

Fundamentos de Inferencia Estadística

Francisco Javier Mercader Martínez

Índice

1	Muestreo y distribuciones muestrales	1
1.1	Introducción	1
1.2	Ejemplos	1
1.3	Surge una pregunta	1
1.4	Esbozo de respuesta: tasa de participación	2
1.5	Realización del experimento: conclusiones	3
1.6	En la práctica	3
1.7	Uso de la distribución muestral	3
1.8	Antes de extraer una muestra:	4
1.9	Otro ejemplo: valores muestrales de una distribución normal	4
1.10	Un resultado importante	5
1.11	Algunos términos	5
1.12	Ejemplos de estadísticos	5
1.13	La media muestral	6
1.13.1	Esperanza y varianza de la media muestral	6
1.14	Consecuencia práctica	7
1.14.1	Analogía con una diana	7
1.15	Varianza muestral	7
1.15.1	Dos apuntes	7
1.16	Esperanza de la varianza muestral	8
1.17	Distribuciones muestrales de \bar{X} y S^2	8
1.18	Distribución de \bar{X} y S^2 para una m.a.s. de una distribución normal	8
1.19	Recordatorio: distribución χ^2 con p grados de libertad	8
1.20	Distribución t-Student	9
1.21	Distribución F de Snedecor para el cociente de varianzas	10
1.22	Si la distribución de X no es Normal	11
1.23	Distribución muestral de la proporción muestral	12
1.24	Simulación y método de Monte-Carlo	13
1.24.1	Muestreo de Monte-Carlo para aproximar esperanzas	13
1.25	Movimiento Browniano	14
1.26	En finanzas, el modelo de Black-Scholes	14
1.27	Simulación y método de Monte-Carlo	16
2	Estimación	36
2.1	Introducción	36
2.2	Ejemplos de estimación paramétrica	36
2.3	Estimación paramétrica: estimación puntual	36
2.4	Métodos de construcción de estimadores	37

2.5	Método de los momentos	37
2.6	Método de máxima verosimilitud	38
2.7	Estimador de máxima verosimilitud	39
2.8	Métodos para evaluar un estimador	41
2.9	Sesgo	41
2.10	Error cuadrático medio	41
2.11	Balance entre sesgo y varianza	42
2.12	Comportamiento asintótico de un estimador	42
2.13	Consistencia	42
2.14	Normalidad asintótica	42
2.15	Estimación no paramétrica	43
2.16	Estimación tipo núcleo	44
2.16.1	Podemos variar el núcleo	45
2.16.2	Se pueden demostrar resultados asintóticos	46
2.17	Ancho de banda	46
2.17.1	Elección óptima del ancho de banda	47
2.17.1.1)	Aproximación gaussiana para el cálculo de $R(f'')$	47
2.17.1.2)	Aproximación no paramétrica para el cálculo de $R(f'')$, implementación en R	48
2.18	Introducción al Bootstrap	49
2.18.1	El error estándar Bootstrap	50
3	Estimación por intervalo	61
3.1	Introducción	61
3.2	Nivel de confianza	62
3.3	Un procedimiento general de construcción	62
3.4	Intervalo de confianza para la media μ de $X \sim \mathcal{N}(\mu, \sigma^2)$, σ conocida	62
3.4.1	Interpretación	63
3.5	Comentarios importantes	64
3.5.1	Factores que afectan a la precisión de la estimación	64
3.5.2	Determinación del tamaño muestral	64
3.5.3	Otros modelos, estadísticos pivotaes	64
3.6	Intervalos de confianza basados en el Bootstrap	66
3.6.1	Aplicación con el ejemplo de los ingresos	67
4	Contrastes de hipótesis	70
4.1	Introducción	70
4.2	Contrastes de hipótesis paramétricos	70
4.2.1	Elementos principales	70
4.2.2	Formulación general del contraste de hipótesis	70
4.2.2.1)	Errores	71
4.3	Métodos útiles para construir tests	73
4.3.1	Un primer procedimiento general	74
4.4	Contraste de hipótesis para la media μ	74
4.5	Procedimiento, hipótesis bilateral	74
4.6	Procedimiento, hipótesis unilateral	75
4.7	El p-valor	76
4.8	Test óptimo	78
4.9	Potencia de un test	78
4.10	Test óptimo como el de máxima potencia	78
4.11	Procedimientos basados en la verosimilitud	79

4.12	Constrastes de hipótesis simples	79
4.13	Hipótesis compuestas	80
4.14	Cociente de verosimilitudes monótono	81
4.15	Constrastes de hipótesis compuesta	81
4.16	Procedimiento basado en la verosimilitud	82
4.17	Test RVG para la media de una población normal con varianza conocida	82
4.18	Test de razón de verosimilitudes generalizada	83
4.19	Test de bondad de ajuste	84
5	Inferencia Bayesiana	89
5.1	Introducción	89
5.2	La inferencia bayesiana	89
5.3	La distribución a posteriori	89
5.4	Distribución a posteriori	91
5.5	Cambiamos la distribución a priori	92
5.5.1	La distribución $Be(a, b)$ es bastante flexible como a priori	92
5.6	Aplicaciones de la inferencia Bayesiana	93
5.6.1	Ilustración para un ejemplo simple de sistema de recomendación:	93

Tema 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 la moneda. X = resultado obtenido.

$$P(X = +) = p, P(X = -) = 1 - p$$

$$¿p = \frac{1}{2}?$$

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: aproximadamente 37 000 000. Escogemos aproximadamente 3000 personas.

Determinación de la concentración de un producto

- Quiero determinar la concentración de un producto.
- Formulo el modelo:
 - Experimento: "llevar a cabo una medición".
 - 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 estimación 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 convencerlos, 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 un conjunto 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 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 poblacion <- c(rep(1, 25900000), rep(0, 11100000))
2 set.seed(314159)
3 p_muestra <- mean(sample(poblacion, 3000, replace = FALSE))
4 p_muestra
```

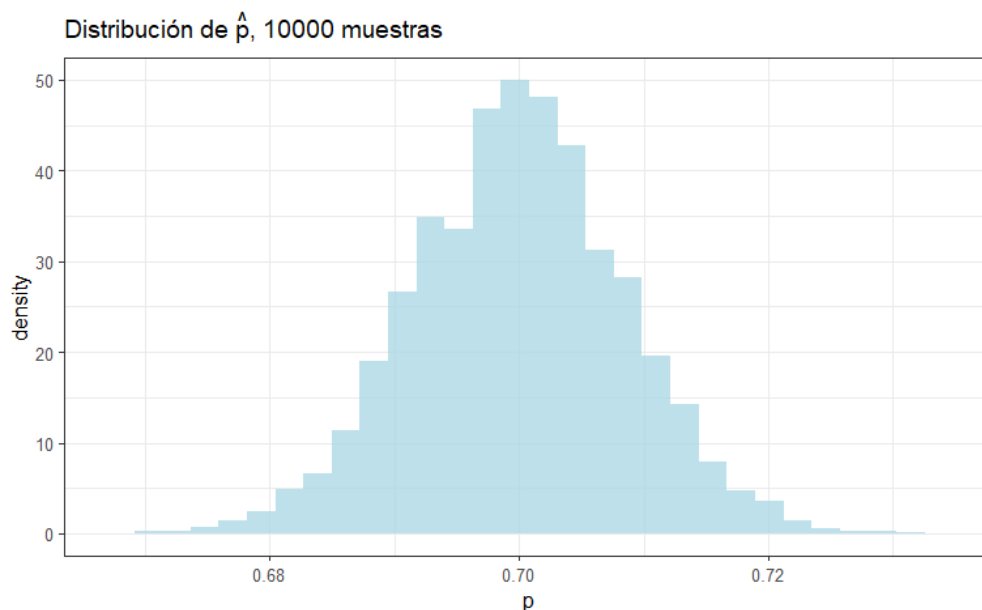
```
## [1] 0.705667
```

Queremos descartar que haya sido suerte. Vamos a repetir muchas veces (10000 veces por ejemplo), la extracción de una muestra de 3000 personas en la población.

```
1 library(tidyverse)
2 lista_muestras <- replicate(
3   10000,
4   sample(poblacion, 3000, replace = FALSE),
5   simplify = FALSE
6 )
7 p_muestras <- map_dbl(lista_muestras, mean)
8 head(p_muestras)
```

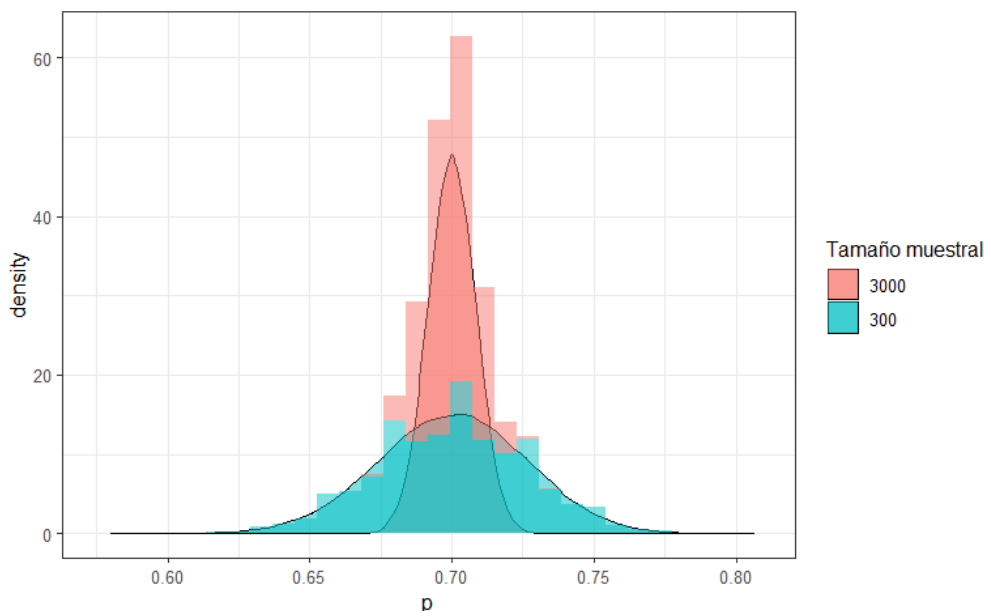
```
## [1] 0.6970000 0.7030000 0.7036667 0.7023333 0.7013333 0.7226667
```

Recogemos los valores obtenidos en un histograma.



