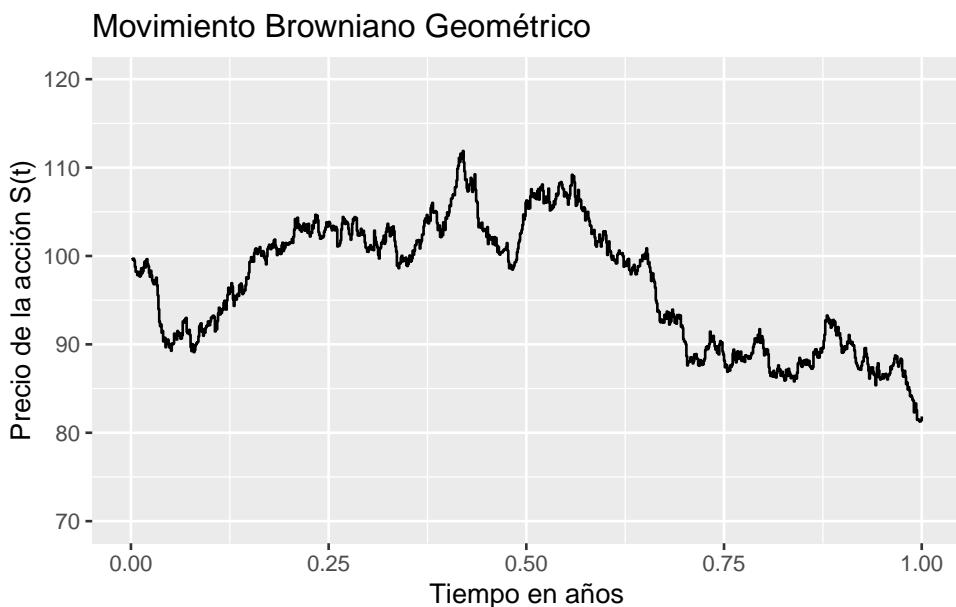
1.26 En finanzas, el modelo de Black-Scholes

i El movimiento Browniano Geométrico:

Lo propusieron Merton, Scholes y Black como modelo teórico para precios de acciones:

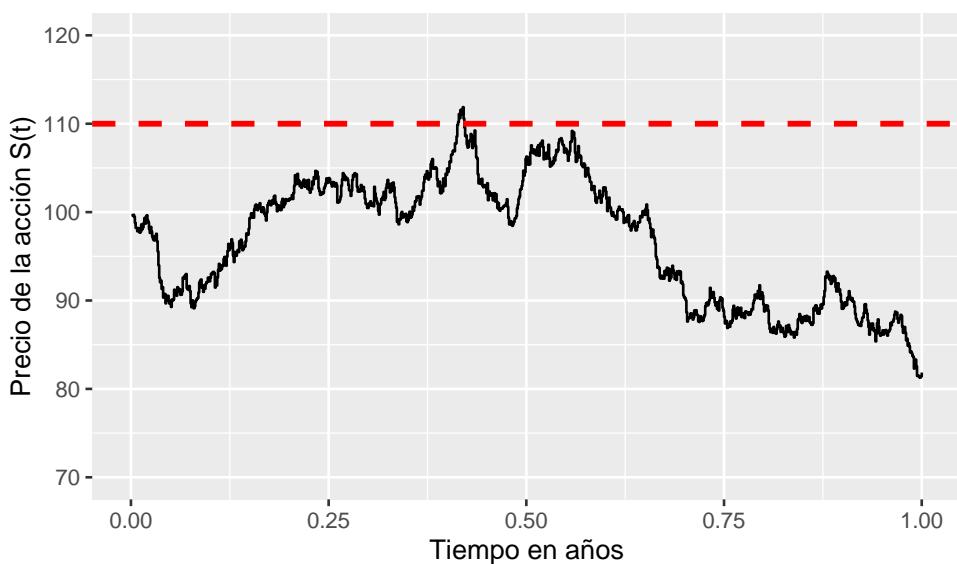
$$S_t = S_0 \exp \left(\left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right).$$

- S_0 : precio inicial de la acción.
- μ : el drift.
- σ^2 : la volatilidad.



¿Cuándo observamos $S(t) \geq 100\text{€}$?

Movimiento Browniano Geométrico

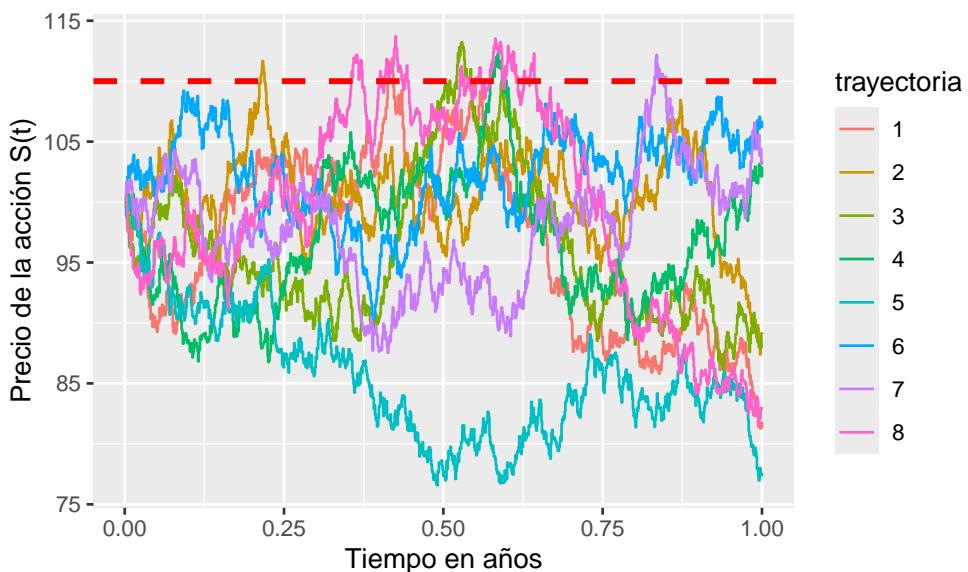


- Podemos simular muchas trayectorias del Movimiento Browniano Geométrico y observa qué pasa con el tiempo en el que supera el umbral 110.
- Esto es Monte-Carlo.
- Luego podremos obtener indicadores de la distribución de este tiempo.

💡 Tip

Representamos las 8 primeras

Movimiento Browniano Geométrico



1.27 Simulación y método de Monte-Carlo

💡 Simulación de variables aleatorias

- Lo que nos planteamos ahora es qué algoritmo podemos utilizar para generar observaciones (simulaciones) de una variable aleatoria.
- Para un gran número de variables aleatorias y modelos probabilísticos, estas simulaciones ya están incorporadas en los paquetes y lenguajes de programación con enfoque matemático y estadístico, como R.
- Aquí describiremos uno de los métodos más básicos, basado en la inversa de la función de distribución.
- El [método de la función inversa](#) es uno de los principales métodos de simulación de variables aleatorias.
- Está basado en la inversa de la función de distribución de una variable aleatoria.
- No todas las funciones de distribución admiten inversa, como la conoceremos usualmente. Para ello es

necesario que sea continua y estrictamente creciente.

