Fundamentos de Inferencia Estadística

Francisco Javier Mercader Martínez

Índice

1 Muestreo y distribuciones muestrales

1.1 Introducción

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

¿Está la moneda trucada?

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

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

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.

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

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 ($\Longleftrightarrow$ el censo electoral).
    - "1" $\Longleftrightarrow$ "la persona piensa ir a votar".
    - "0" $\Longleftrightarrow$ "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 extración 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.

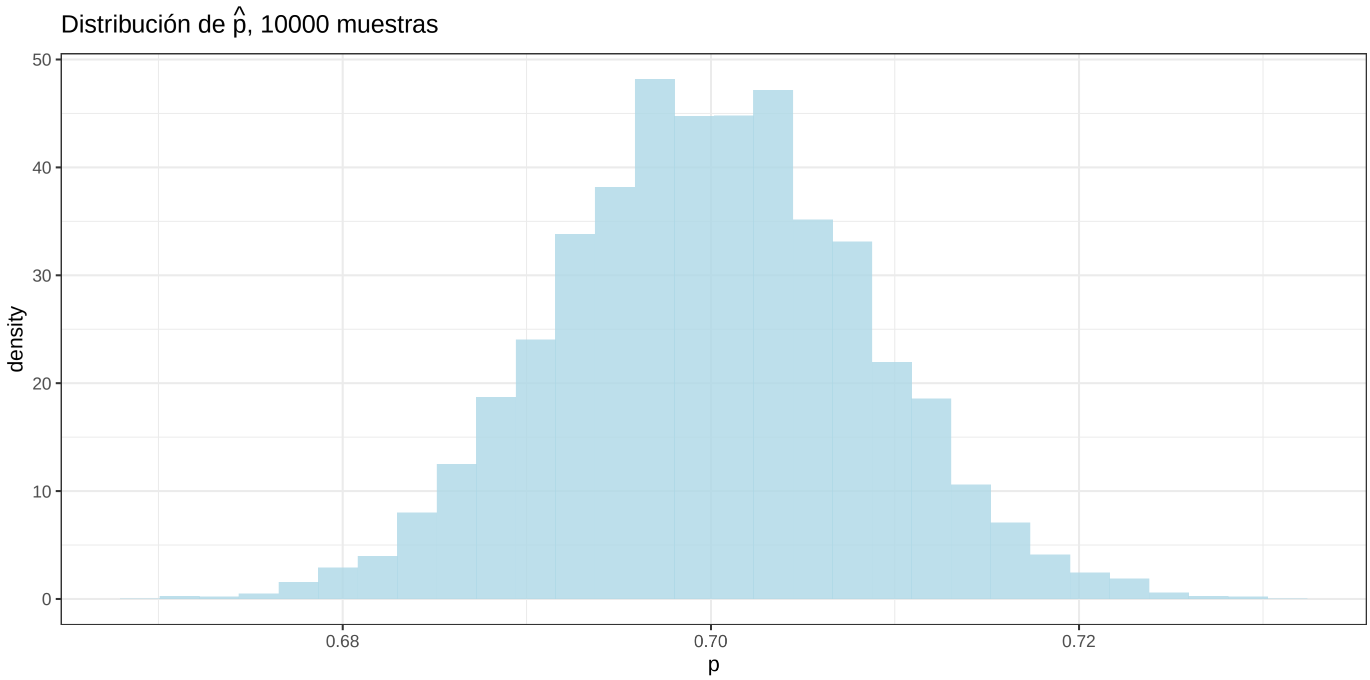
```r
library(tidyverse)
lista_muestras <- replicate(
    10000,
    sample(poblacion, 3000, replace = FALSE),
    simplify = FALSE
)
p_muestras <- map_dbl(lista_muestras, mean)
head(p_muestras)
```

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

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

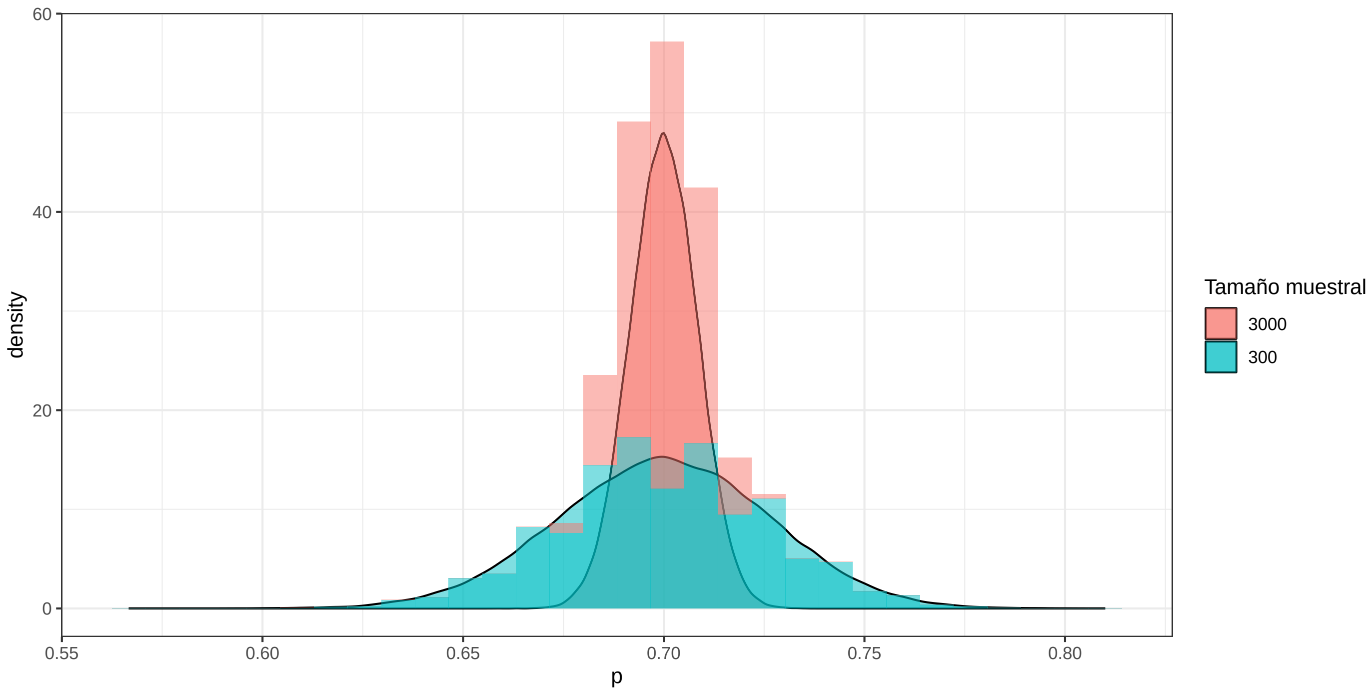
```
[1] 0.7090000 0.7080000 0.6946667 0.7216667 0.6906667 0.6990000
```

Recogemos los valores obtenidos en un histograma.



1.5 Realizacción 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

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 37 000 000, 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 $\hat{p}$?