1.5) Realización del experimento: conclusiones

- La enorme mayoría de las muestras de 3000 individuos proporcionan una tasa de partición 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 variables aleatoria: proporción de "1"s en la muestra escogida.
- ¿Cuál es la distribución de valores de \hat{p} ?

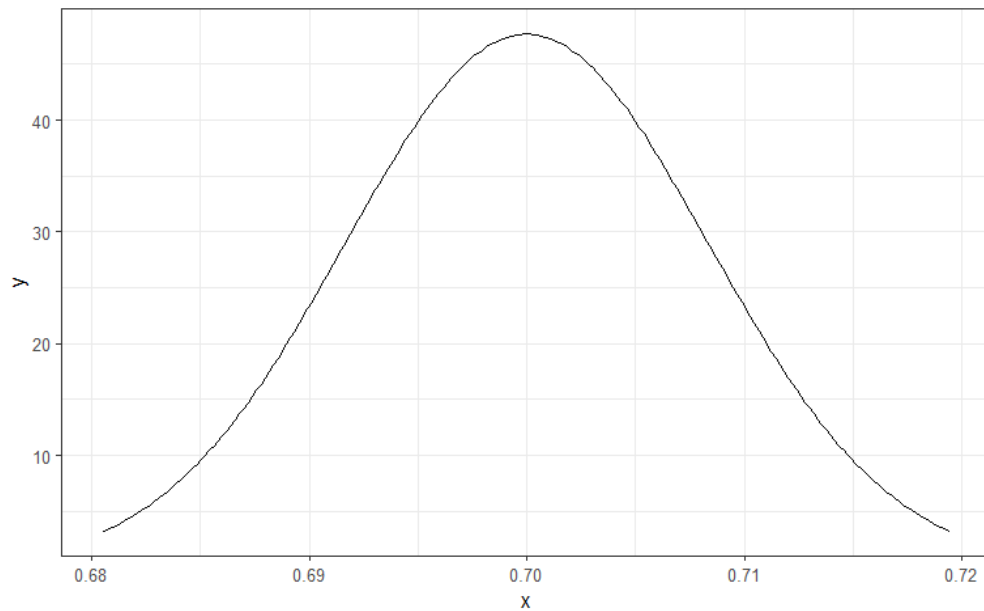
$$\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

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

1.7) Uso de la distribución muestral

La distribución muestral de \hat{p} :

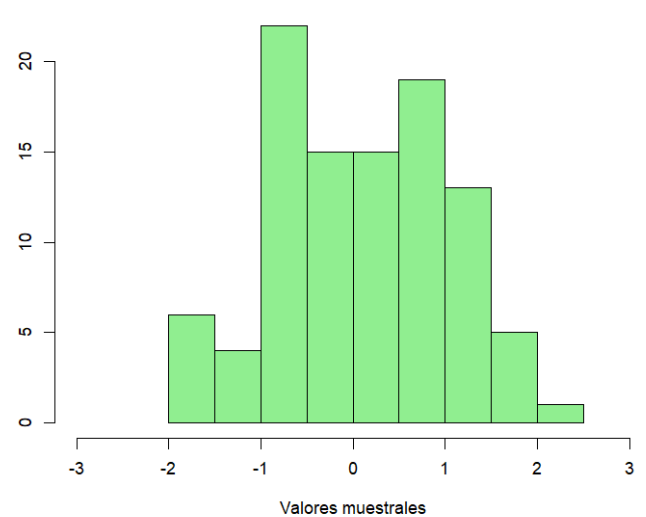
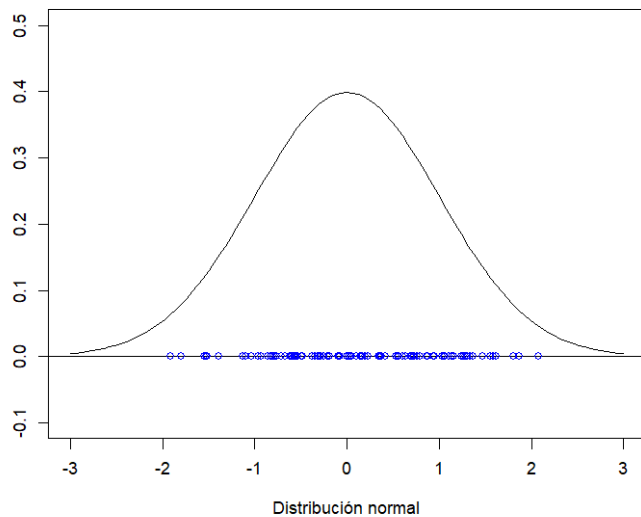
Es la distribución esperada de los valores de \hat{p} respecto a todas las muestras de ese tamaño que podría extraer.

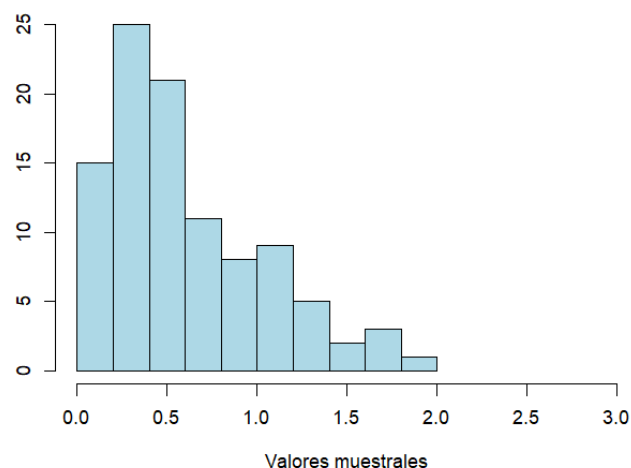
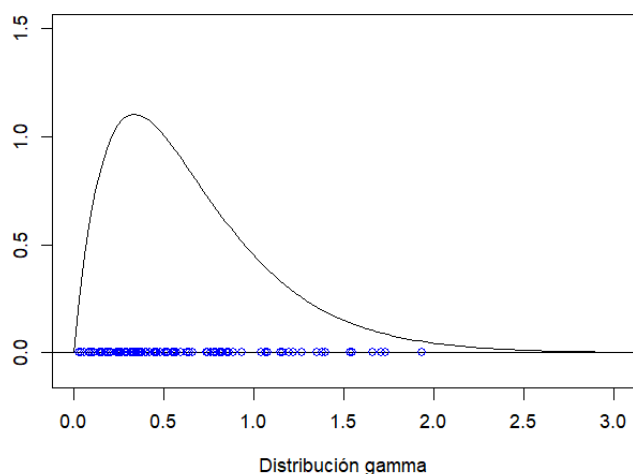


1.8) Antes de extraer una muestra:

- ¿Es suficiente el tamaño de la muestra para el riesgo asumible y la precisión requerida?
- Una vez extraída 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 aleatorias (v.v.aa) independientes con la misma 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

- 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 un 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 :

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

- 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 $\sigma^2 = \text{Var}(X)$.

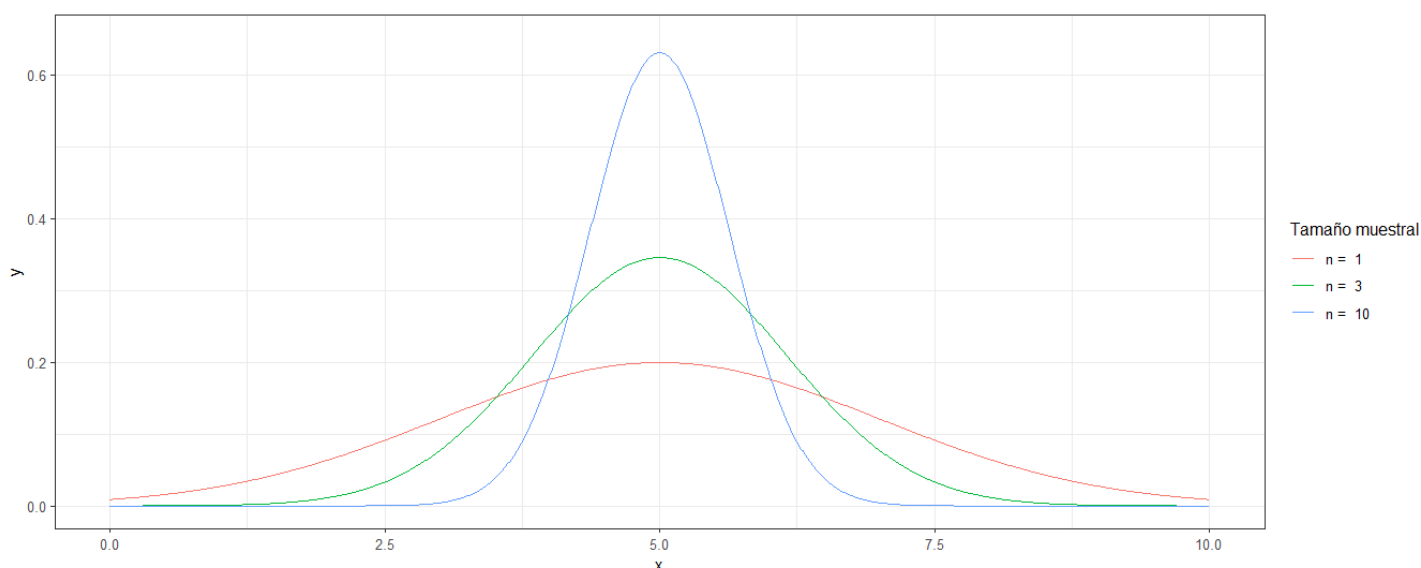
- Tenemos

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

→ 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 $n = 10$.