i Inversa generalizada de la función de distribución o función cuantil

Dada una función de distribución F , su [función cuantil \(inversa generalizada\)](#) se define como

$$F^{-1}(p) = \inf\{x | F(x) \geq p\}, \text{ para todo } p \in (0, 1).$$

i Propiedades de la función cuantil

- 1) $F(F^{-1}(p)) \geq p$, para todo $p \in (0, 1)$.
- 2) $F^{-1}(F(x)) \leq x$, para todo x donde $F(x) > 0$.
- 3) Si $U \sim U(0, 1)$ entonces $F^{-1}(U)$ tiene como distribución F .
- 4) Dada una variable aleatoria X , con media finita y función de distribución F se verifica que

$$E[X] = \int_0^1 F^{-1}(u) du.$$

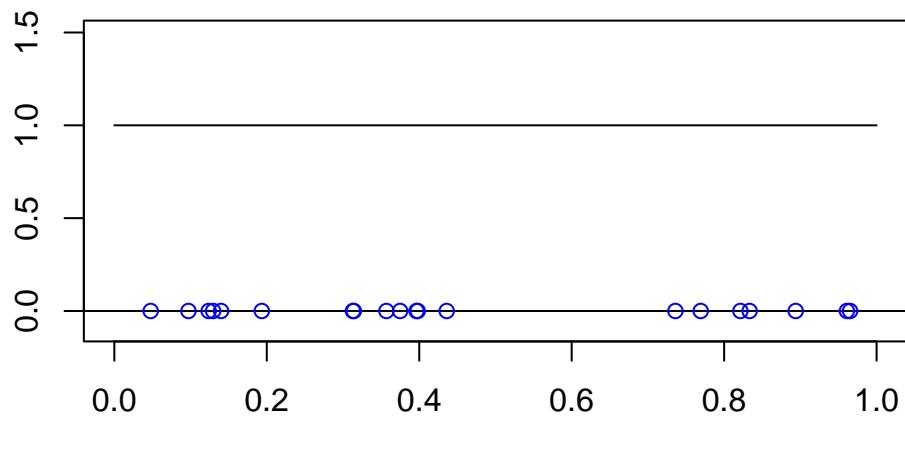
i Método de la función inversa

- Dada una variable aleatoria X con función de distribución F , si generamos observaciones independientes $U \sim U(0, 1)$, u_1, u_2, \dots, u_n , entonces los valores

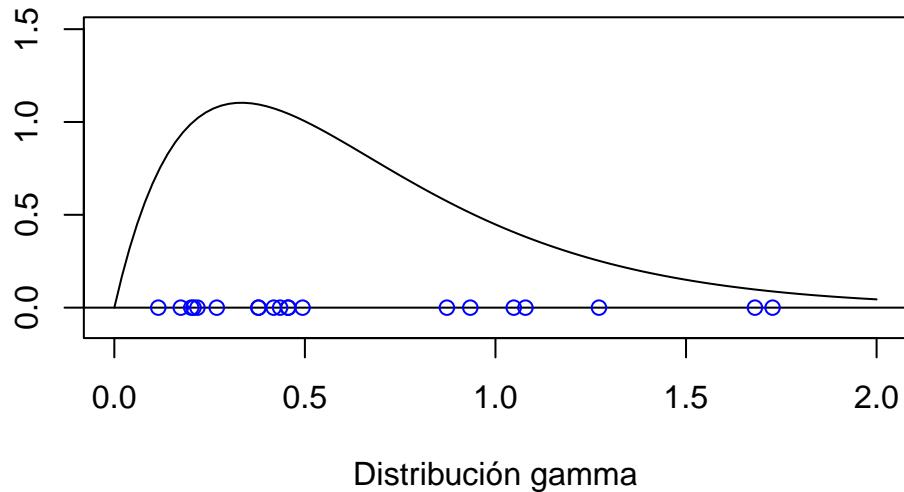
$$x_1 = F^{-1}(u_1), x_2 = F^{-1}(u_2), \dots, x_n = F^{-1}(u_n)$$

constituyen una observación de la muestra aleatoria simple de tamaño n de la variable aleatoria X .

Generación de valores de una distribución uniforme



Generación de valores de una distribución gamma



Transformación de una variable aleatoria

Planteamiento del problema

En muchas ocasiones, tenemos una variable aleatoria continua X , de la que conocemos la función de densidad, pero nos interesa conoer cómo se comporta una transformación de X , $Y = \varphi(X)$.

Teorema 1.27.1 Sea X una variable aleatoria con función de densidad f_X definida en un intervalo abierto $(a, b) \subseteq \mathbb{R}$. Sea $\varphi : (a, b) \rightarrow \mathbb{R}$ tal que:

- Es continua.
- Es estrictamente creciente o decreciente.
- φ^{-1} es diferenciable.

Entonces, la variable aleatoria $Y = \varphi(X)$ tiene función de densidad

$$f_Y(y) = \begin{cases} f_X(\varphi^{-1}(y)) \left| \frac{d}{dx}(\varphi^{-1}(y)) \right|, & \text{si } y \in \varphi(a, b) \\ 0, & \text{en otro caso} \end{cases}$$

Función características

Definición 1.27.1 Sea X una variable aleatoria cualquiera. La *función característica* de X se define como

$$\phi(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

en donde $i = \sqrt{-1}$.

Tip

Nota: $\phi(t) = \int_{-\infty}^{\infty} \cos(tx) dF(x) + i \int_{-\infty}^{\infty} \sin(tx) dF(x)$.

Propiedades

- **Existencia:** $|\phi(t)| \leq 1$, para todo $t \in \mathbb{R}$.
- Si X e Y son variables aleatorias **independientes**, entonces

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t),$$

para todo $t \in \mathbb{R}$.

Desigualdades

i Desigualdad de Markov

Si Z es una variable aleatoria no negativa con media finita $E[Z]$ y $\varepsilon > 0$, entonces

$$\varepsilon \Pr[Z \geq \varepsilon] = \varepsilon \int_{[\varepsilon, \infty)} dF_Z(x) \leq \int_{[\varepsilon, \infty)} x dF_Z(x) \leq \int_{[0, \infty)} x dF_Z(x) = E(Z)$$

(donde $F_Z(x) = \Pr[Z \leq x]$ es su función de distribución), es decir

$$\Pr[Z \geq \varepsilon] \leq \frac{E[Z]}{\varepsilon}.$$

i Desigualdad de Chebyshev

Si X es una variable aleatoria con media finita $\mu = E[X]$ y varianza $\sigma^2 = \text{Var}(X) > 0$, entonces tomando $Z = \frac{(X - \mu)^2}{\sigma^2} \geq 0$ y aplicando la desigualdad de Markov, tenemos

$$\Pr\left[\frac{(X - \mu)^2}{\sigma^2} \geq \varepsilon\right] \leq \frac{1}{\varepsilon}$$

para todo $\varepsilon > 0$.

También se puede escribir como

$$\Pr[(X - \mu)^2 < \varepsilon\sigma^2] \geq 1 - \frac{1}{\varepsilon}$$

o como

$$\Pr[|X - \mu| < r] \geq 1 - \frac{\sigma^2}{r^2},$$

para todo $r > 0$.

2 Estimación

2.1 Introducción

- Hemos modelizado un experimento con una variable aleatoria X .
- La [estimación](#) hace referencia al proceso de conseguir información sobre la distribución de X a partir de los valores de una muestra, aproximando valores asociados a la distribución mediante el valor de un estadístico en un muestra concreta.

! Dos situaciones

- Nuestro modelo supone que la distribución de X pertenece a una familia paramétrica de distribuciones: tienen una determinada forma con unos parámetros variables.
 - Buscamos información sobre el valor de los parámetro. [Estimación paramétrica](#).
- No limitamos la familia de distribuciones a la que pertenece nuestro modelo.
 - Buscamos información sobre la distribución en sí (función de distribución, de densidad o función de probabilidad). [Estimación no paramétrica](#).
- A lo largo de las prácticas veremos también la estimación de parámetros que no necesitan de una familia paramétrica, como es el caso de la mediana.

2.2 Ejemplos de estimación paramétrica

i Sondeo sobre intención de participación en unas elecciones

- Queremos estimar la tasa de participación antes de unas elecciones generales.
- Formulamos un modelo: experimento: “escoger una persona al azar en el censo”. X : variable dicotómica (“Sí” o “No”). $p = P(X = \text{Si})$.
- El modelo pertenece a la familia paramétrica de Bernoulli.

i Determinación de la concentración de un producto

- Quiero determinar la concentración.
- Formulo el modelo: experimento=“llevar a cabo una medición”. X : “valor proporcionado por el aparato”. $X \sim \mathcal{N}(\mu, \sigma^2)$.
- El modelo pertenece a la familia paramétrica de las distribuciones normales.

2.3 Estimación paramétrica: estimación puntual

i Ingredientes del modelo

- Experimento aleatorio
- Variable aleatoria X con una distribución f que pertenece a una familia paramétrica $\{f_\theta, \theta \in \Theta\}$.
- Disponemos de una muestra de la distribución de X .