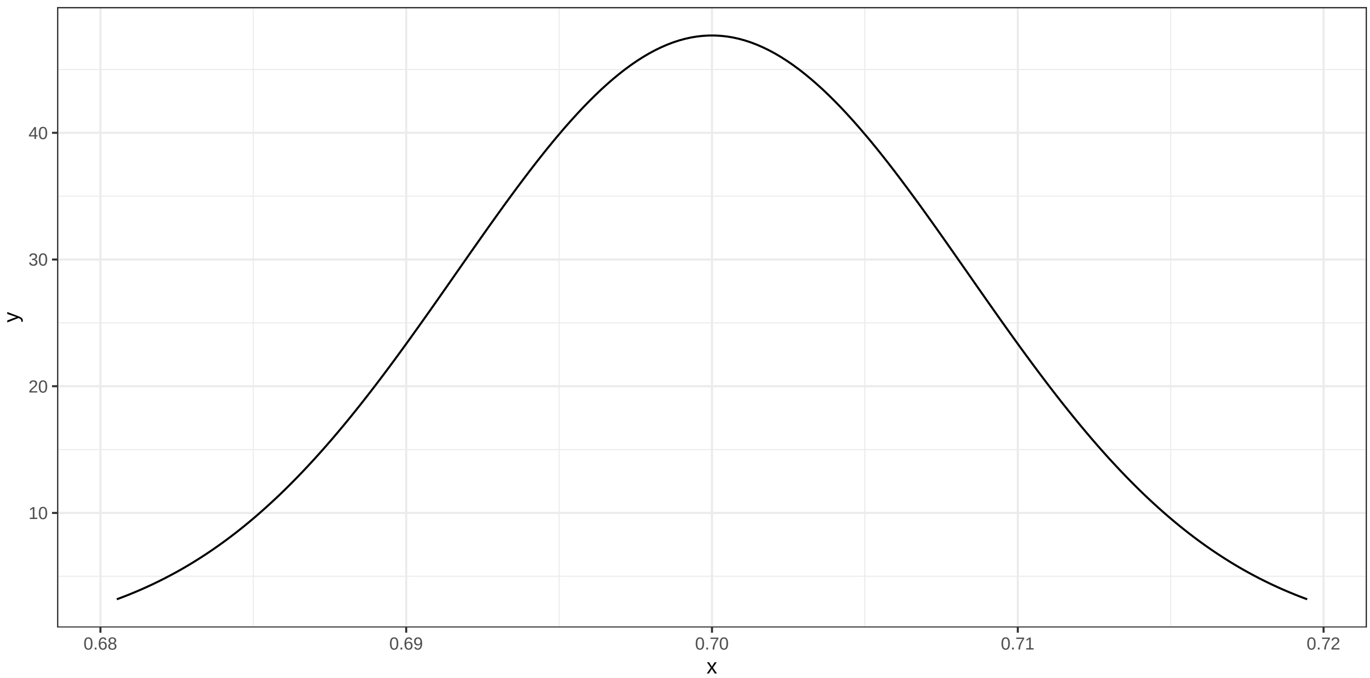
$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Es lo que llamamos la distribución muestral de $\hat{p}$.

1.7 Uso de la distribución muestral

La distribución muestral de $\hat{p}$:

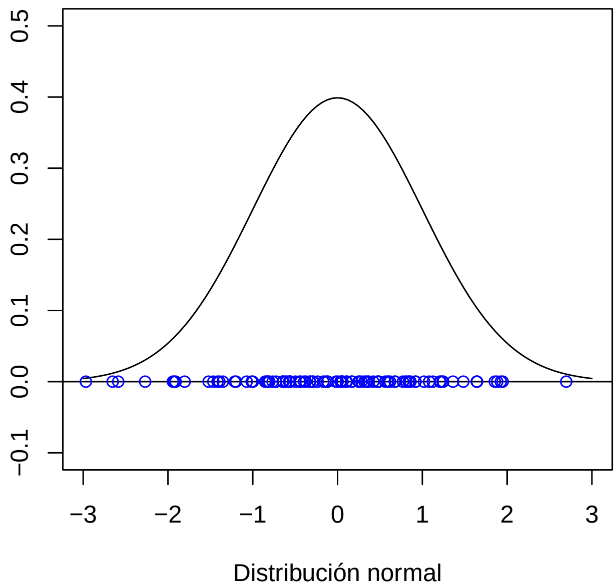
Es la distribución esperada de los valores de $\hat{p}$ respecto a todas las muestras de ese tamaño que podría extraer

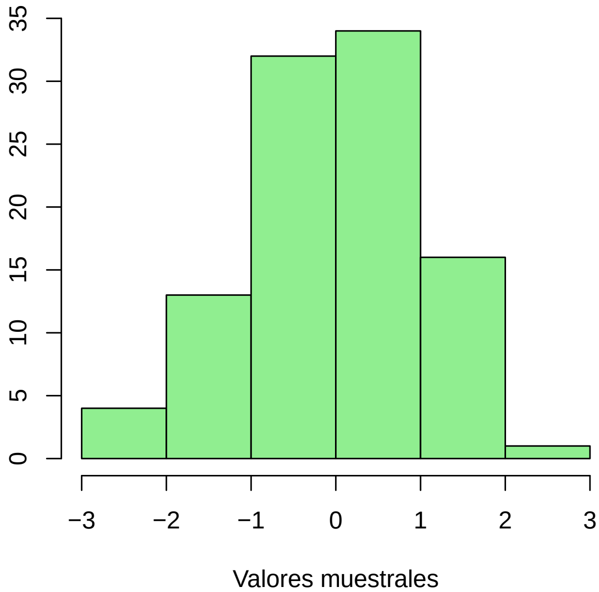


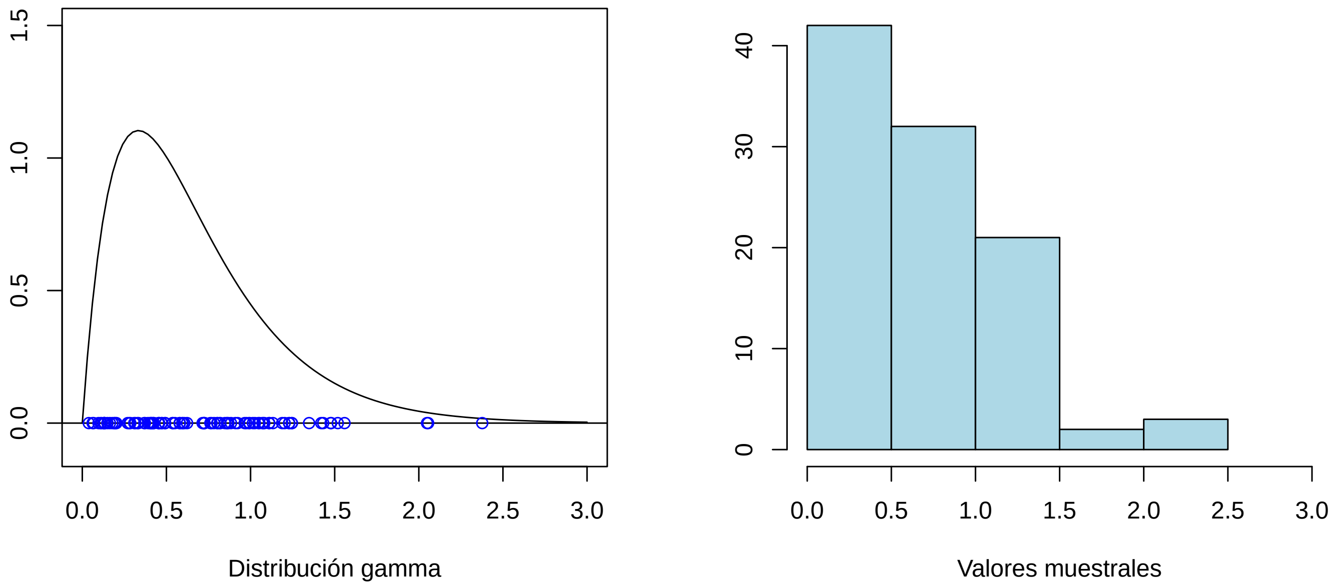
1.8 Antes de extraer una muestra:

- ¿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 $X_1, X_2, \ldots, X_n, \ldots$ una sucesión de variables (vv.aa) independientes con la distribución que X, entonces