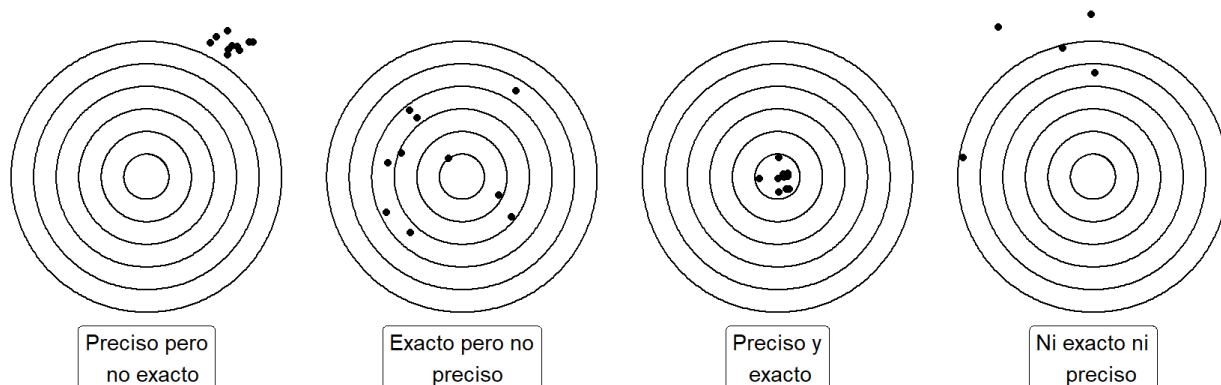


1.14) Consecuencia práctica

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

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

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

1.16) Esperanza de la varianza muestral

Proposición

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 contrastes 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

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

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

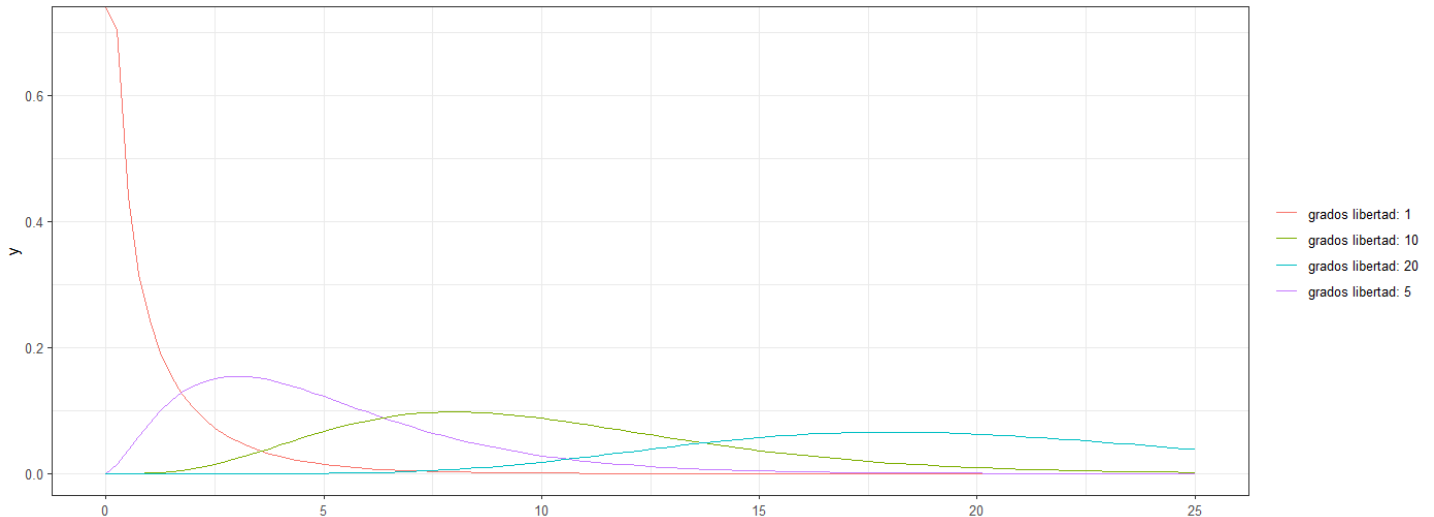
Si Z_1, \dots, Z_p son p variables aleatorias independientes, con $Z_i \sim \mathcal{N}(0, 1)$, entonces la variable aleatoria X definida como

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

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

Hemos visto, si X es Normal:

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

Si queremos centrarnos en μ es natural sustituir en ella σ por S_n .

Proposición

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

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

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 un t-Student con k grados de libertad:

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

donde Γ denota la función Gamma.

Caracterización de la t-Student como cociente

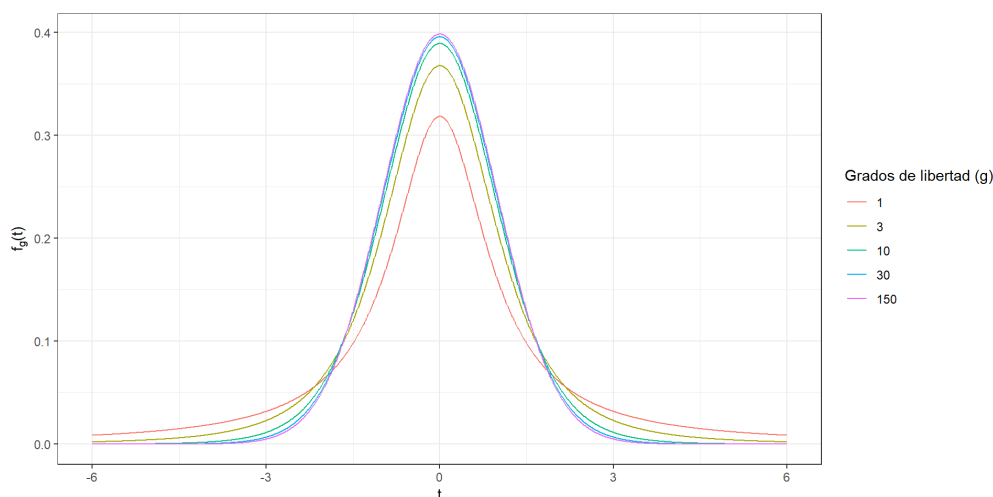
Si Z e Y son dos variables aleatorias independientes, con $Z \sim \mathcal{N}(0, 1)$ e $Y \sim \chi_p^2$, el cociente

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

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

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

Tiene colas más pesadas que una normal



1.21) Distribución F de Snedecor para el cociente de varianzas

Proposición

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

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

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

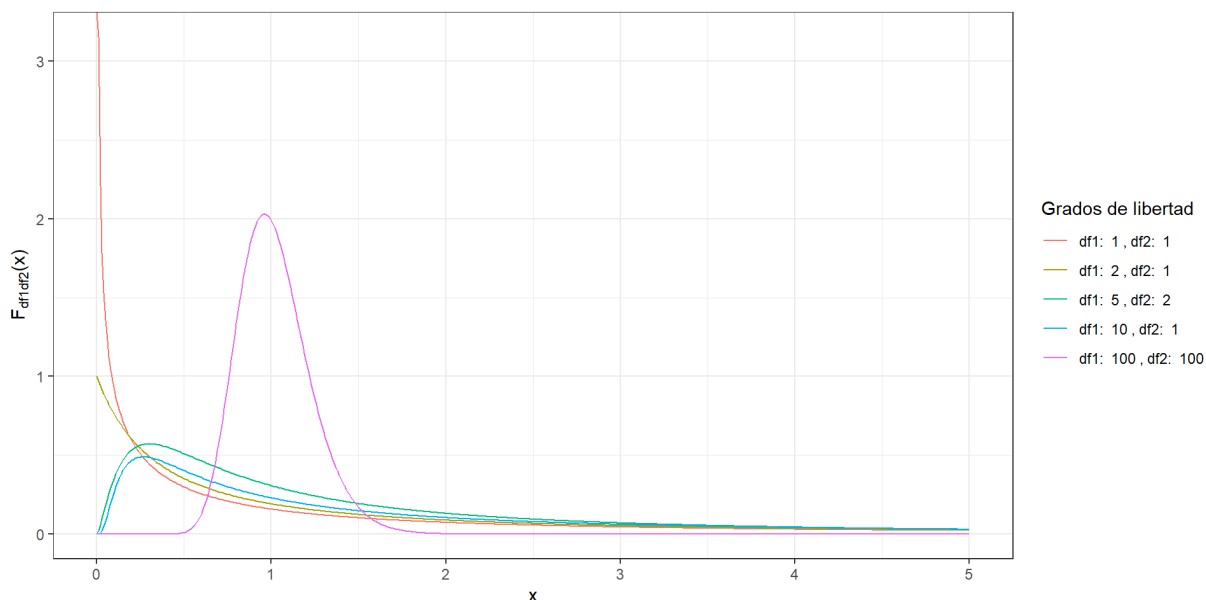
Consecuencia

Consideremos X e Y variables aleatorias normales independientes con varianzas σ_X^2 y σ_Y^2 , así como X_1, \dots, X_{n_x} e Y_1, \dots, Y_{n_y} dos muestras aleatorias simples de X e Y , respectivamente. Deducimos que

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

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

Depende mucho 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 $\frac{\sigma^2}{n}$:

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

Formulación matemática

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

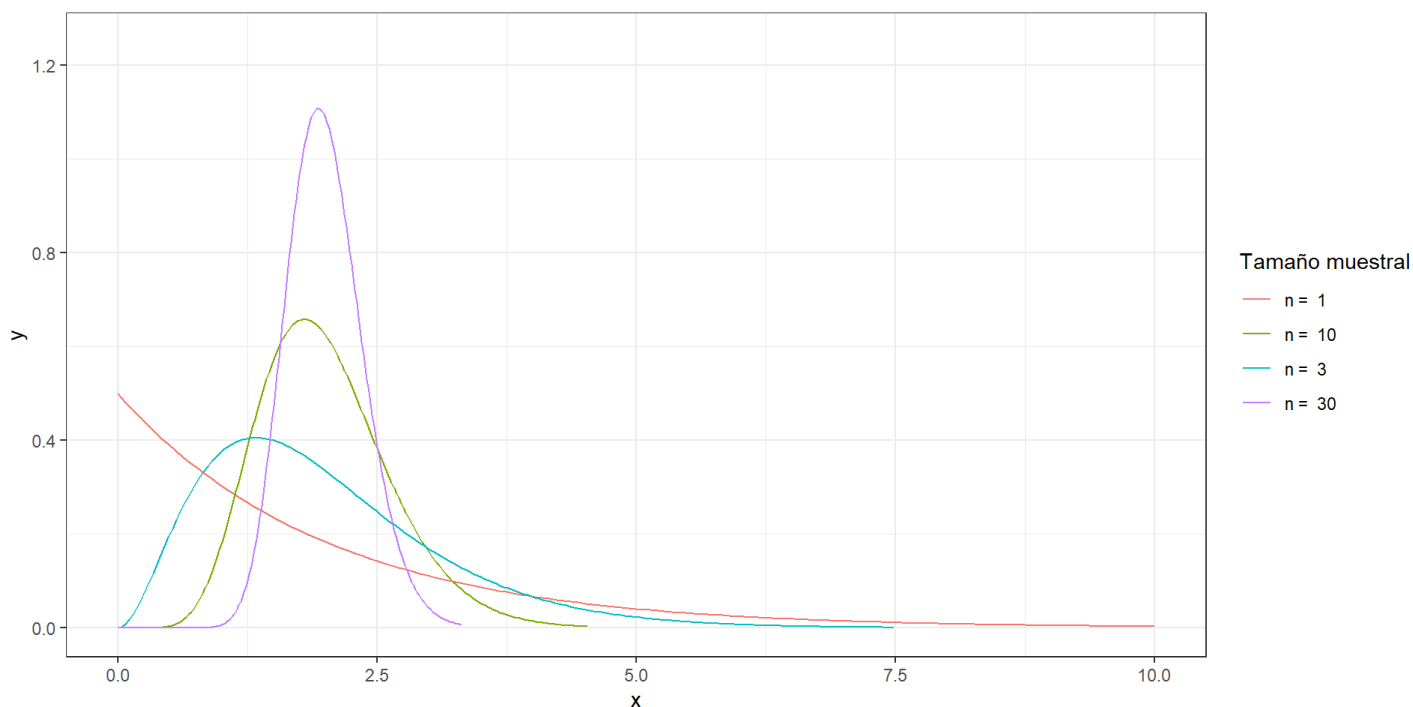
- ¿Cuándo considerar que n es grande?