Definición 2.3.1 Cualquier estadístico diseñado para aproximar el valor de un parámetro θ del modelo, se llama [estimador puntual](#) del parámetro θ .

i Ejemplos de estimadores paramétricos

θ	Estimador
μ	\bar{X} , media muestral
σ^2	S^2 , varianza muestral
p	\hat{p} , proporción muestral

⚠ A tener en cuenta:

- Un estimador es un variable aleatoria, su valor depende de la muestra concreta escogida.
- Para controlar bondad de nuestra estimación, nos basaremos en el estudio de la distribución del estimador.

2.4 Métodos de construcción de estimadores

Para μ, σ^2 o p , es fácil pensar en estimadores naturales, pero para modelos o parámetros más sofisticados, vamos a ver métodos generales.

Veremos dos métodos en este tema:

- El método de los momentos.
- El método de la máxima verosimilitud.

2.5 Método de los momentos

Contexto

- Experimento con una variable aleatoria X , suponemos $f_X \in \{x \mapsto f_\theta(x), \theta \in \Theta\}$.
- El parámetro θ tiene dimensión p .
- Consideramos una muestra aleatoria simple X_1, X_2, \dots, X_n de X .

Si θ tiene dimensión p , igualamos los p primeros momentos de f_θ con los equivalentes muestrales.

Sea $\mu_k(\theta)$ el momento de orden k de la distribución $f_{\theta, \mu_k} = E[X^k]$. Resolvemos:

$$\begin{aligned}\mu_1(\theta) &= \bar{X}, \\ \mu_2(\theta) &= \bar{X^2}, \\ &\vdots \\ \mu_p(\theta) &= \bar{X^p}.\end{aligned}$$

Ejemplos

! Calculad los estimadores usando el método de los momentos en los dos casos:

- Modelo normal: $X \sim \mathcal{N}(\mu, \sigma^2)$, donde $\theta = (\mu, \sigma^2)$.

$$\begin{aligned}X &\sim \mathcal{N}(\mu, \sigma^2) & \mu &= E[X] \\ \theta &= (\mu, \sigma^2), \quad p = 2 & \sigma^2 &= \text{Var}[X] = E[X^2] - (E[X])^2 \rightarrow E[X^2] = \sigma^2 + \mu^2 \\ \mu_1(\theta) &= E[X] = \bar{x} & \hat{\mu} &= \bar{x} \\ \mu_2(\theta) &= E[X^2] = \bar{x^2} & \hat{\sigma}^2 &= \bar{x^2} - (\bar{x})^2\end{aligned}$$

- Modelo de Bernoulli: $X \sim \text{Bernoulli}(p)$, donde desconocemos p .

$$\begin{aligned}X &\sim \text{Bernoulli}(p), \quad 0 < p < 1 \\ \theta &= p \\ \mu(\theta) &= E[X] = \bar{x} = p \\ \hat{p} &= \bar{x} \text{ (proporción muestral)}\end{aligned}$$

2.6 Método de máxima verosimilitud

- El método más utilizado de construcción de un estimador puntual.
- Se basa en lo que se conoce como [función de verosimilitud](#).

Definición 2.6.1

- Sea X una variable aleatoria, con distribución $x \mapsto f_X(x; \theta)$ ([función de densidad](#) o [función puntual de probabilidad](#)), donde θ es de dimensión $p : \theta \in \Theta \subset \mathbb{R}^p$.
- Para un valor concreto de una muestra aleatoria simple (X_1, \dots, X_n) , que denotamos por (x_1, \dots, x_n) , consideramos la función de θ :

$$L_n : \left\{ \begin{array}{l} \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}^+ \\ \theta \mapsto L_n(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta). \end{array} \right.$$

- La función L_n es la [función de verosimilitud](#). Nos dice lo creíbles (verosímiles) que son las observaciones para ese valor del parámetro.

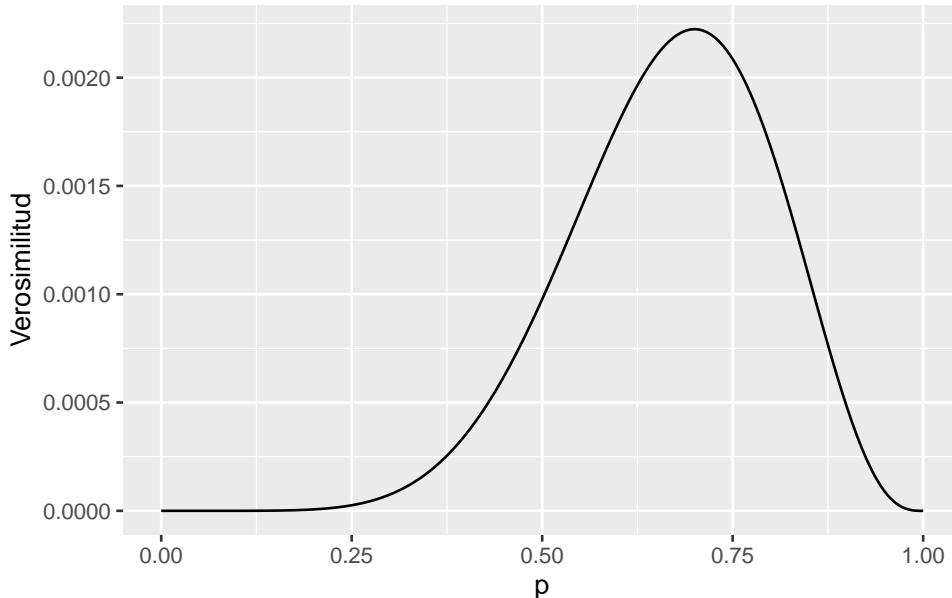
Ejemplo de cálculo de la verosimilitud

- Tiramos 10 veces una moneda (1 es cara, 0 es cruz), y obtenemos: 0,0,1,0,1,1,1,1,1.
- La verosimilitud asocia a cada p el valor de

$$\Pr(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1, X_6 = 1, X_7 = 1, X_8 = 1, X_9 = 1, X_{10} = 1),$$

por lo que

$$L_n(p) = (1-p)(1-p)p(1-p)p^6 = (1-p)^3 \cdot p^7.$$



2.7 Estimador de máxima verosimilitud

Definición 2.7.1 El *estimador de máxima verosimilitud* $\hat{\theta}$ de θ es cualquier valor de θ que maximiza $\theta \mapsto L_n(\theta)$, es decir,

$$\hat{\theta} \operatorname{argmax}_{\theta \in \Theta} L_n(\theta).$$

! Nota

- La maximización se realiza sobre todos los valores admisibles para el parámetro θ .
- Podría haber de un máximo.

Estimación de la proporción

Retomamos el ejemplo de las 10 monedas:

$$p \mapsto L_n(p) = (1-p)(1-p)p(1-p)p^6 = (1-p)^3 \cdot p^7.$$

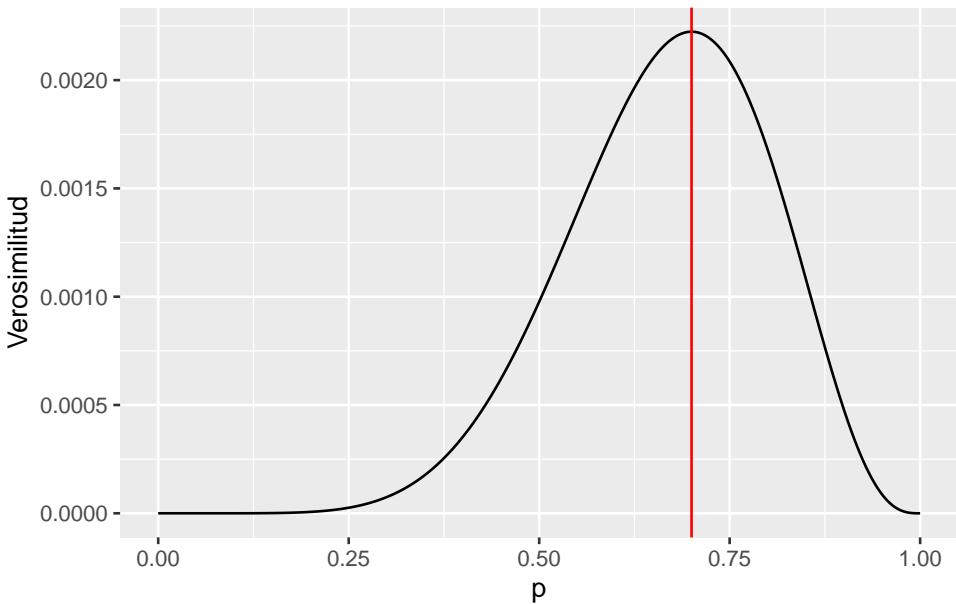


Figure 3: Verosimilitud correspondiente al ejemplo de 10 tiradas de una moneda.