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

1.11 Algunos términos

Definición 1.11.1

- *Sea una variable aleatoria X. Consideramos n variables aleatorias independientes e idénticamente distribuidas $X_1, X_2, \ldots, X_n$, que se distribuyen como X. La variable aleatoria multidimensional $(X_1, X_2, \ldots, X_n)$ 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

Contexto

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

$$x_1, x_2, \ldots, x_n.$$

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

Ejemplo:

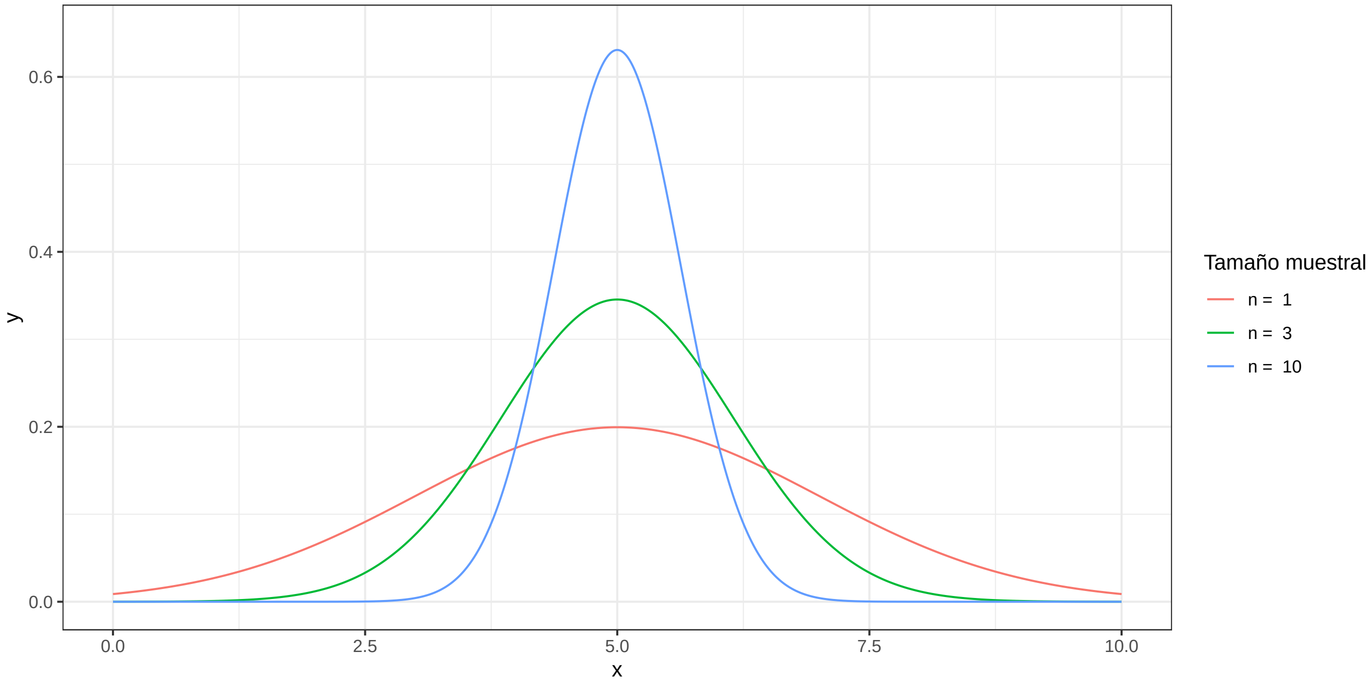
- Quiero medir una cantidad. Hay variabilidad en las mediciones.
- Introduzco una variable aleatoria $X =$ “valor proporcionado por el aparato”.
- μ representa el centro de los valores.
- Extraigo una muestra de tamaño 5 del valor de X

1.13.1 Esperanza y varianza de la media muestral

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

- Tenemos
$$\boxed{E[\bar{X}] = \mu.}$$
- Es decir que el centro de la distribución muestral de $\bar{X}$ coincide con el centro de la distribución X.
- Tenemos $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$, 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 $n = 10$

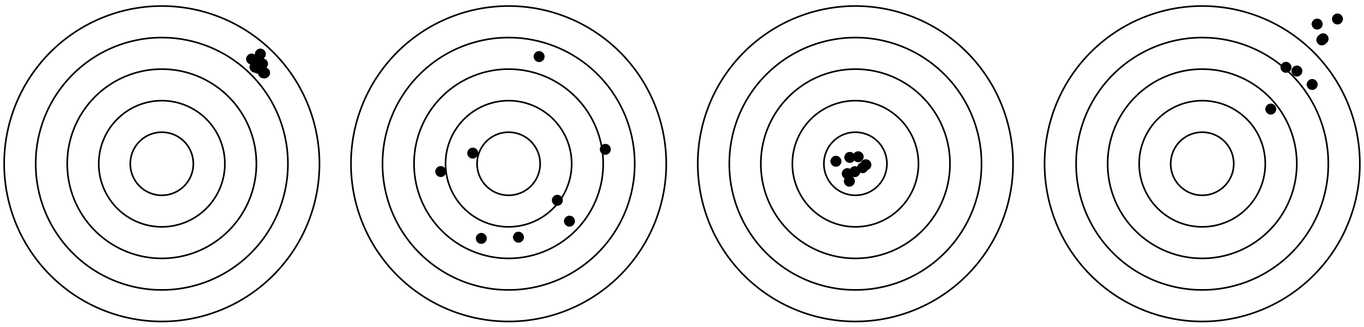


1.14 Consecuencia práctica

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

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

Fórmula alternativa para S_n^2:

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

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

1.15.1 Dos apuntes

En algunos textos en castellano

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

En estas fórmulas

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

1.16 Esperanza de la varianza muestral

Proposición 1.16.1 *Si $(X_1, X_2, \ldots, X_n)$ 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 normla, 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) $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. 3) $\dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$.

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 ξ^2 es igual a

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

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

Caracterización de la χ^2

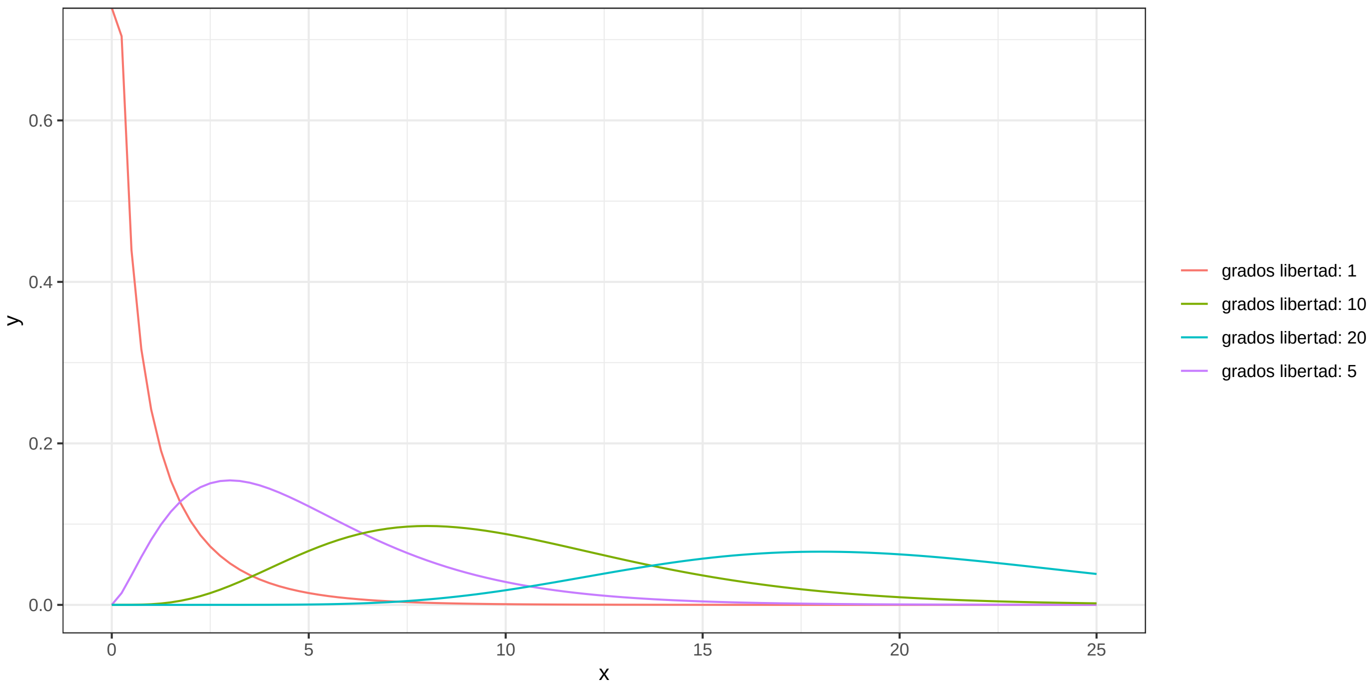
Si $Z_1, \ldots, Z_p$ son p variables aleatorias independientes, con $Z_i \sim \mathcal{N}(0,1)$, entonces la variable aleatoria X definida como