Depende de la forma de la distribución de X :

- Si X casi Normal: n pequeño es suficiente.
- Si X es 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

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

Contamos N el número de "1"s.

- 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 $N = X_1 + X_2 + \dots + X_n$, 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:

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

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 hacen 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 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

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,

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

Ejemplos

- $g(x) = x$: $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = E[X]$, es decir $\bar{X}_n \rightarrow E[X] = \mu_X$.

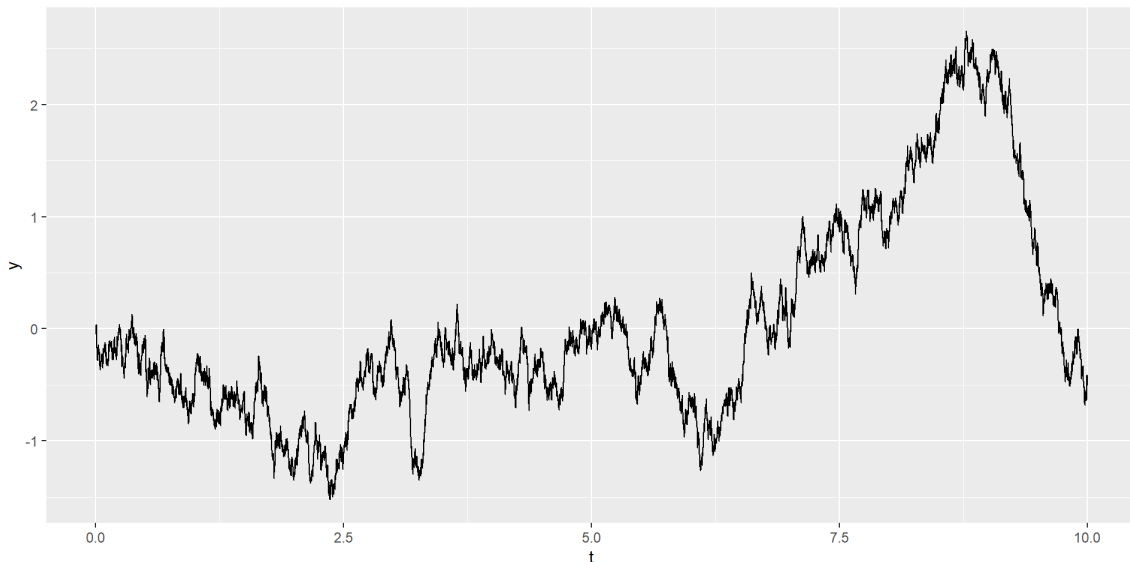
- $g(x) = (x - \mu_X)^2 : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)^2 = E[(X - \mu_X)^2] = \text{Var}(X)$.
- Si combinamos las dos convergencias anteriores: $\lim_{n \rightarrow \infty} \frac{1}{n} (X_i - \bar{X}_n)^2 = \text{Var}(X)$, es decir, $S_n^2 \rightarrow \text{Var}(X)$.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

Aplicaciones

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



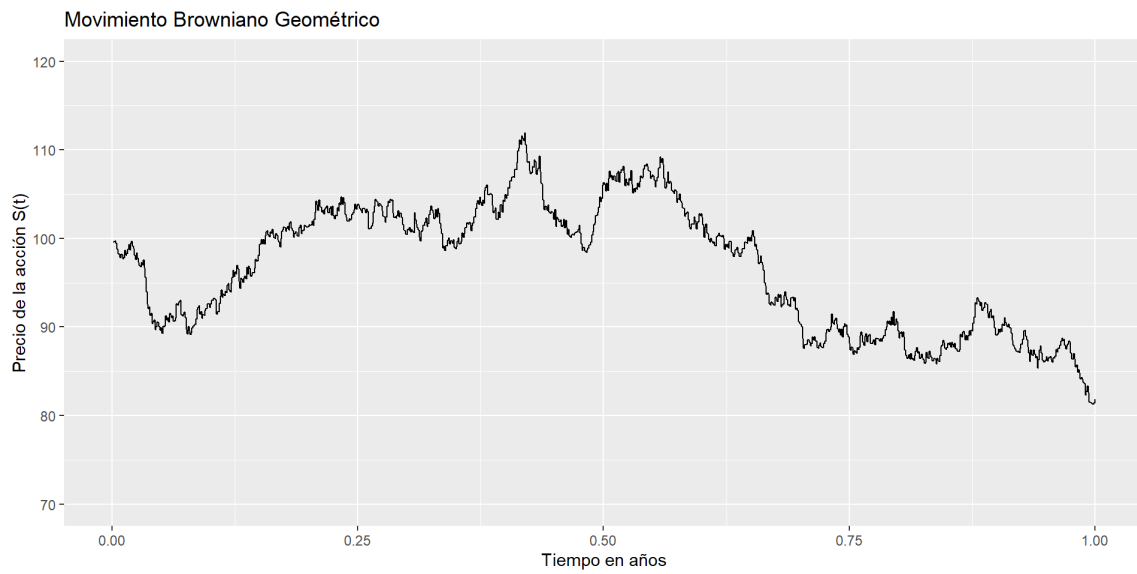
1.26) En finanzas, el modelo de Black-Scholes

El movimiento Browniano Geométrico:

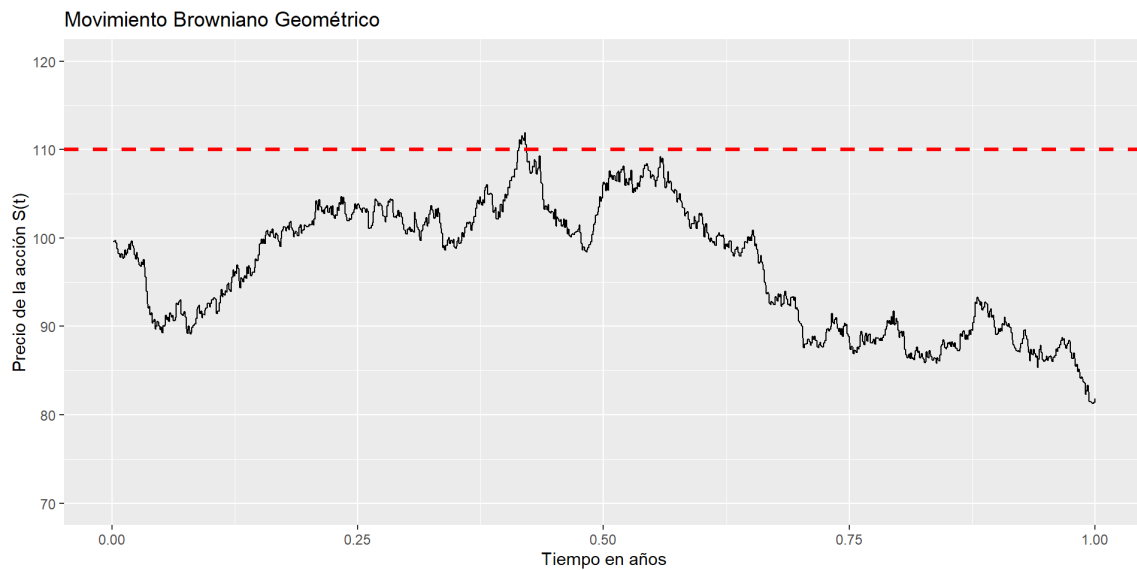
Lo propusieron Merton, Scholes y Black como modelo teórico para precios de acciones:

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

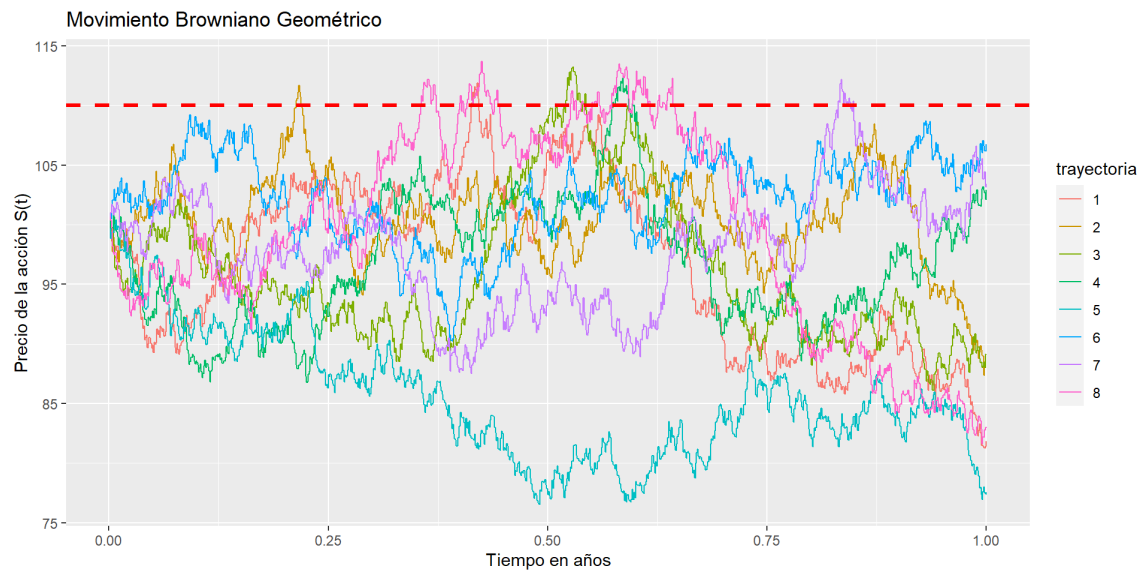
- S_0 : precio inicial de la acción.