Ejemplos

! Ejemplos de estimación por máxima verosimilitud

Calculad los estimadores usando el método de máxima verosimilitud en los dos casos:

- $X \sim \text{Bernoulli}(p)$, donde desconocemos p .
- $X \sim \mathcal{N}(\mu, \sigma^2)$, donde desconocemos $\theta = (\mu, \sigma^2)$.

- En el primer caso anterior, calcular la distribución de Bernoulli, donde desconocemos p .

$X \sim \text{Bernoulli}(p)$, $0 < p < 1$

(X_1, \dots, X_N) m.a.s. de tamaño n

$$L_n(\theta) = f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta)$$

$$L_n(p) = \prod_{i=1}^n f_X(x_i; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i}$$

$$\ell(p) = \log L_n(p) = \log p^{\sum_i x_i} + \log (1-p)^{n - \sum_i x_i} = \left(\sum_i x_i \right) \cdot \log p + \left(n - \sum_{i=1}^n x_i \right) \cdot \log (1-p)$$

$$\begin{aligned} \frac{\partial \ell(p)}{\partial p} &= \left(\sum_{i=1}^n x_i \right) \cdot \frac{1}{p} + \left(n - \sum_{i=1}^n x_i \right) \cdot \left(-\frac{1}{1-p} \right) = (1-p) \cdot \sum_{i=1}^n x_i - \left(n - \sum_{i=1}^n x_i \right) \cdot p \\ &= \sum_{i=1}^n x_i - p \cdot \sum_{i=1}^n x_i - np + p \cdot \sum_{i=1}^n x_i = 0 \implies \hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \end{aligned}$$

- En el segundo caso anterior, calculad la esperanza del estimador de máxima verosimilitud de σ^2 .

La función de densidad de una variable aleatoria X con distribución normal es:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Si tenemos una muestra de tamaño n , es decir, X_1, X_2, \dots, X_n que son observaciones independientes y distribuidas como $\mathcal{N}(\mu, \sigma^2)$, entonces la función de verosimilitud $L(\mu, \sigma^2)$ es el producto de las funciones de densidad de cada observación

$$L(\mu, \sigma^2) = \prod_{i=1}^n f(x_i|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Para simplificar los cálculos, se toma el logaritmo de la función de verosimilitud, lo que da la función de log-verosimilitud $\ell(\mu, \sigma^2)$:

$$\begin{aligned}\ell(\mu, \sigma^2) &= \log L(\mu, \sigma^2) = \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) + \sum_{i=1}^n \log \left(\exp \left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right) \\ &= \sum_{i=1}^n \log \left((2\pi\sigma^2)^{-\frac{1}{2}} \right) + \sum_{i=1}^n -\frac{(x_i - \mu)^2}{2\sigma^2} = \sum_{i=1}^n -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\ &= -\frac{n}{2} \cdot \log(2\pi) - \frac{n}{2} \cdot \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\end{aligned}$$

Para encontrar los estimadores de máxima verosimilitud, derivamos $\ell(\mu, \sigma^2)$ con respecto a μ y σ^2 , e igualamos a cero:

$$\begin{aligned}\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} &= \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \rightarrow \sum_{i=1}^n (x_i - \mu) = 0 \rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\ \frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \rightarrow -n\sigma^2 + \sum_{i=1}^n (x_i - \mu)^2 = 0 \rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2\end{aligned}$$

2.8 Métodos para evaluar un estimador

💡 Recordar

- Un estimador es una variable aleatoria.
- Es valioso disponer de conocimiento sobre la distribución del estimador (su [distribución en el muestreo](#))
⇒ permite manejar el riesgo y el error que podemos cometer al aproxima θ por $\hat{\theta}$ asociado.

Consideraremos dos aspectos de la distribución muestral de $\hat{\theta}$

- Su localización: [sesgo](#).
- Su variabilidad: [error cuadrático medio](#).
- Mencionaremos su comportamiento cuando $n \rightarrow \infty$.

2.9 Sesgo

Definición 2.9.1 Consideramos para un estimador $\hat{\theta}$ de un parámetro θ : $E_{\theta}[\hat{\theta}] - \theta$.

Esta diferencia se llama el [sesgo](#).

💡 Una propiedad deseable para un estimador

Si el sesgo de un estimador es nulo para todo valor de θ , decimos que el estimador es [insesgado](#).

2.10 Error cuadrático medio

Para medir la variabilidad en el muestreo de un estimador.

Definición 2.10.1 El [error cuadrático medio del estimador](#) $\hat{\theta}$ es la función de θ definida por

$$\theta \mapsto E_{\theta}[(\hat{\theta} - \theta)^2]$$

💡 Para practicar

Calculad el error cuadrático medio del estimador de máxima verosimilitud (e.m.v.) de μ para una muestra aleatoria simple de $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$X \sim \underbrace{\mathcal{N}(\mu, \sigma^2)}_{\text{desconocidos}}$$

$\hat{\mu} = \bar{x}$ ([estimador basado en los movimientos](#))

$E[\bar{X}] = \mu \rightarrow \hat{\mu} = \bar{X}$ es un estimador insesgado para μ

$$\left. \begin{aligned} \text{Var}[\bar{X}] &= \frac{\sigma^2}{n} \\ \text{Var}[\bar{X}] &= E[(\bar{X} - \mu)^2] \end{aligned} \right\} \rightarrow E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n} \quad (\text{E.C.M.})$$

2.11 Balance entre sesgo y varianza

💡 Sesgo y varianza

El error cuadrático medio se puede descomponer

$$E_\theta[(\hat{\theta} - \theta)^2] = \text{Var}(\hat{\theta}) + [\text{sesgo}(\hat{\theta})]^2$$

En ocasiones, se consigue un menor error cuadrático medio con un estimador sesgado y estamos dispuestos a sacrificar el sesgo, por conseguir una menor varianza.

Demostración 2.11.1

$$\begin{aligned} E_\theta[(\hat{\theta} - \theta)^2] &= E_\theta[(\hat{\theta} - E_\theta[\hat{\theta}] + E_\theta[\hat{\theta}] - \theta)^2] \\ &= E_\theta[(\hat{\theta}) - E_\theta[\hat{\theta}] + (E_\theta[\hat{\theta}] - \theta)^2 + 2(\hat{\theta} - E_\theta[\hat{\theta}])(E_\theta[\hat{\theta}] - \theta)] \\ &= \text{Var}[\hat{\theta}] + [\text{sesgo}(\hat{\theta})]^2 \end{aligned}$$

2.12 Comportamiento asintótico de un estimador

💡 Cuando el tamaño muestral crece

Muchos de los resultados en inferencia se obtienen en un contexto asintótico: nos interesa comprobar el comportamiento de nuestro estimador cuando crece el número de observaciones.

- ¿Converge $\hat{\theta}$ hacia el verdadero valor del parámetro?
- ¿Cómo se comporta el error $\hat{\theta} - \theta$?

2.13 Consistencia

Definición 2.13.1 Una sucesión de estimadores $(\hat{\theta}_n)_n$ es consistente si converge hacia θ en probabilidad, para todo θ . Es decir:

$$\forall \varepsilon > 0, \quad P_\theta \left[|\hat{\theta}_n - \theta| > \varepsilon \right] \xrightarrow{n \rightarrow \infty} 0$$

💡 Note

Si el error cuadrático medio de una sucesión de estimadores tiende a cero, esta sucesión es consistente.

2.14 Normalidad asintótica

- Cuando un estimador es consistente, $(\hat{\theta}_n - \theta)$ converge hacia cero en probabilidad.
- Se busca entonces la velocidad de convergencia hacia cero.

💡 Un resultado que se puede demostrar para muchos modelos, para el e.m.v:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, A)$$

(la distribución de $\sqrt{n}(\hat{\theta}_n - \theta)$ se aproxima a la distribución $\mathcal{N}(0, A)$).