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

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

¿Cómo es su función de desidad?

Depende de los grados de libertad.



1.20 Distribución t-Student

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

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

La distribución que admit 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:

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

donde Γ dentoa la función Gamma.

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 $Y \sim \chi^2_p$, 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

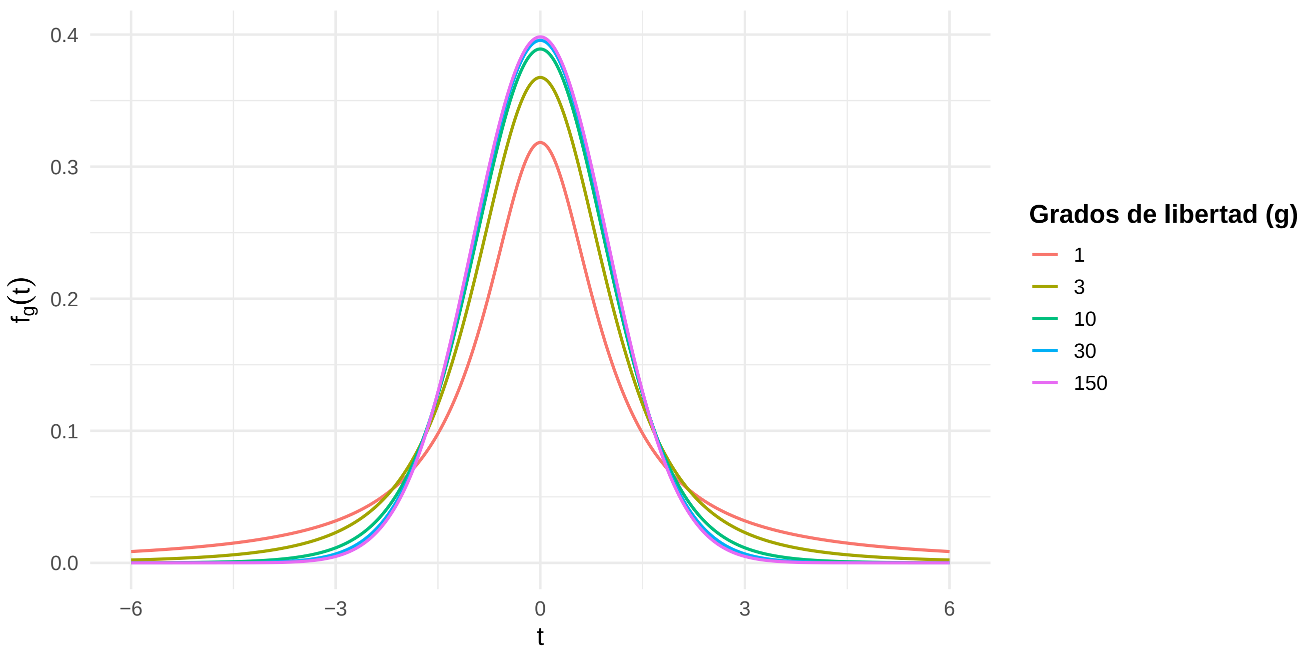


Figure 1: Densidad de la ditribución t de Student para varios valores de los grados de libertad

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 $F = \dfrac{U_1/p_1}{U_2/p_2}$ admite la densidad

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

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

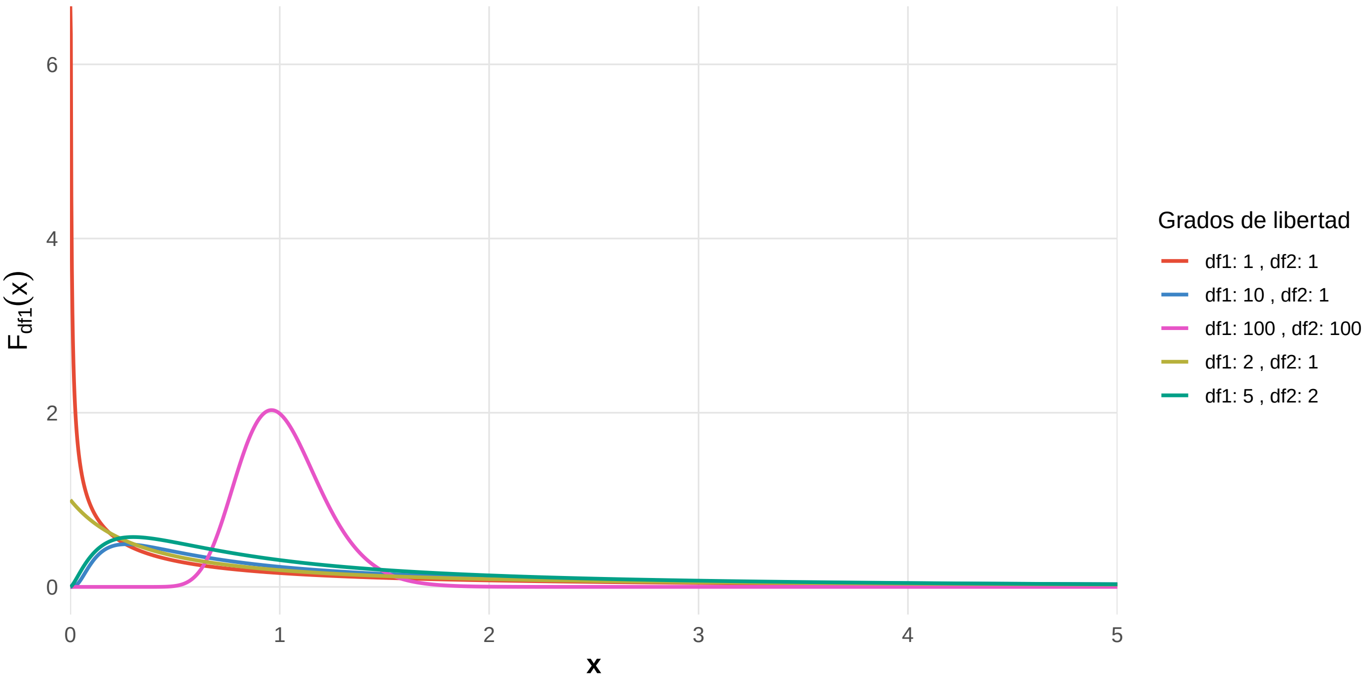


Figure 2: Densidad de la distribución F de Snedecor para varios valores de los grados de libertad