- μ : el drift.
- σ^2 : la volatilidad



¿Cuán observaremos $S(t) \geq 110\text{€}$?



- 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.
- Representamos las 8 primeras



- ¿Cuál es la probabilidad de que alcance el umbral antes de un año?

Por Monte-Carlo:

Simulamos 1000 trayectorias hasta el año y aproximamos la probabilidad que nos interesa por la proporción de trayectorias que alcanzan los 110 euros.

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 conocemos usualmente. Para ello es necesaria 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

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

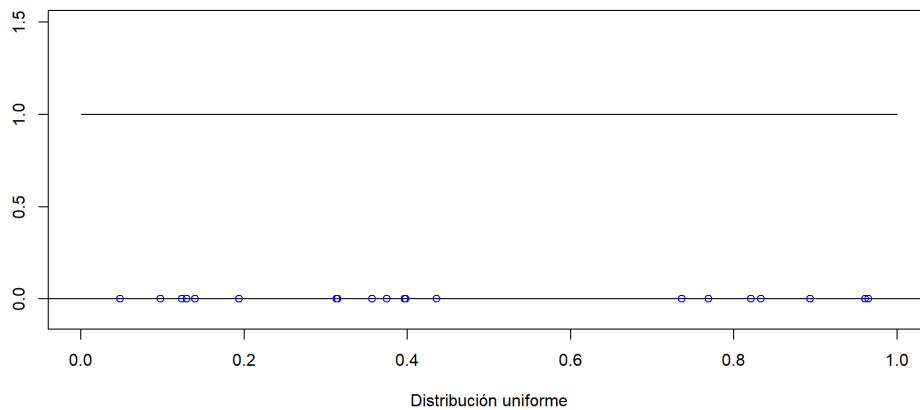
Método de la función inversa

- 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

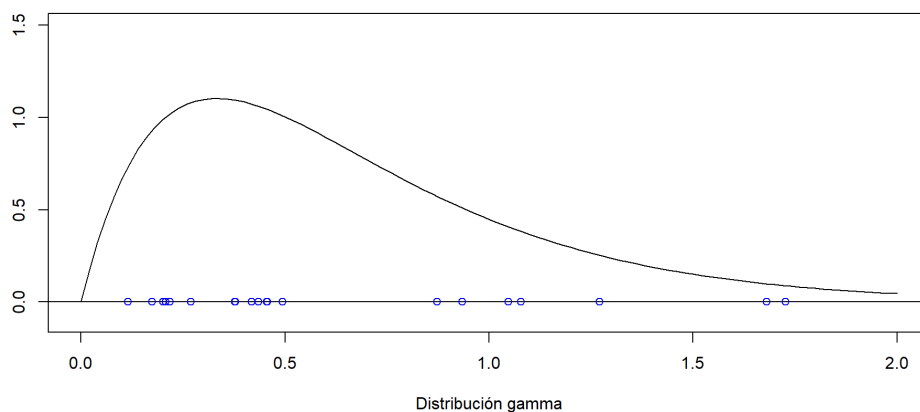
$$x_1 = F^{-1}(u_1), x_2 = F^{-1}(u_2), \dots, x_n = F^{-1}(u_n)$$

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 conocer cómo se comporta una transformación de X , $Y = \varphi(X)$.

Teorema

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 característica

Definición

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 $i = \sqrt{-1}$

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

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

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

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

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 $r > 0$.

Hoja de ejercicios Tema 1: Muestreo y distribuciones muestrales

- 1) Sea X variable aleatoria con distribución Bernoulli, de parámetro $p(X \sim b(p))$ y sea X_1, X_2, X_3 una muestra aleatoria simple (m.a.s) de X . Se pide:

- a) Estudiar la distribución del vector (X_1, X_2, X_3) .

Dado que X_1, X_2, X_3 son una muestra aleatoria simple de $X \sim \text{Bernoulli}(p)$, entonces las siguientes propiedades son ciertas:

- 1) X_1, X_2, X_3 son independientes e idénticamente distribuidas.
- 2) Cada $X_i \sim \text{Bernoulli}(p)$, es decir, la probabilidad de éxito $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$.

El vector (X_1, X_2, X_3) tiene una **distribución multinomial** con 2 posibles resultados (0 o 1) para cada componente. La distribución conjunta es:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = p^{x_1+x_2+x_3}(1-p)^{3-(x_1+x_2+x_3)},$$

donde $x_1, x_2, x_3 \in \{0, 1\}$.

Esto corresponde a la **distribución conjunta** de 3 variables Bernoulli independientes.

- b) Estudiar la distribución en el muestreo del estadístico $\frac{X_1 + X_2 + X_3}{3}$.

Es estadístico $\bar{X} = \frac{X_1 + X_2 + X_3}{3}$ es el **promedio muestral** de las 3 variables. Para analizar su distribución:

- 1) $S = X_1 + X_2 + X_3$ sigue una **distribución binomial** porque es la suma de $n = 3$ variables Bernoulli independientes:

$$S \sim B(n = 3, p).$$

La función de probabilidad de S es:

$$P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

- 2) El estadístico $\bar{X} = \frac{S}{3}$ simplemente escala los valores posibles de S dividiéndolos por 3. Los valores posibles de \bar{X} son:

$$\bar{X} \in \left\{0, \frac{1}{3}, \frac{2}{3}, 1\right\}.$$

- 3) La probabilidad de cada valor de \bar{X} es proporcional a la probabilidad de los valores correspondientes de S :

$$P\left(\bar{X} = \frac{k}{3}\right) = P(S = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k = 0, 1, 2, 3.$$

Por lo tanto, la distribución de \bar{X} es discreta y está determinada por la distribución binomial de S .

- 2) Sea X variable aleatoria con distribución $\text{Exp}(\lambda)$. Sea (X_1, \dots, X_n) una m.a.s de X , estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Dado que $X \sim \text{Exp}(\lambda)$ (exponencial con parámetro $\lambda > 0$), y que (X_1, \dots, X_n) es una muestra aleatoria simple (m.a.s) de X , podemos analizar la distribución del estadístico $S = \sum_{j=1}^n X_j$.

Propiedades relevantes:

- 1) **Distribución de la suma de variables exponenciales independientes:** Si X_1, \dots, X_n son variables aleatorias

independientes e idénticamente distribuidas ($X_i \sim \text{Exp}(\lambda)$), entonces la suma:

$$S = \sum_{j=1}^n X_j$$

sigue una distribución **Gamma** con parámetros n y λ . Esto se denota como:

$$S \sim \text{Gamma}(n, \lambda),$$

donde:

- n es el parámetro de forma.
- λ es el parámetro de escala.

Distribución Gamma:

La función de densidad de probabilidad de una variables aleatoria $S \sim \text{Gamma}(n, \lambda)$ está dada por:

$$f_S(s) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} s^{n-1} e^{-\lambda s} & s > 0 \\ 0 & s \leq 0 \end{cases}$$

donde:

- $\Gamma(n)$ es la función gamma (para $n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$).
- s^{n-1} y $e^{-\lambda s}$ controlan la forma y decaimiento de la densidad.

Propiedades del estadístico S :

1) Esperanza ($E[S]$):

Si $S \sim \text{Gamma}(n, \lambda)$, entonces:

$$E[S] = \frac{n}{\lambda}.$$

2) Varianza ($\text{Var}(S)$):

La varianza de S está dada por:

$$\text{Var}(S) = \frac{n}{\lambda^2}.$$

3) Caso especial ($n = 1$):

Cuando $n = 1$, la distribución Gamma coincide con la distribución exponencial. Es decir:

$$\text{Gamma}(1, \lambda) = \text{Exp}(\lambda).$$

3) Sea X variable aleatoria con distribución $\mathcal{N}(\mu, \sigma^2)$. Sea (X_1, \dots, X_n) una m.a.s de X , estudiar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Dado que $X \sim \mathcal{N}(\mu, \sigma^2)$, y que (X_1, \dots, X_n) es una muestra aleatoria simple (m.a.s) de X , las X_i son independientes e idénticamente distribuidas con $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Queremos analizar la distribución en el muestreo de $S = \sum_{j=1}^n X_j$.