Ejemplo: distribución gamma(α, β)

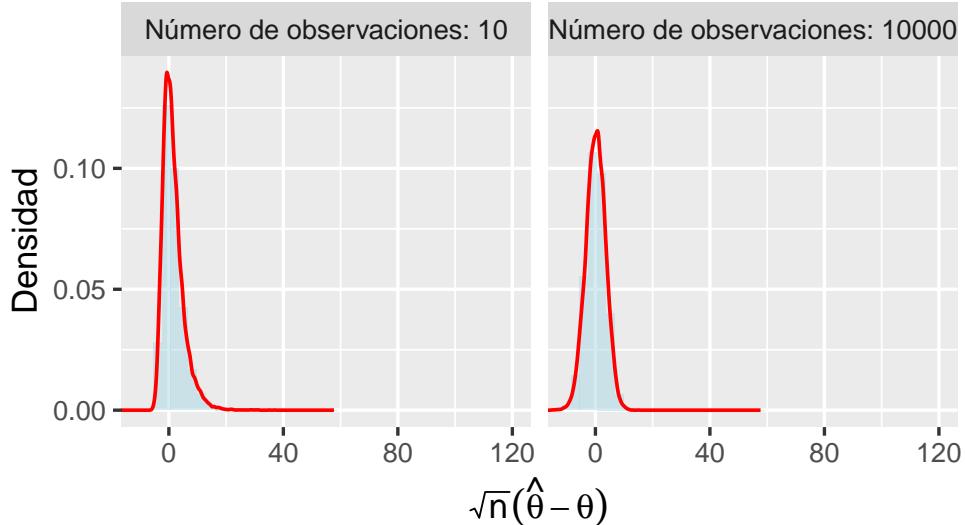
En práctica, veremos que los estimadores de los momentos para (α, β) son

$$\hat{\alpha} = \frac{(\hat{X})^2}{\frac{X^2 - (\bar{X})^2}{n}}$$

$$\hat{\beta} = \frac{\frac{X^2 - (\bar{X})^2}{n}}{\hat{X}}$$

Simulamos 10000 muestras de tamaño 10 y 10000 muestras de tamaño 10000.

Estimador momentos de α



2.15 Estimación no paramétrica

i Estimación de funciones asociadas a una variable aleatoria

En muchas ocasiones no interesa suponer un modelo específico para la variable aleatoria, y por lo tanto, no necesitamos conocer el valor de los parámetros que los caracterizan, sino que nos interesa aproximar el valor de alguna función asociada a la variable. Estas funciones suelen ser:

- Función de distribución: $F(x) = P(X \leq x)$ para todo $x \in \mathbb{R}$.
- Función puntual de probabilidad: $p(x) = P(X = x)$, para todo x en el soporte de X (en el caso discreto).
- Función de densidad: $f(x) = F'(x)$, para todo x en el soporte de X (en el caso continuo).
- En estas situaciones es posible considerar lo que se conoce como **estimadores no paramétricas** de las funciones anteriores en un valor concreto de x .
- En los dos primeros casos, puesto que se trata de estimar probabilidades, podremos usar la proporción muestral para hacer aproximaciones (lo veremos a continuación).
- En el caso de la función de densidad recurriremos a lo que se conoce como **estimadores tipo núcleo**.

i Estimación de la función de distribución

Dado que la función de distribución en un punto x es una probabilidad, $F(x) = P(X \leq x)$, tenemos como estimador la correspondiente proporción muestral que vendrá dada por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i),$$

que se conoce como la **función de distribución empírica**.

i Estimación de la función puntual de probabilidad

De la misma forma que antes el estimador $p(x) = P(X = x)$ vendrá dado por

$$\hat{p}(x) = \frac{\text{número de veces que aparece el valor } x \text{ en la muestra}}{n}$$

Primera opción: nos basamos en frecuencias relativas

- La estimación de la función de densidad no es tan sencilla.

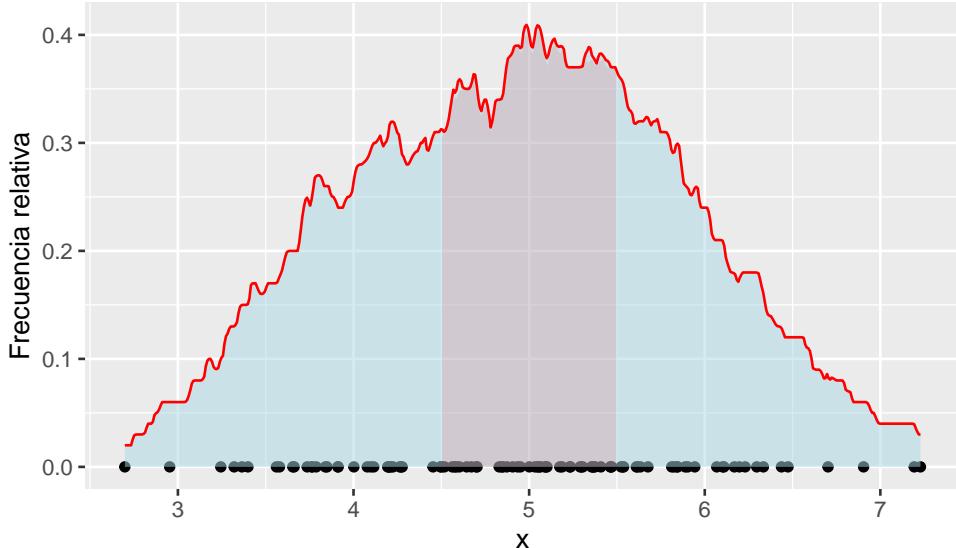
- A partir de la expresión de la función de densidad como derivada de la función de distribución, y dada una muestra aleatoria simple X_1, \dots, X_n de una variable aleatoria X , para cada x , estimaciones $F_X(x)$ por la frecuencia relativa de un vecindario de x :

$$\hat{f}_n(x) = \frac{1}{2nh} \sum_{i=1}^n I_{(x-h, x+h)}(x_i),$$

h es la mitad del ancho del vecindario, “suficientemente pequeño”.

Es una especie de histograma con ventana móvil, que se va deslizando.

100 valores observados, ventana móvil con $h = 0.5$



2.16 Estimación tipo núcleo

Podemos escribir

$$\hat{f}_n(x) = \frac{1}{2nh} \sum_{i=1}^n I_{(x-h, x+h)}(x_i),$$

como

$$\begin{aligned}\hat{f}_n(x) &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), \text{ con } K(z) = \frac{1}{2}I_{(-1,1)}(z). \\ \hat{f}_n(x) &= \frac{1}{n} \sum_{i=1}^n K_h(x - x - i), \text{ con } K_h(z) \frac{1}{h}K\left(\frac{z}{h}\right) = \frac{1}{2h}I_{(-h,h)}(z)\end{aligned}$$

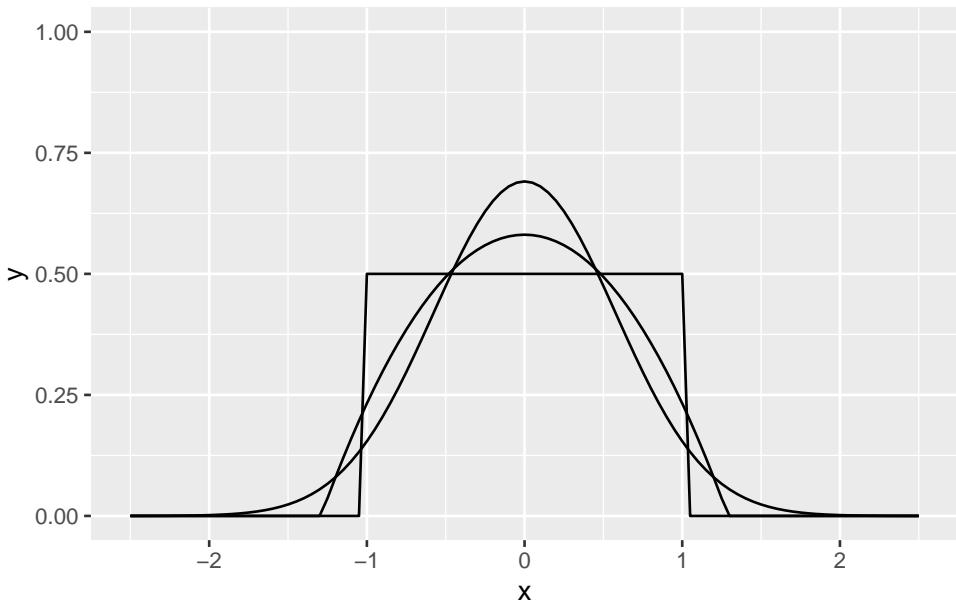
- En la expresión anterior hay dos elementos principales, la función K , que en el ejemplo es la función de densidad de una distribución uniforme, y el parámetro h , que he dejado en 0.5 en el ejemplo, y en función de su valor tiene en cuenta valores de la muestra cercanos o alejados al valor x .
- En el primer caso hablamos sobre la [función núcleo \(kernel\)](#) y en el segundo caso del [ancho de banda \(bandwith\)](#).
- La principal característica de la función K es que al ser la densidad de una distribución uniforme da el mismo peso a todos los puntos de la muestra que se encuentran en un entorno de valor x donde estimamos la densidad.
- Podríamos pensar en otra función K con la propiedad de ser simétrica y unimodal en el 0, como en el caso anterior, pero con distintos pesos para las observaciones.