1.22 Si la distribución de X no es Normal

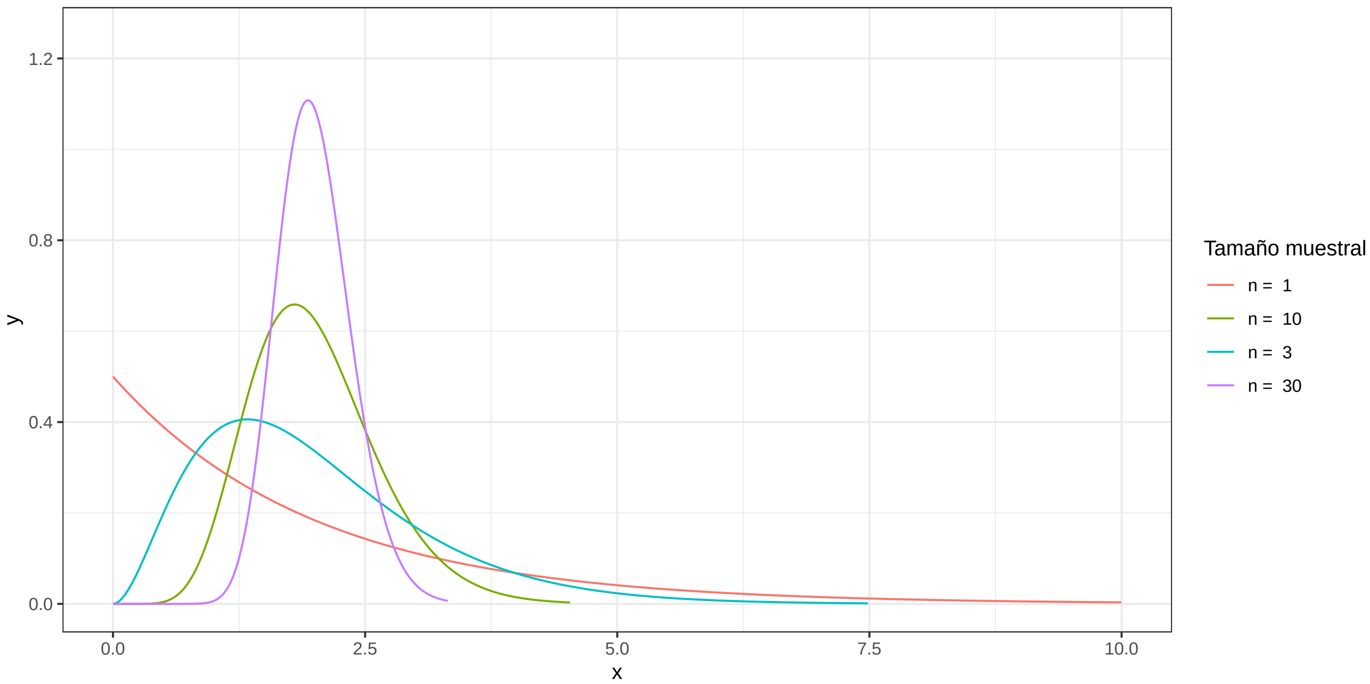
No podemos decir nada en general, excepto si n es grande...

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

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

> **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

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:

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

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...

Distribución aproximada de $\hat{p}$

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

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

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

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 $X_1, X_2, \dots, X_n$.

¿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.

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 utilziar 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

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,

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

1.24.1.1 Ejemplos

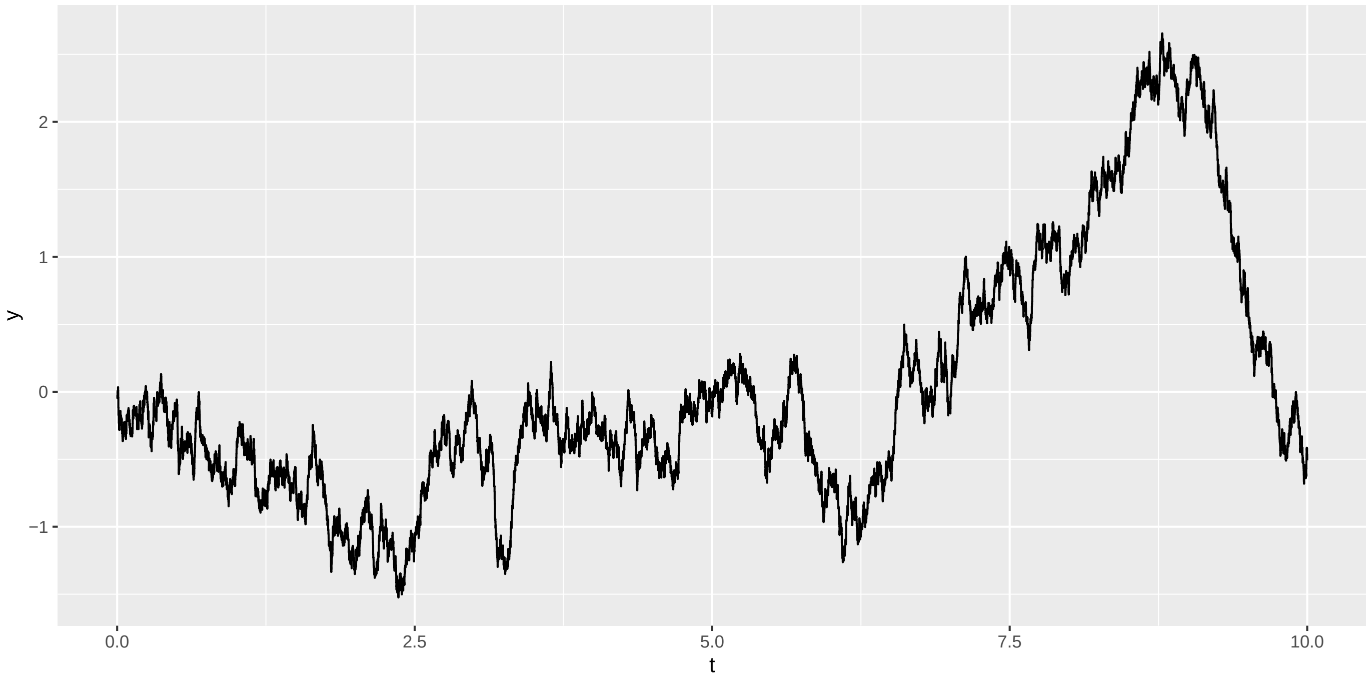
- $g(x) = x$: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = E[X]$, es decir $\bar{X}_n \longrightarrow E[X] = \mu_X$.
- $g(x) = (x-\mu_X)^2$: $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (X_i-\mu_X)^2 = E[(X-\mu_X)^2] = \mathrm{Var}(X)$.
- Si combinamos las dos convergencias anteriores: $\lim_{n\to\infty}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

1.24.1.2 Aplicaciones

Ejemplo: el movimiento Browniano

- Es proceso muy usado en la predicción de precios (opciones) en matemáticas financieras.
- Una caracterízación 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