Propiedades relevantes:

- 1) **Suma de variables normales independientes:** Si X_1, \dots, X_n son independientes y $X_i \sim \mathcal{N}(\mu, \sigma^2)$, entonces la suma:

$$S = \sum_{j=1}^n X_j$$

sigue una distribución normal:

$$S \sim N(n\mu, n\sigma^2).$$

Derivación:

- 1) **Esperanza** ($E[S]$): La esperanza de S es la suma de las esperanzas de las X_i :

$$E[S] = E\left[\sum_{j=1}^n X_j\right] = \sum_{j=1}^n E[X_j] = \sum_{j=1}^n \mu = n\mu.$$

- 2) **Varianza** ($\text{Var}(S)$): La varianza de S es la suma de las varianzas de las X_i , ya que son independientes:

$$\text{Var}(S) = \text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) = \sum_{j=1}^n \sigma^2 = n\sigma^2.$$

- 3) **Distribución**: Dado que una combinación lineal de variables normales independientes también sigue una distribución normal, se concluye que:

$$S \sim N(n\mu, n\sigma^2).$$

- 4) Sea X una variable aleatoria con función de densidad:

$$f(x, \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0, +\infty)}(x).$$

Obtener la distribución en el muestreo estadístico:

$$T(X_1, X_2, \dots, X_n) = \sum_{j=1}^n X_j^2.$$

Obtener su media y su varianza.

Paso 1: Verificar la distribución de X

La función de densidad de X es:

$$f(x; \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0, \infty)}(x).$$

Esta densidad corresponde a una **distribución Rayleigh generalizada** con parámetro de escala θ . Para esta distribución:

- X^2 sigue una distribución exponencial con parámetro $\lambda = \frac{1}{\theta}$.

Entonces:

$$Y = X^2 \sim \text{Exp}\left(\lambda = \frac{1}{\theta}\right).$$

Paso 2: Distribución del estadístico $T = \sum_{j=1}^n X_j^2$

Dado que $Y_j = X_j^2 \sim \text{Exp}\left(\frac{1}{\theta}\right)$, y las Y_j son independientes, la suma de n variables exponenciales independientes sigue una distribución **Gamma**.

Por lo tanto, el estadístico:

$$T = \sum_{j=1}^n X_j^2 = \sum_{j=1}^n Y_j$$

sigue la distribución:

$$T \sim \text{Gamma}\left(n, \lambda = \frac{1}{\theta}\right),$$

donde:

- n es el parámetro de forma.
- $\lambda = \frac{1}{\theta}$ es el parámetro de escala.

La densidad de la distribución Gamma es:

$$f_T(t; n, \lambda) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{\Gamma(n)}, \quad t > 0.$$

Paso 3: Esperanza y Varianza del estadístico T

Para una distribución Gamma con parámetros (n, λ) , las propiedades son:

1) Esperanza:

$$E[T] = \frac{n}{\lambda}.$$

2) Varianza:

$$\text{Var}(T) = \frac{n}{\lambda^2}.$$

En este caso, como $\lambda = \frac{1}{\theta}$:

$$\begin{aligned} E[T] &= \frac{n}{\frac{1}{\theta}} = n\theta, \\ \text{Var}(T) &= \frac{n}{\left(\frac{1}{\theta}\right)^2} = n\theta^2. \end{aligned}$$

5) Sea X una variable aleatoria con función de densidad:

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \dots, X_n) = \frac{\sum_{j=1}^n \ln(1+X_j)}{n}.$$

Obtener su media y su varianza.

1) Identificación de la distribución de X

La función de densidad de X viene dada por:

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x), \quad \theta > 0.$$

Observemos que, para $x > 0$, la forma $\frac{1}{(1+x)^{1+\theta}}$ sugiere una transformación logarítmica conveniente:

$$Y = \ln(1+X).$$

Vamos a encontrar la distribución de Y .

2) Transformación $Y = \ln(1+X)$

1) Relación entre X y Y :

$$Y = \ln(1+X) \longleftrightarrow X = e^Y - 1.$$

2) Soporte:

Dado que $x > 0$, entonces $1+x > 1$ y por ende $Y > \ln(1) = 0$. De modo que $Y \in (0, +\infty)$.

3) Derivada $\frac{dx}{dy}$:

$$\frac{dx}{dy} = \frac{d}{dy}(e^Y - 1) = e^Y.$$

4) Función de densidad de Y :

Partiendo de que $f_X(x, \theta)$ es la densidad de X , la densidad de Y se obtiene mediante:

$$f_Y(y) = f_X(x(y)) \cdot \left| \frac{dx}{dy} \right|.$$

Sustituyendo $x = e^Y - 1$, obtenemos:

$$f_X(e^Y - 1, \theta) = \frac{\theta}{(1 + (e^Y - 1))^{1+\theta}} = \frac{\theta}{(e^Y)^{1+\theta}} = \theta e^{-(1+\theta)Y}.$$

Por último, multiplicamos por $\frac{dx}{dy} = e^Y$:

$$f_Y(y) = \left[\theta e^{-(1+\theta)Y} \right] \cdot e^Y = \theta e^{-(1+\theta)Y} e^Y = \theta e^{-\theta Y}, \quad y > 0.$$

Esta es precisamente la **densidad de una Exponencial** con parámetro θ . Por tanto,

$$Y = \ln(1 + X) \sim \text{Exp}(\theta).$$

3) Distribución del estadístico

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^n \ln(1 + X_j).$$

Definamos

$$Y_j = \ln(1 + X_j).$$

Cada Y_j es $\text{Exp}(\theta)$ y son independientes e idénticamente distribuidas al provenir de una muestra aleatoria simple de X . Entonces

$$T = \frac{1}{n} \sum_{j=1}^n Y_j.$$

- La suma $S = \sum_{j=1}^n Y_j$ sigue una $\text{Gamma}(n, \theta)$.

- El estadístico $T = \frac{1}{n}S$ es simplemente la suma escalada por $\frac{1}{n}$.

1) Si $S \sim \Gamma(n, \theta)$, entonces $T = \frac{1}{n}S$ tiene parámetro de forma n y de **tasa** $n\theta$. En otras palabras,

$$T \sim \Gamma(n, n\theta).$$

Explícitamente, su función de densidad es:

$$f_T(t) = \frac{(n\theta)^n}{\Gamma(n)} t^{n-1} e^{-n\theta t}, \quad t > 0.$$

4) Media y varianza de T

Para una $\Gamma(k, \lambda)$, se sabe que:

$$E[W] = \frac{k}{\lambda}, \quad \text{Var}(W) = \frac{k}{\lambda^2}.$$

En nuestro caso, $T \sim \Gamma(n, n\theta)$. Luego:

1) Media de T :

$$E[T] = \frac{n}{n\theta} = \frac{1}{\theta}$$

2) Varianza de T :

$$\text{Var}(T) = \frac{n}{(n\theta)^2} = \frac{1}{n\theta^2}.$$

6) Sea X una variable aleatoria con función de densidad en todo \mathbb{R} :

$$f(x, \theta) = \exp(-(x - \theta)) \exp(-\exp(-(x - \theta))).$$

Obtener la distribución en el muestreo del estadístico:

$$T(X_1, X_2, \dots, X_n) = \frac{\sum_{j=1}^n \exp(-X_j)}{n}.$$

Obtener su media y su varianza.

1) Identificar la distribución de X

La función de densidad que se nos da es, para todo $x \in \mathbb{R}$,

$$f(x, \theta) = \exp(-(x - \theta)) \exp(-\exp(-(x - \theta))).$$

Obsérvese que si definimos

$$Z = X - \theta,$$

la densidad Z queda

$$f_Z(z) = \exp(-z) \exp(-e^{-z}), \quad z \in \mathbb{R}.$$

Esta es la **distribución Gumbel** estándar (con parámetro de localización 0 y escala 1) pa el máximo. Por lo tanto, X se distribuye como una $\text{Gumbel}(\theta, 1)$ con localización θ y escala 1.

En resumen:

$$X \sim \text{Gumbel}(\theta, 1) \longleftrightarrow Z = X - \theta \sim \text{Gumbel}(0, 1).$$

2) Analizar el estadístico

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{j=1}^n \exp(-X_j).$$

Definamos

$$Y_j = \exp(-X_j).$$

El objetivo es estudiar la distribución de

$$T = \frac{1}{n} \sum_{j=1}^n Y_j = \frac{1}{n} \sum_{j=1}^n e^{-X_j}.$$

2.1) Distribución de $Y_j = e^{-X_j}$

Usaremos el hecho de que $Z_j = X_j - \theta$ es Gumbel estándar. Entonces

$$X_j = Z_j + \theta \longrightarrow e^{-X_j} = e^{-\theta} e^{-Z_j}.$$

Definamos $W_j = e^{-Z_j}$. Veamos la distribución de W_j :

- Si $Z_j \sim \text{Gumbel}(0, 1)$, la función de distribución acumulativa (CDF) de Z_j es

$$F_{Z_j}(z) = e^{-e^{-z}}, \quad z \in \mathbb{R}.$$

- Entonces

$$W_j = e^{-Z_j} > 0.$$

Para $w > 0$,

$$\{W_j \leq w\} \equiv \{e^{-Z_j} \leq w\} \equiv \{Z_j \geq -\ln w\}.$$

Por tanto,

$$F_{W_j}(w) = \Pr(Z_j \geq -\ln w) = 1 - \Pr(Z_j < -\ln w) = 1 - F_{Z_j}(-\ln w).$$

Usando $F_{Z_j}(z) = e^{-e^{-z}}$, se obtiene

$$F_{Z_j}(-\ln w) = e^{-\exp(-(-\ln w))} = e^{-w}.$$

Por lo tanto,

$$F_{W_j}(-\ln w) = e^{-w}, \quad w > 0,$$

que es la función de distribución acumulativa de una **Exponencial**.

En consecuencia,

$$W_j \sim \text{Exp}(1)$$

Como

$$Y_j = e^{-\theta} W_j,$$

entonces $Y - j$ es simplemente W_j escalada por $e^{-\theta}$.

- Si $W_j \sim \text{Exp}(1)$, entonces la variable $c \cdot W_j$ con $c > 0$ es $\text{Exp}\left(\frac{1}{c}\right)$.