2.16.1 Podemos variar el núcleo

i Función núcleo

- En la literatura hay distintas propuestas de la función K , [función núcleo](#).
- Cualquier estimador de la densidad en la forma anterior, con una función núcleo K , recibe el nombre de [estimador tipo núcleo](#).
- Los casos más usados de función núcleo son los siguientes:
 - **Núcleo normal:** K es la función de densidad de una distribución normal estándar. Suele ser el núcleo que más se utiliza.
 - **Núcleo de Epanechnikov:** $K(z) = (1 - z^2)I_{(-1,1)}$. Se considera que es el núcleo más eficiente.
- El caso considerado en la introducción se conoce como el [núcleo rectangular](#).

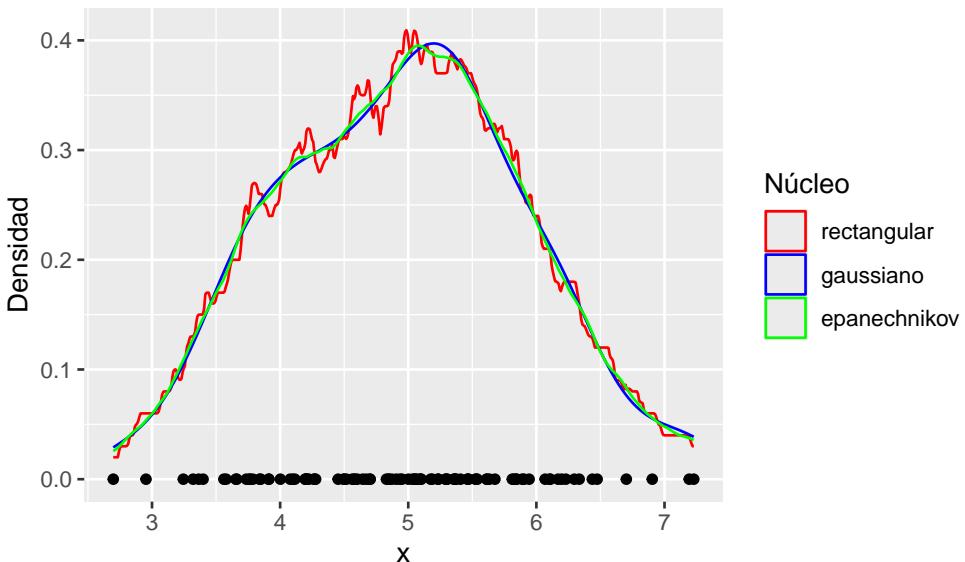
Por ejemplo que tenga más peso cuando el valor muestral esté cerca del valor x . Este argumento parece bastante razonable, es decir, que en la estimación tengan más influencia los valores muestrales cercanos al valor x en el intervalo $x \pm h$.



- K uniforme.
- K gaussiano.
- K Kernel Epanechnikov:

$$K(z) = \frac{3}{4}(1 - z^2)I_{(-1,1)}(z)$$

100 valores observados, estimación núcleo



2.16.2 Se pueden demostrar resultados asintóticos

Función núcleo

Vamos a justificar, mediante el comportamiento asintótico, que los valores del estimador tipo núcleo dan buenas aproximaciones de la función de densidad.

Para el principal resultado se supondrán las siguientes condiciones sobre la función de densidad, la función núcleo y el ancho de banda:

- La densidad al cuadrado es integrable, admite derivada continua de segundo orden y esta al cuadrado es integrable. Seguiremos la notación $R(f'') = \int (f''(x))^2 dx$.
- El núcleo es una función de densidad simétrica y acotada, con momento de orden dos finito y al cuadrado es integrable. Seguiremos en este caso la notación $R(K) = \int (K(x))^2 dx$, y por $\mu_2(K) = \int x^2 K(x) dx$ el valor del momento de orden 2 asociado a la densidad K .
- Conforme aumentamos el tamaño de muestra el ancho de banda, h_n , es una secuencia de valores positivos con límite 0 y nh_n tiende a $+\infty$.

Si suponemos:

- f_X es L^2 , es derivable dos veces y f''_X es L^2 .
- K es simétrica, acotada, K es L^2 y admite momento de orden 2.
- Consideramos $h_n \rightarrow 0$ cuando $n \rightarrow \infty$, y además $nh_n \rightarrow \infty$.

i Resultado

Para todo x ,

$$E[\hat{f}_n(x) - f(x)] = \frac{1}{2}\mu_2(K)f''(x)h_n^2 + o(h_n^2) \text{ y } \text{Var}[\hat{f}_n(x)] = \frac{R(K)}{nh_n}f(x) + o((nh_n)^{-1}).$$

- Este resultado junto con las suposiciones que hemos hecho anteriormente nos lleva a concluir que se producen buenas aproximaciones en términos del sesgo y la varianza, pues ambos tienden a 0 cuando aumentamos el tamaño de muestra.

2.17 Ancho de banda

i Si variamos el ancho de banda:

- Cuánto mayor es la ventana o vecindario, la curva estimada es más suavizada.
- Si la ventana es pequeña, la curva estimada es “wiggly”.
- Puesto que el ancho de banda se puede hacer tan grande como queramos nos planteamos cómo elegir un ancho de banda que sea óptimo en algún sentido.
- En general el problema de la selección del ancho de banda no es sencillo.
- Se han propuesto elecciones óptimas del ancho de banda, que minimicen la distancia entre la función de densidad f_X y el estimador \hat{f}_X .

! Elección óptima del ancho de banda

$$h_n = \left(\frac{R(K)}{\mu_2^2(K)R(f'')n} \right)^{\frac{1}{5}}.$$

2.17.1 Elección óptima del ancho de banda

En la fórmula

$$h_n = \left(\frac{R(K)}{\mu_2^2(K)R(f'')n} \right)^{\frac{1}{5}}$$

aparece $R(f'')$ que es desconocida.

! Veremos dos maneras de aproximar $R(f'')$

- Aproximación Gaussiana de f_X .
- Estimar f_X de manera no paramétrica.

2.17.1.1 Aproximación gaussiana para el cálculo de $R(f'')$

i Suponemos que f es la función de densidad de una distribución normal con desviación típica σ :

$$h_n = \left(\frac{8\pi^{\frac{1}{2}}R(K)}{3\mu_2^2(K)n} \right)^{\frac{1}{5}} \sigma$$

En la práctica, si la desviación típica σ es desconocida la aproximamos por la cuasi-desviación típica muestra S :

$$\hat{h}_n = \left(\frac{8\pi^{\frac{1}{2}}R(K)}{3\mu_2^2(K)n} \right)^{\frac{1}{5}} S. \text{ Normal scales bandwidth selector}$$

i Si además consideramos que el núcleo K es gaussiano:

$$\hat{h}_{n,RP} = \left(\frac{4}{3n}\right)^{\frac{1}{5}} S.$$

Se llama la “rule of thumb” (la regla del pulgar).