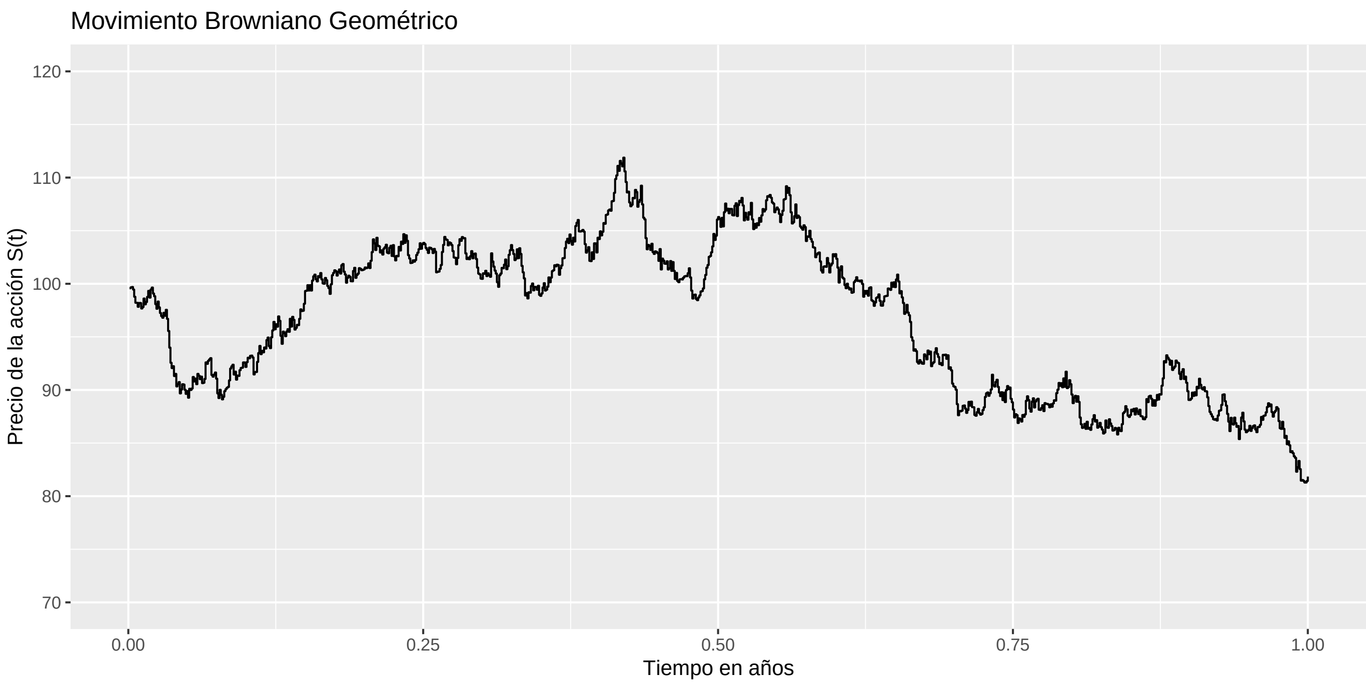
1.26 En finanzas, el modelo de Black-Scholes

El movimiento Browniano Geométrico:

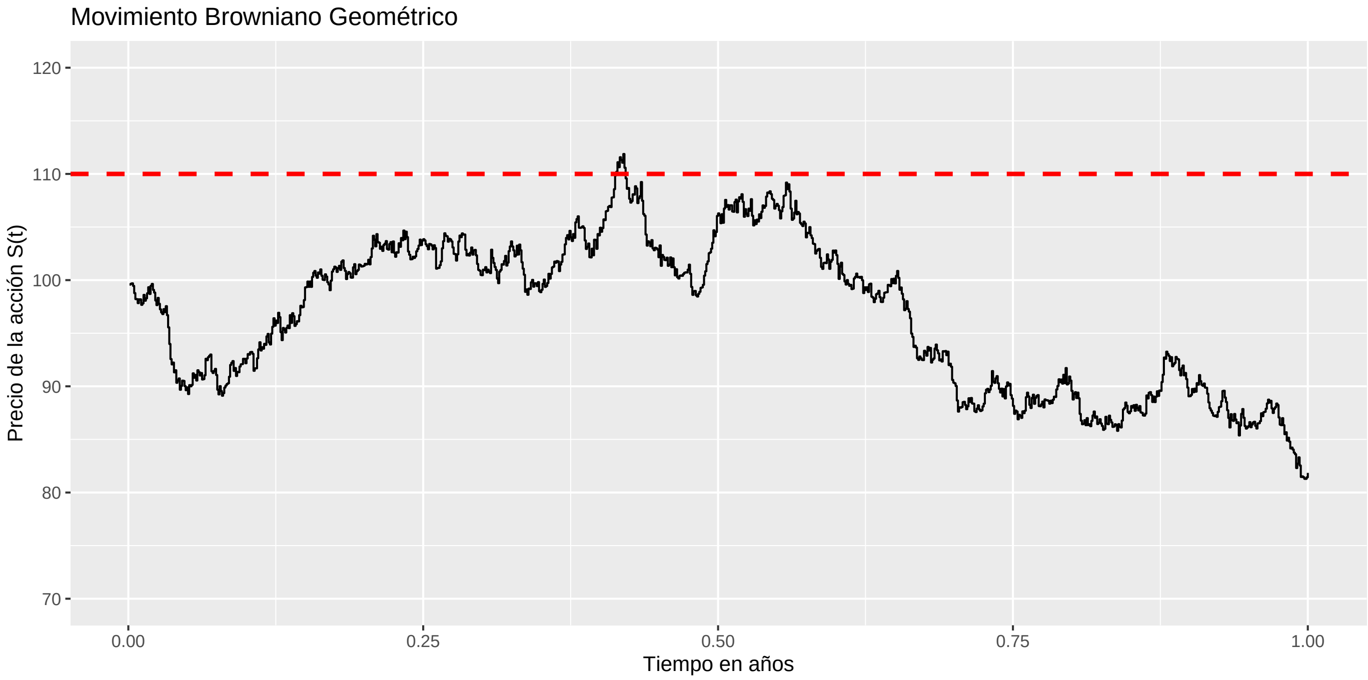
Lo propusieron Merton, Scholes y Black como modelo teórico para precios de acciones:

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

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

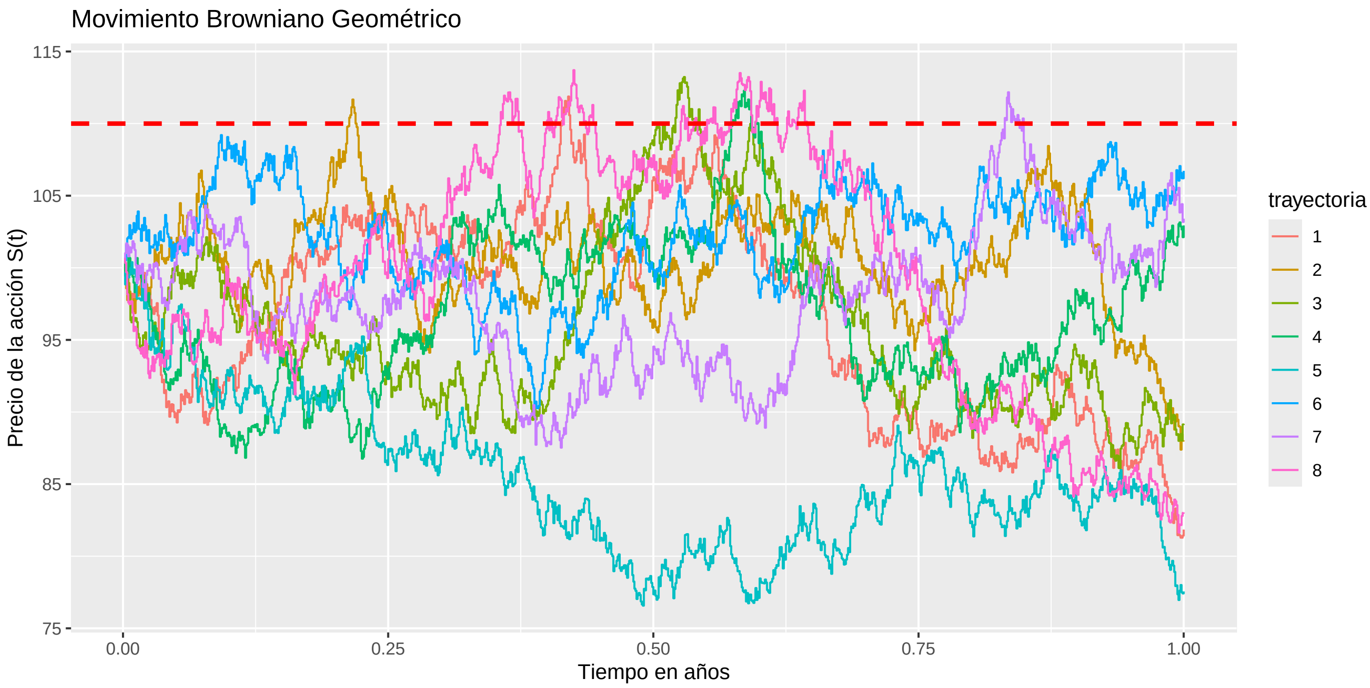


¿Cuándo observamos $S(t) \geq 100€$?