- Aquí $c = e^{-\theta} \longrightarrow \frac{1}{c} = e^{\theta}$.

Por tanto:

$$Y_j = e^{-\theta} W_j \sim \text{Exp}(e^{\theta}).$$

Es decir, cada Y_j tiene tasa e^{θ} .

2.2) Distribución de la media muestral T

Dado que los Y_j son independientes e idénticamente distribuidos $\text{Exp}(e^{\theta})$, la suma

$$S = \sum_{j=1}^n Y_j$$

sigue una distribución **Gamma** con forma n y tasa e^{θ} ; escribimos

$$S \sim \Gamma(n, e^{\theta}).$$

El estadístico

$$T = \frac{S}{n}$$

es simplemente la suma S escalada por $\frac{1}{n}$. Se conoce la siguiente propiedad de la distribución Gamma:

- Si $S \sim \Gamma(n, \lambda)$, entonces $\alpha S \sim \Gamma\left(n, \frac{\lambda}{\alpha}\right)$.

En nuestro caso, $\alpha = \frac{1}{n}$. Por tanto,

$$T = \frac{S}{n} \sim \Gamma(n, ne^{\theta}).$$

3) Media y varianza de T

Sea $T \sim \Gamma(k, \lambda)$ con $k = n$ y $\lambda = ne^\theta$. Recordemos que:

$$E[\Gamma(k, \lambda)] = \frac{k}{\lambda}, \quad \text{Var}[\Gamma(k, \lambda)] = \frac{k}{\lambda^2}.$$

Por tanto:

1) Media de T :

$$E[T] = \frac{n}{ne^\theta} = \frac{1}{e^\theta}.$$

2) Varianza de T :

$$\text{Var}(T) = \frac{n}{(ne^\theta)^2} = \frac{n}{n^2 e^{2\theta}} = \frac{1}{ne^{2\theta}}.$$

7) Sea X_1, X_2, \dots, X_n una m.a.s de una variable X con distribución $\mathcal{N}(\mu, \sigma^2)$. Sean \bar{X} y S^2 su media y cuasi-varianzas muestrales, respectivamente. Sea X_{n+1} una nueva observación de X independiente de X_1, X_2, \dots, X_n . Obtener la distribución en el muestreo del estadístico:

$$\frac{X_{n+1} - \bar{X}}{S} \sqrt{\frac{n}{n+1}}.$$

1) Distribución de $X_{n+1} - \bar{X}$

1) La media muestral $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ es independiente de X_{n+1} (porque X_{n+1} es una nueva observación independiente).

2) Cada X_i (incluyendo X_{n+1}) se distribuye como $\mathcal{N}(\mu, \sigma^2)$.

3) Se sabe que

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \quad X_{n+1} \sim \mathcal{N}(\mu, \sigma^2),$$

y son independientes. Por ende,

$$X_{n+1} - \bar{X} \sim \mathcal{N}\left(0, \sigma^2 + \frac{\sigma^2}{n}\right) = \mathcal{N}\left(0, \sigma^2 \cdot \frac{n+1}{n}\right).$$

Es decir,

$$\text{Var}(X_{n+1} - \bar{X}) = \sigma^2 \cdot \frac{n+1}{n}.$$

2) Relación con el cociente Normal-Chi-cuadrado

Sabemos además que la cuasi-varianza S^2 satisface

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2,$$

y es independiente de \bar{X} (y por tanto también independiente de X_{n+1}).

Para simplificar la notación, definamos

$$Y = X_{n+1} - \bar{X}.$$

Entonces $Y \sim \mathcal{N}\left(0, \sigma^2 \cdot \frac{n+1}{n}\right)$. Podemos escribir

$$T = \frac{Y}{S} \sqrt{\frac{n}{n+1}} = \frac{\frac{Y}{\sigma}}{\sqrt{\frac{n+1}{n}}} \bigg/ \underbrace{\frac{S}{\sigma}}_{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}.$$

- La variable $U = \frac{\frac{Y}{\sigma}}{\sqrt{\frac{n+1}{n}}} \sim \mathcal{N}(0, 1)$.
- $\frac{S^2}{\sigma^2}$ es $\frac{1}{n-1}$ veces una χ^2 con $n-1$ grados de libertad.
- U y $\frac{S^2}{\sigma^2}$ son independientes.

La razón

$$\frac{U}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}$$

sigue una distribución **t de Student** con $n-1$ grados de libertad.

Por lo tanto,

$$T = \frac{Y}{S} \sqrt{\frac{n}{n+1}} = \frac{U}{\frac{S}{\sigma}} = \frac{U}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \sim t_{n-1}.$$

8) Sean X_1, X_2, \dots, X_{n_1} e Y_1, Y_2, \dots, Y_{n_2} muestras aleatorias simples independientes de dos poblaciones $X \sim \mathcal{N}(\mu_1, \sigma^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, respectivamente. Obtener la distribución en el muestreo estadístico:

$$\frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}},$$

siendo α y β dos números reales fijos.

Tenemos dos muestras aleatorias simples e independientes:

- X_1, X_2, \dots, X_{n_1} de una población $X \sim \mathcal{N}(\mu_1, \sigma^2)$.
- Y_1, Y_2, \dots, Y_{n_2} de otra población $Y \sim \mathcal{N}(\mu_2, \sigma^2)$.

Se denotan:

- $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ la media muestral de la primera población.
- $\bar{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j$ la media muestral de la segunda población.
- S_1^2 la cuasi-varianza muestral de la primera muestra.
- S_2^2 la cuasi-varianza muestral de la segunda muestra.

Sabemos que ambas poblaciones comparten la misma varianza σ^2 .

El estadístico a estudiar es:

$$T = \frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}},$$

donde $\alpha, \beta \in \mathbb{R}$ son constantes dadas.

1) Distribución del numerador

Sea

$$Z = \alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2).$$

1) Propiedad de $\bar{X} - \mu_1$:

\bar{X} es $\mathcal{N}\left(\mu_1, \frac{\sigma^2}{n_1}\right)$. Por lo tanto,

$$\bar{X} - \mu_1 \sim \mathcal{N}\left(0, \frac{\sigma^2}{n_1}\right).$$

2) Propiedad de $\bar{Y} - \mu_2$:

\bar{Y} es $\mathcal{N}\left(\mu_2, \frac{\sigma^2}{n_2}\right)$. Entonces

$$\bar{Y} - \mu_2 \sim \mathcal{N}\left(0, \frac{\sigma^2}{n_2}\right).$$

3) Independencia de las dos muestras:

Como las dos muestras (de X y de Y) son independientes, también lo son $\bar{X} - \mu_1$ y $\bar{Y} - \mu_2$.

Por consiguiente, la variable

$$Z = \alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)$$

es normal de media 0 y varianza

$$\text{Var}(Z) = \alpha^2(\bar{X} - \mu_1) + \beta^2\text{Var}(\bar{Y} - \mu_2) = \alpha^2\frac{\sigma^2}{n_1} + \beta^2\frac{\sigma^2}{n_2}.$$

Por tanto,

$$Z \sim \mathcal{N}\left(0, \sigma^2\left(\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}\right)\right).$$

2) Distribución de la varianza combinada en el denominador

En el denominador aparece el factor

$$\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

Este es el **estimador combinado** de la desviación típica σ . Recordemos que si las dos muestras provienen con la **misma** varianza σ^2 , entonces

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi_{n_1 + n_2 - 2}^2.$$

Además, este estimador combinado de la varianza es independiente de las medias muestrales \bar{X} y \bar{Y} , y por ende, es independiente del numerador Z .

3) El estadístico T

Reescribimos el estadístico:

$$T = \frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}} = \frac{\frac{Z}{\sigma}}{\frac{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2(n_1 + n_2 - 2)}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}}{\sigma}}.$$

• La variable

$$\frac{Z}{\sigma} = \frac{Z}{\sqrt{\text{Var}(Z)}} \cdot \sqrt{\frac{\text{Var}(Z)}{\sigma^2}} \sim \mathcal{N}\left(0, \frac{\text{Var}(Z)}{\sigma^2}\right) \quad \text{con} \quad \text{Var}(Z) = \sigma^2\left(\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}\right).$$

Dividiendo Z por $\sigma\sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}$ se obtiene un $\mathcal{N}(0, 1)$.

• La parte

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \sim \chi_{n_1 + n_2 - 2}^2 \text{ e independiente de } Z.$$

Por la definición de la distribución t de Student con k grados de libertad,

$$t_k = \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{\chi_k^2}{k}}},$$

nuestro estadístico T encaja exactamente con esta forma, con $k = n_1 + n_2 - 2$.

4) Conclusión

El estadístico

$$T = \frac{\alpha(\bar{X} - \mu_1) + \beta(\bar{Y} - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{\alpha^2}{n_1} + \frac{\beta^2}{n_2}}}$$

sigue, en el muestreo, una distribución **t de Student** con $n_1 + n_2 - 2$ grados de libertad. Es decir,

$$T \sim t_{n_1+n_2-2}.$$

9) Una empresa de agua produce botellas que deberían contener 300ml pero que presentan en la práctica una variabilidad modelada por una distribución Normal con media $\mu = 298\text{ml}$ y desviación típica $\sigma = 3\text{ml}$.

a) ¿Cuál es la probabilidad de que una botella elegida al azar de la producción contenga menos de 295ml?

Tenemos botellas cuyo contenido (X) está modelado por una distribución normal con media $\mu = 298\text{ml}$ y desviación típica $\sigma = 3\text{ml}$. Es decir,

$$X \sim \mathcal{N}(298, 3^2).$$

Queremos calcular

$$P(X < 295).$$

Definimos la variable tipificada

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 298}{3}.$$

Entonces,

$$P(X < 295) = P\left(Z < \frac{295 - 298}{3}\right) = P(Z < -1) = P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587.$$

b) ¿Cuál es la probabilidad de que el contenido medio de un paquete de seis botellas sea inferior a 295ml?