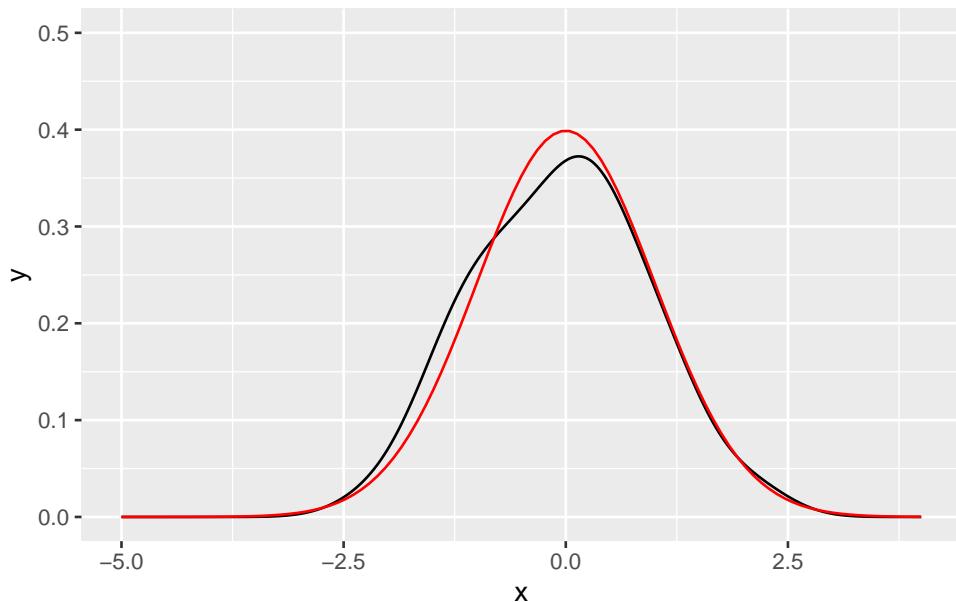
La forma más complicada es dar una estimación no paramétrica de $R(f'')$. La metodología va más allá de los contenidos de esta asignatura y no veremos los detalles, solo indica que podemos seleccionar con **R** y lo veremos en prácticas.

Implementación en R

```
# más de una distribución normal de tamaño 100
set.seed(314159)
muestra <- rnorm(100, 0, 1)
# valor del ancho de banda segun la regla del pulgar
bw.nrd(x = muestra)

[1] 0.4047312

# gráfica del estimador de la función de densidad
tibble(x = muestra) |>
  ggplot(aes(x = x)) +
  geom_density(kernel = "gaussian", bw = "nrd") +
  geom_function(fun = dnorm, col = "red") +
  xlim(-5, 4) + ylim(0, 0.5)
```



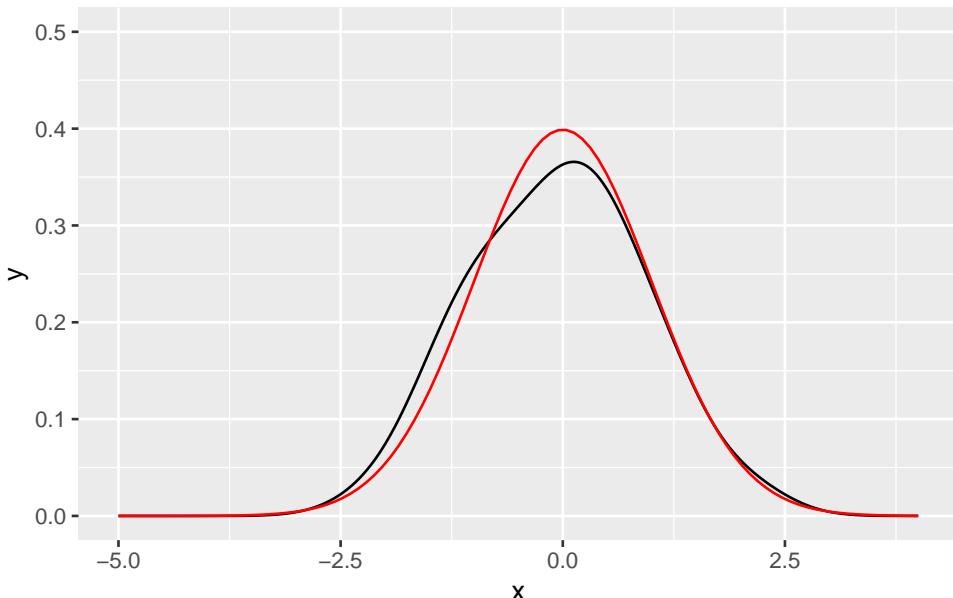
2.17.1.2 Aproximación no paramétrica para el cálculo de $R(f'')$, implementación en R

Se usa la especificación “SJ” (Sheather & Jones, 1991) para el bandwidth.

```
# valor del ancho de banda con estimación no paramétrica
bw.SJ(x = muestra)
```

```
[1] 0.442874
```

```
# grafica del estimador de la función de densidad
tibble(x = muestra) |>
  ggplot(aes(x = x)) +
  geom_density(kernel = "gaussian", bw = "SJ") +
  geom_function(fun = dnorm, col = "red") +
  xlim(-5, 4) + ylim(0, 0.5)
```



2.18 Introducción al Bootstrap

i **Bootstrap:** una manera de evaluar la incertidumbre asociada a un estimador

Permite obtener información acerca de la distribución muestral de un estadístico relevante, usando remuestreo de la muestra disponible, es decir, submuestra de la muestra original.

- Permite estimar la distribución de un estimador (como la media, mediana, varianza, etc.) a partir de muestras obtenidas de los datos disponibles.
- Se utiliza especialmente cuando la distribución poblacional es desconocida o cuando el tamaño de la muestra es pequeño.
- Es muy útil para hacer inferencias estadísticas sin hacer suposiciones sobre la distribución subyacente de los datos.
- Se aplica incluso cuando el modelo no permite obtener la distribución muestral.
- Nos permite aproximar el error estándar al estimador, lo que nos da una idea de su precisión.

? ¿De dónde viene el nombre?

Está asociado a la expresión: “*To pull oneself up by one’s bootstrap*”, que hace referencia a situaciones en las que uno cuenta con sus propios medios para salir de una dificultad, sin contar con ayuda externa.

- Se basa en la aproximación de la función de distribución mediante la función de distribución empírica y los métodos de simulación.
- Para utilizar el método de simulación necesitamos conocer la distribución F . Como en muchas situaciones no conocemos F , lo que vamos a hacer es generar muestras a partir de F_n .

i Determinación del error muestral de un estadístico

- Recordamos que el error estándar asociado a un estadístico es su desviación típica muestral.
- Para un estimador, nos indica la variabilidad de los valores que toma respecto a todas las muestras posibles. Es, por lo tanto, un indicador valioso de su precisión.
- Por ejemplo, para una población normal, σ desconocida, el error estándar de \bar{X} es $\frac{\sigma}{\sqrt{n}}$.
- En general, dado un parámetro θ y un estadístico $\hat{\theta}$ que approxima (estima) el valor de θ , su error muestral viene dado por

$$E[(\hat{\theta} - \theta)^2].$$

- La idea es simular muestras de F_n para obtener en cada una de ellas el valor del estadístico.
- Con cada una de ellas calculamos el equivalente empírico de $E[(\hat{\theta} - \theta)^2]$, mediante el promedio de las desviaciones al cuadrado de los valores del estadístico respecto del promedio de todas las estimaciones.

i La distribución empírica asociada a una muestra

- Consideramos una muestra concreta $\mathbf{x} = (x_1, \dots, x_n)$, asociada a una distribución F en una población.
- La distribución empírica \hat{F} se construye únicamente a partir de esos valores observados y es la distribución que asigna una probabilidad $\frac{1}{n}$ a cada valor $x_i, i = 1, 2, \dots, n$.
- Es por lo tanto una distribución discreta, sus valores posibles son los de x_1, \dots, x_n .
- Es decir \hat{F} asocia a cualquier evento A , la probabilidad *empírica*:

$$\widehat{\text{Prob}}(A) = \frac{\#\{x_i \in A\}}{n}.$$

! Tened en cuenta

Se pueden repetir los valores en la muestra $\mathbf{x} = (x_1, \dots, x_n)$.

i La distribución empírica asociada a una muestra

- Es una distribución discreta, sus valores posibles son los de x_1, \dots, x_n .
- Es decir \hat{F} asociado cualquier evento A , la probabilidad *empírica*: $\widehat{\text{Prob}}(A) = \frac{\#\{x_i \in A\}}{n}$.
- Podemos calcular la esperanza de cualquier función g :

$$\hat{E}[g] = \frac{1}{n} \sum_{i=1}^n g(x_i).$$

- Por ejemplo:

$$\begin{aligned}\text{La esperanza de } \hat{F} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \\ \text{La esperanza de } \hat{F} &= \frac{(n-1)}{n} s^2.\end{aligned}$$

i El principio de plug-in

- Es un método simple para estimar parámetros a partir de una muestra.
- Estimamos un parámetro θ que se puede calcular como una función $t(F)$ de la distribución F , a través de su equivalente empírico:

$$\hat{\theta} = t(\hat{F}).$$

! Nota

- Por ejemplo, estimamos μ_F usando la esperanza de \hat{F} , que es \bar{x} .
- Es la idea que usamos en el método de los momentos también.

i Muestra Bootstrap

Extraer una muestra Bootstrap es extraer una muestra de la distribución empírica \hat{F} .