- 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



1.27 Simulación y método de Monte-Carlo

> **Simulación de variables aleatorias**
>
> - Lo que nos planteamos ahora es qué algoritmo podemos utilziar 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.

Inversa generalizada de la función de distribución o función cuantil

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

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

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

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

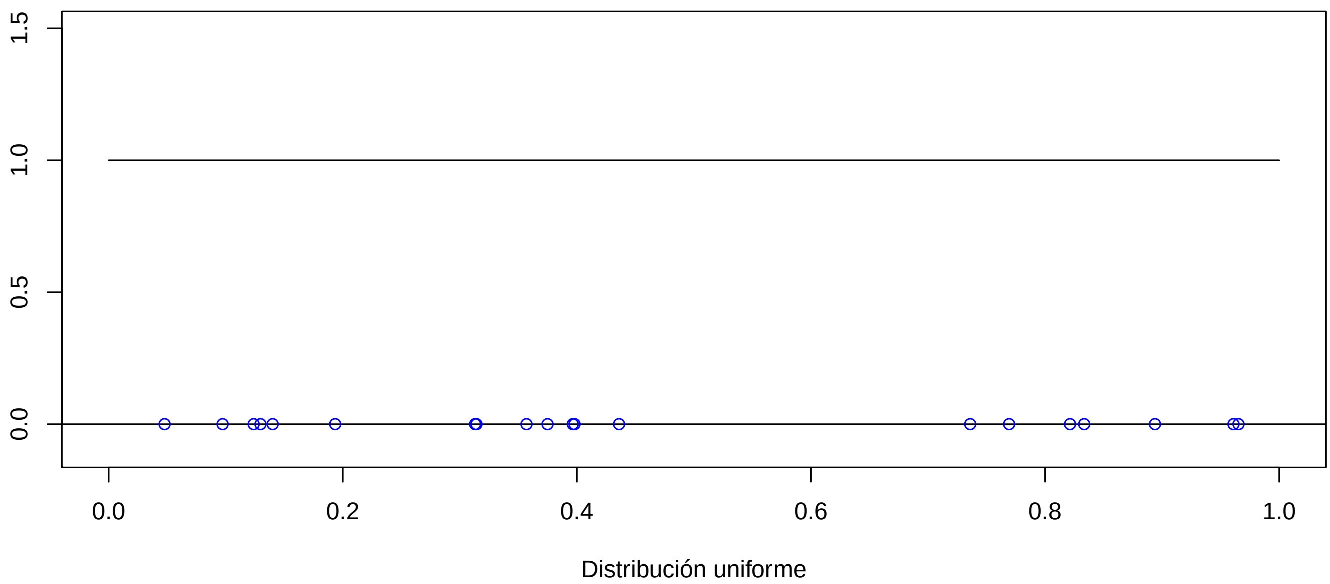
Método de la función inversa

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

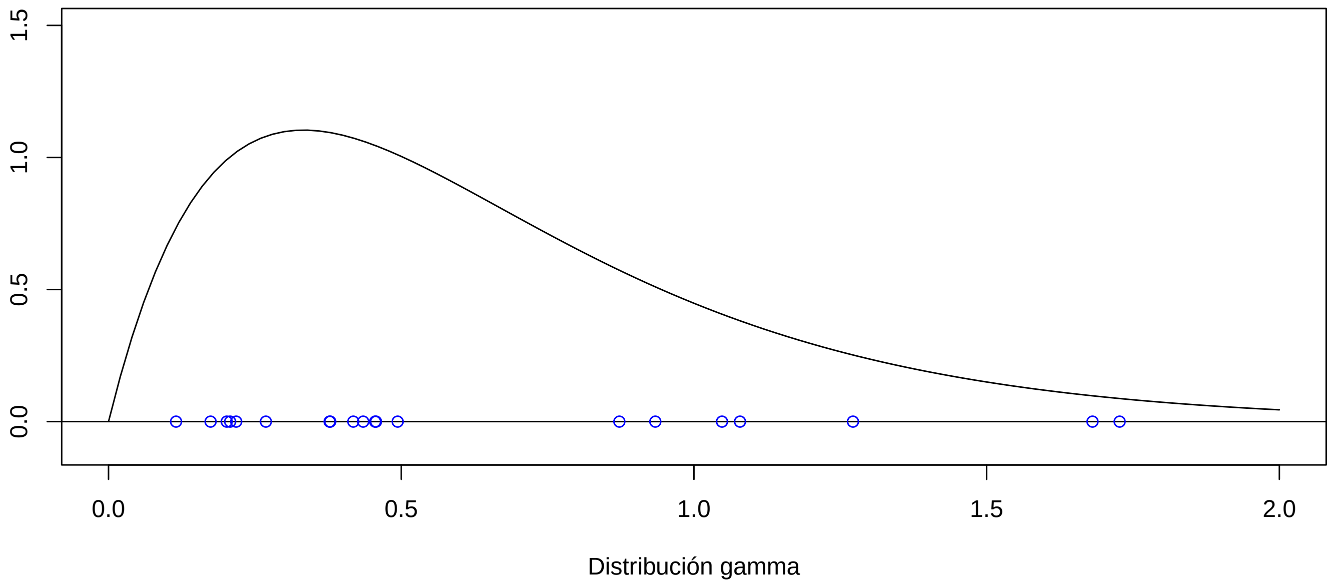
$$x_1 = F^{-1}(u_1), x_2 = F^{-1}(u_2), \ldots, x_n = F^{-1}(u_n)$$

  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 concoer 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 $\varphi : (a, b) \to \mathbb{R}$ tal que:*

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

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

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

Función características

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

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

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

Tip

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

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

Desigualdad de Markov

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

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

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

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

Desigualdad de Chebyshev

Si X es una variable aleatoria con media finita $\mu = E[X]$ y varianza $\sigma^2 = \mathrm{Var}(X) > 0$, entonces tomando $Z = \dfrac{(X-\mu)^2}{\sigma^2} \geq 0$ 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 concrentación 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 contrucción 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

i 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:

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

Ejemplos

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

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

$$\begin{array}{ll} X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) & \mu = E[X] \\ \theta = (\mu, \sigma^2), \quad p = 2 & \sigma^2 = \mathrm{Var}[X] = E[X^2] - (E[X])^2 \to E[X^2] = \sigma^2 + \mu^2 \\ \mu_1(\theta) = E[X] = \overline{x} & \hat{\mu} = \overline{x} \\ \mu_2(\theta) = E[X^2] = \overline{x^2} & \hat{\sigma}^2 = \overline{x^2} - (\overline{x})^2 \end{array}$$

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

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

2.6 Método de máxima verosimilitud

- El método más utilziado 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:*

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

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

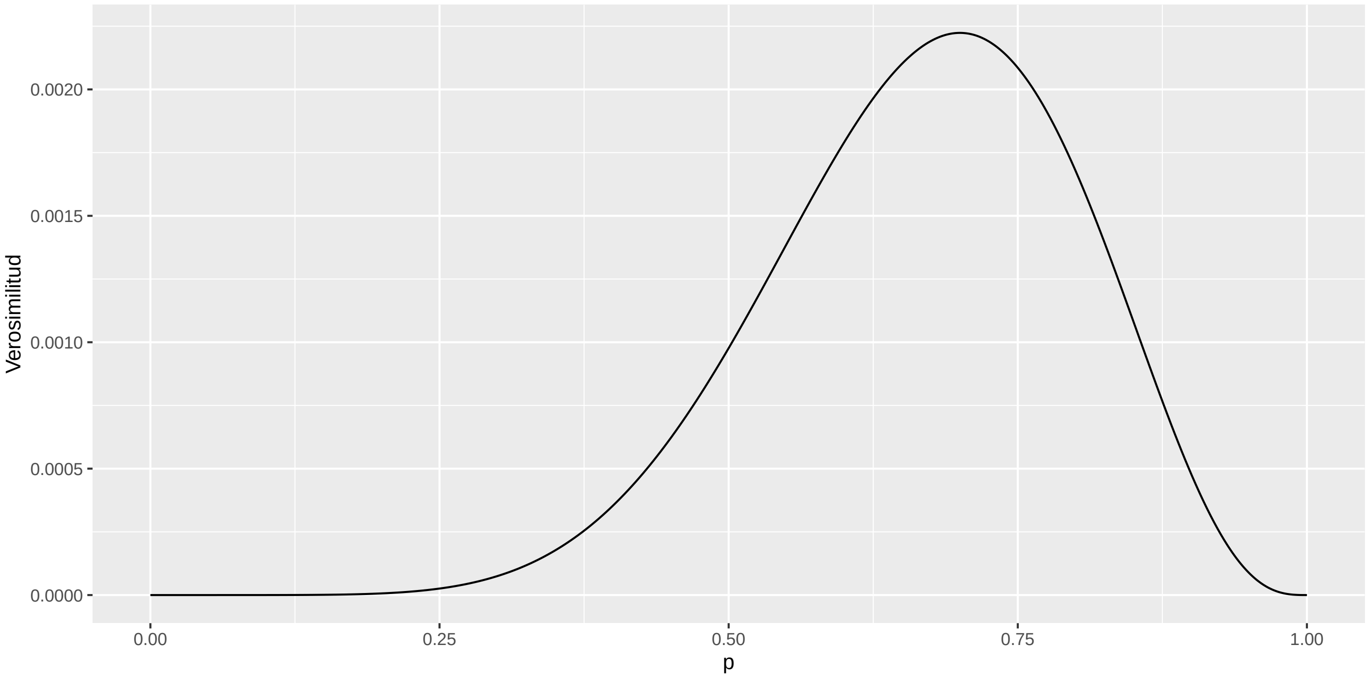
Ejemplo de cálculo de la verosimilitud

- Tiramos 10 veces una moneda (1 es cara, 0 es cruz), y obtenemos: 0,0,1,0,1,1,1,1,1,1.
- 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,*

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

> **!** 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:

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

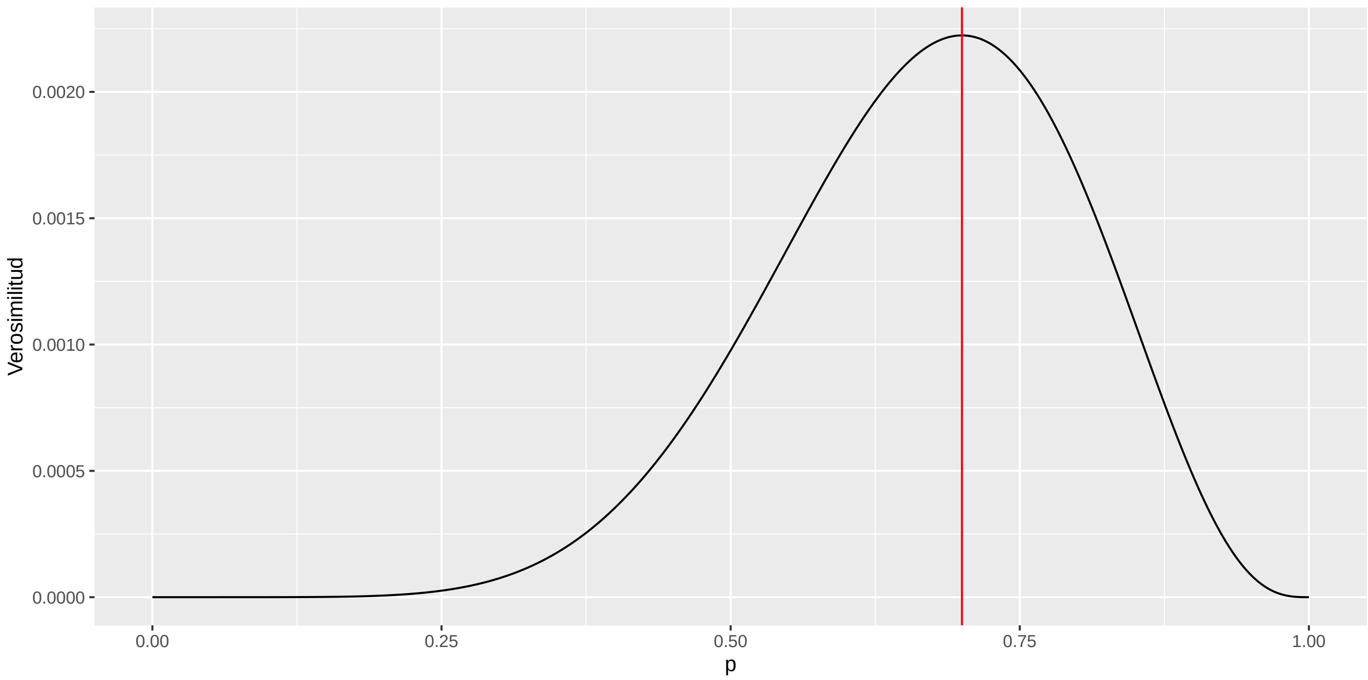


Figure 3: Verosimilitud correspondiente al ejemplo de 10 tiradas de una moneda.

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.

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

$(X_1, \ldots, X_N)$ m.a.s. de tamaño n

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

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

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

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

$$= \sum_{i=1}^{n} x_i - \cancel{p \cdot \sum_{i=1}^{n} x_i} - np + \cancel{p \cdot \sum_{i=1}^{n} x_i} = 0 \implies \boxed{\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}}$$

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

La función de denisdad 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, $X_1, X_2, \ldots, X_n$ 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:

$$\frac{\partial\ell(\mu,\sigma^2)}{\partial\mu}=\frac{\cancel{2}}{\cancel{2}\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0\to\sum_{i=1}^{n}(x_i-\mu)=0\to\boxed{\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}}$$

$$\frac{\partial\ell(\mu,\sigma^2)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0\to-n\sigma^2+\sum_{i=1}^{n}(x_i-\mu)^2=0\to\boxed{\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2}$$

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) $\implies$ permite manejar el riesgo y el error que podemos cometer al aproxima θ por $\hat{\theta}$ asociado.

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

- Su localización: sesgo.
- Su variabilidad: error cuadrático medio.
- Mencionaremos su comportamiento cuando $n\to\infty$.

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 $X\sim\mathcal{N}(\mu,\sigma^2)$.

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

$\widehat{\mu} = \bar{x}$ (estimador basado en los movimientos)

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

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