- Consiste en extraer al azar, de manera equiprobable y con reemplazo, n elementos del vector $\mathbf{x} = (x_1, \dots, x_n)$.
- Denotamos la muestra extraída por $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$.
- Al ser con reemplazo, algunos elementos de \mathbf{x} pueden aparecer repetidos y otros pueden no aparecer.

! Con R

```
x <- c(5.3, 4.8, 5.4, 5.0, 4.8)
set.seed(3.14159)
(x_bootstrap <- sample(x, replace = TRUE))
[1] 4.8 4.8 5.0 5.0 4.8
```

2.18.1 El error estándar Bootstrap

i Recordad

- El error estándar asociado a un estadístico es su desviación típica muestral.
- Para un estimador, nos indica la variabilidad de los valores que toma respecto a todas las muestras posibles, es por lo tanto un indicador valioso de su precisión.
- Por ejemplo, para una población normal, σ desconocida, el error estándar de \bar{X} es S/\sqrt{n} .

i Idea del error estándar Bootstrap

- Consideramos un estimador $\hat{\theta}$ que se expresa como $T(\mathbf{x})$.
- Vamos a aplicar el principio plug-in para aproximar el error estándar de $\hat{\theta}$, usando \hat{F} .
- El principio plug-in en este caso se describe como:
 - Consideramos el proceso de extraer una muestra de $\hat{F} : \mathbf{x}^*$.
 - Consideramos el estimador $T(\mathbf{x}^*)$.
 - Calculamos la desviación típica de la distribución muestral del estimador anterior, respecto a todas las muestras de \hat{F} que podría extraer.
- Consideramos una variable aleatoria X y una muestra x_1, \dots, x_n de X .
- Hemos observado \mathbf{x} , hemos calculado el estadístico $\hat{\theta} = T(\mathbf{x})$. Procederemos de la manera siguiente:
 - Extraemos B muestras Bootstrap $\mathbf{x}^{*1}, \dots, \mathbf{x}^{*n}$. Estas muestras independientes con reemplazamiento (pueden aparecer valores repetidos) de tamaño n de x_1, \dots, x_n .
 - Para cada muestra Bootstrap, calculamos el valor del estadístico y lo denotamos por $\hat{\theta}^{*b} = T(\mathbf{x}^{*b})$, para $b = 1, \dots, B$.
- Aproximamos el error estándar de $\hat{\theta}$, que denotamos por $\hat{s}\hat{e}$, como la desviación típica muestral de $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$, esto es, como

$$\hat{s}\hat{e} = \sqrt{\sum_{b=1}^B \frac{(\hat{\theta}^{*b} - \bar{\hat{\theta}}^{*b})^2}{B-1}}$$

3 Estimación por intervalos

3.1 Introducción

💡 Estimación paramétrica

En el contexto de la estimación paramétrica, queremos aprovechar nuestro conocimiento sobre la distribución muestral del estimador, para proporcionar margen de error y riesgo.

💡 Bootstrap

También construiremos intervalos de confianza en percentiles Bootstrap

❗ Idea básica

Supongamos:

- $X \sim \mathcal{N}(\mu, 4)$.
- Consideramos una muestra aleatoria simple X_1, \dots, X_4 , estimaciones μ con \bar{X} .

Deducimos:

- Sabemos que $\bar{X} \sim \mathcal{N}(\mu, 1)$.
- Implica: el 95% de los valores de \bar{X} está aproximadamente entre $\mu \pm 2$.
- Si sé donde está μ , sé que, con probabilidad 0.95, \bar{X} se encuentre a menos de 2 unidades.
- Ahora al revés: si he observado un valor de \bar{X} , ¿dónde está μ ? la probabilidad de que μ se encuentre a menos de 2 unidades de \bar{X} es 0.95.

La probabilidad de que, habiendo observado un valor de \bar{X} , μ se encuentre a menos de 2 unidades es 0.95.

$$\mu \in \bar{X} \pm 2, \text{ con probabilidad 0.95}$$

- $\bar{X} \pm 2$ es un intervalo aleatorio. - Para una muestra concreta que extraigamos, será $\bar{x} \pm 2$. - ± 2 expresa el margen de error. - “con probabilidad 0.95” expresa la confianza que tenemos en que nuestra información sea cierta. - Corremos un riesgo de error de 0.05 al afirmar, $\mu \in \bar{X} \pm 2$.

Definición 3.1.1

- (X_1, \dots, X_n) una muestra asociada a f_θ .
- Un *intervalo de estimación* está asociado por dos estadísticos $I(X_1, \dots, X_n)$ (extremo inferior) y $D(X_1, \dots, X_n)$ (extremo superior) y persigue capturar el valor del parámetro θ .

❗ Un intervalo de estimación es aleatorio

- Sus extremos dependen de la muestra concreta escogida.
- $\theta \in [I(X_1, \dots, X_n), D(X_1, \dots, X_n)]$ define un suceso aleatorio.
- La probabilidad de este suceso $\mathbb{P}_\theta[\theta \in [I(X_1, \dots, X_n), D(X_1, \dots, X_n)]]$ se llama **nivel de cobertura** del intervalo para este valor de θ .
- El nivel de cobertura depende de θ .

3.2 Nivel de confianza

3.2.1 Nivel de cobertura

$$\theta \mapsto \mathbb{P}_\theta[\theta \in [I(X_1, \dots, X_n), D(X_1, \dots, X_n)]]$$

El **nivel de confianza** es el valor mínimo del nivel de cobertura calculado cuando varía θ .

💡 Confianza vs Precisión

Deberemos buscar una alta confianza pero sin sacrificar en exceso la precisión.

3.3 Un procedimiento general de construcción

Se aplica en muchas situaciones de modelización.

i Procedimiento

- Nos fijamos en el “nivel de riesgo”, $0 < \alpha < 1$, por ejemplo, 0.1, 0.5, o 0.01.
- Buscamos $T(X_1, \dots, X_n)$ que se **pivota**, es decir, que su distribución no depende de θ .
- Escogemos dos cotas a y b :

$$\mathbb{P}_\theta[a \leq T(X_1, \dots, X_n) \leq b] = 1 - \alpha, \quad \forall \theta \in \Theta.$$

- Procuramos despejar θ :

$$\mathbb{P}_\theta[I(X_1, \dots, X_n) \leq \theta \leq D(X_1, \dots, X_n)] = 1 - \alpha.$$

?

Nota

- T es pivotal \implies el nivel de cobertura $1 - \alpha$ de $[I(X_1, \dots, X_n)]$ no depende de θ .
- $1 - \alpha$ es el nivel de confianza.

3.4 Intervalo de confianza para la media μ de $X \sim \mathcal{N}(\mu, \sigma^2)$, σ conocida

Consideramos una muestra aleatoria simple de una distribución $X \sim \mathcal{N}(\mu, \sigma^2)$. El valor de σ^2 es conocido

- Usaremos \bar{X} para estimar μ .

Estadístico pivotal:

$$T(X_1, \dots, X_n; \theta) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- Dibujamos en la densidad del estadístico pivotal $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, una región central que represente el $100(1 - \alpha)\%$ del área total.