Sea \bar{X}_6 la media muestral de 6 botellas independientes.

$$\bar{X}_6 = \frac{1}{6} \sum_{i=1}^6 X_i.$$

Dado que cada $X_i \sim \mathcal{N}(\mu, \sigma^2)$ y son independientes, la media \bar{X}_6 sigue una distribución normal con:

- Media $\mu_{\bar{X}_6} = \mu = 298$.
- Desviación típica $\sigma_{\bar{X}_6} = \frac{\sigma}{\sqrt{6}} = \frac{3}{\sqrt{6}}$.

Por lo tanto,

$$\bar{X}_6 \sim \mathcal{N}\left(298, \left(\frac{3}{\sqrt{6}}\right)^2\right).$$

Queremos

$$\begin{aligned} P(\bar{X}_6 < 295) &= P\left(Z < \frac{295 - 298}{\frac{3}{\sqrt{6}}}\right) = P\left(Z < \frac{-3}{\frac{3}{\sqrt{6}}}\right) = P(Z < -\sqrt{6}) \\ &= P(Z > \sqrt{6}) = 1 - P(Z < \sqrt{6}) \simeq 1 - P(Z < 2.45) = 1 - 0.9929 = 0.0071. \end{aligned}$$

10) Se realiza una medición de peso en un laboratorio, sabiendo que la desviación típica de las mediciones es $\sigma = 10\text{mg}$. La medición se repite 3 veces, se calcula la media \bar{x} , y este es el resultado proporcionado como estimación del peso.

a) ¿Cuál es la desviación típica del resultado calculado?

Si cada una de las mediciones X_i tiene varianza $\sigma^2 = 10^2 \text{mg}^2$, y las mediciones son independientes, entonces la varianza de la **media muestral** de n mediciones es

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

En consecuencia, la **desviación típica** de la media (o error estándar de la media) es

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

Para $n = 3$ mediciones,

$$\sqrt{\text{Var}(\bar{X})} = \frac{10}{\sqrt{3}} \approx 5.77 \text{mg}.$$

b) ¿Cuántas veces se debe repetir la medición para que la desviación típica del valor medio se reduzca a 5?

Queremos que la desviación típica de \bar{X} sea 5 mg.

Usando

$$\sqrt{\text{Var}(\bar{X})} = \frac{10}{\sqrt{n}} = 5,$$

resolvemos para n :

$$\frac{10}{\sqrt{n}} = 5 \longrightarrow \sqrt{n} = 2 \longrightarrow n = 4.$$

Por lo tanto, **necesitamos 4 mediciones** para reducir la desviación típica de la media a 5 mg.

11) El resultado de una encuesta fue que el 59% de la población española opina que el contexto económico es bueno o muy bueno. Supongamos que, extrapolando al conjunto de la población, efectivamente la proporción de todos los españoles que piensan que la situación es buena o muy buena es del 0.59.

Tenemos una población en la que la proporción real de personas que piensan que la situación económica es buena o muy buena es $p = 0.59$.

Cuando el tamaño de la muestra n es suficientemente grande, podemos aproximar la distribución de \hat{p} mediante una Normal con media p y desviación típica

$$\sqrt{\frac{p(1-p)}{n}}.$$

En este problema:

- $p = 0.59$
- $q = 1 - p = 0.41$
- Desviación típica de \hat{p} :

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.59 \cdot 0.41}{n}}.$$

Nos interesa

$$P(0.56 \leq \hat{p} \leq 0.62) = P(\hat{p} - 0.59 \in [-0.03, 0.03]).$$

Definamos la variable tipificada

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \sim \mathcal{N}(0, 1).$$

Entonces

$$P(|\hat{p} - 0.59| \leq 0.03) = P(-0.03 \leq \hat{p} - 0.59 \leq 0.03) = P(-z_0 \leq Z \leq z_0),$$

donde

$$z_0 = \frac{0.03}{\sqrt{\frac{0.59 \cdot 0.41}{n}}}$$

- a) Sabemos que las encuestas incluyen márgenes de error que son aproximadamente ± 3 puntos. ¿Cuál es la probabilidad de que una muestra aleatoria de 300 españoles presente una proporción muestral que caiga dentro del intervalo 0.59 ± 0.03 ?

- 1) **Cálculo de la desviación típica de \hat{p} :**

$$\sqrt{\frac{0.59 \cdot 0.41}{300}} = \sqrt{\frac{0.2419}{300}} \simeq 0.0284.$$

- 2) **Cálculo de z_0 :**

La amplitud del intervalo es 0.03. Dividimos por la desviación típica:

$$z_0 = 0. \frac{03}{0.0284} \simeq 1.06$$

- 3) **Probabilidad asociada:**

$Z \sim \mathcal{N}(0, 1)$. Entonces

$$P(-1.06 \leq Z \leq 1.06).$$

Usando tablas o función de distribución normal,

$$\Phi(1.06) \simeq 0.8554,$$

de modo que

$$P(-1.06 \leq Z \leq 1.06) = \Phi(1.06) - (1 - \Phi(1.06)) = 0.8554 - (1 - 0.8554) = 0.8554 - 0.1446 = 0.7108 \approx 0.71$$

Por lo tanto, **la probabilidad es aproximadamente 0.71** (un 71%).

- b) Contesta a la pregunta anterior en el caso en que la muestra consta de 600 personas y cuando la muestra consta de 1200 personas. ¿Cuál es el efecto de aumentar el tamaño de la muestra?

Caso $n = 600$:

- 1) **Desviación típica de \hat{p} :**

$$\sqrt{\frac{0.2419}{600}} \simeq 0.0201.$$

- 2) **Cálculo de z_0 :**

$$z_0 = \frac{0.03}{0.0201} \simeq 1.49$$

- 3) **Probabilidad:**

$$P(-1.49 \leq Z \leq 1.49).$$

Tenemos que $\Phi(1.49) \approx 0.9319$. Por lo tanto,

$$P(-1.49 \leq Z \leq 1.49) = 0.9319 - (1 - 0.9319) = 0.9319 - 0.0681 = 0.8638 \approx 0.86$$

Por lo tanto, **la probabilidad es aproximadamente 0.86** (un 86%).

Caso $n = 1200$:

- 1) **Desviación típica de \hat{p} :**

$$\sqrt{\frac{0.2419}{1200}} \approx 0.0142.$$

- 2) **Cálculo de z_0 :**

$$z_0 = \frac{0.03}{0.0142} \approx 2.11$$

3) Probabilidad:

$$P(-2.11 \leq Z \leq 2.11).$$

Tenemos que $\Phi(2.11) \approx 0.9826$. Entonces,

$$P(-2.11 \leq Z \leq 2.11) = 0.9826 - (1 - 0.9826) = 0.9826 - 0.0174 = 0.9652 \approx 0.965.$$

Por tanto, la probabilidad es aproximadamente **0.965** (un 96.5%).

- 12)** Un aparato de medición es exacto (el valor proporcionado medio es el valor auténtico de la señal) y la desviación típica del valor medido es 0.1 unidades. La distribución del valor medido es aproximadamente normal. ¿Cuál es la probabilidad de que el valor de una medición se aleje de la señal auténtica en más de 0.1 unidades? ¿Y si se repite la medición 5 veces y se toma la media de los 5 valores obtenidos?

Tenemos un aparato de medición que, al medir una señal (valor esperado μ), proporciona un valor X que es:

- **Exacto en promedio:** $E[X] = \mu$.
- **Normalmente distribuido** con desviación típica $\sigma = 0.1$. Es decir,

$$X \sim \mathcal{N}(\mu, 0.1^2).$$

Queremos calcular:

- 1) $P(|X - \mu| > 0.1)$, es decir, la probabilidad de que **una medición individual** se aleje más de 0.1 unidades de la señal real.
 - 2) $P(\bar{X}_5 - \mu > 0.1)$, donde \bar{X}_5 es la **media de 5 mediciones independientes**.
- 1) Probabilidad de que una sola medición difiera de μ es más de 0.1

Sea X la lectura, $X \sim \mathcal{N}(\mu, 0.1^2)$. Entonces

$$P(|X - \mu| > 0.1) = P(X - \mu > 0.1 \text{ o } X - \mu < -0.1).$$

Tipificamos usando $Z = \frac{X - \mu}{0.1}$, que sigue $\mathcal{N}(0, 1)$. Así,

$$P(|X - \mu| > 0.1) = P(|Z| > 1) = 2(1 - \Phi(1)).$$

Tenemos que $1 - \Phi(1) \approx 0.1587$. Por tanto,

$$P(|Z| > 1) = 2 \cdot 0.1587 = 0.3174.$$

- 2) Probabilidad de que la media de 5 mediciones difiera que μ en más de 0.1

Sea $\bar{X}_5 = \frac{1}{5} \sum_{i=1}^5 X_i$. Dado que cada $X_i \sim \mathcal{N}(\mu, 0.1^2)$ y son independientes, se tiene

- $E[\bar{X}_5] = \mu$
- $\text{Var}(\bar{X}_5) = \frac{\sigma^2}{5} = \frac{0.1^2}{5}$
- $\sigma(\bar{X}_5) = \frac{0.1}{\sqrt{5}}$

La probabilidad de interés es

$$P(|\bar{X}_5 - \mu| > 0.1).$$

Tipificamos con

$$Z = \frac{\bar{X}_5 - \mu}{\frac{0.1}{\sqrt{5}}} \sim \mathcal{N}(0, 1).$$

Entonces,

$$P(|\bar{X}_5 - \mu| > 0.1) = P\left(|Z| > \frac{0.1}{\frac{0.1}{\sqrt{5}}}\right) = P(|Z| > \sqrt{5}) \approx 1 - \Phi(Z < 2.24) = 0.0127.$$

De modo que

$$P(|Z| > 2.24) = 2 \cdot (1 - \Phi(2.24)) \approx 2 \cdot 0.0127 = 0.0254.$$

13) En condiciones normales, una máquina produce piezas con una tasa de defectuosas del 1%. Para controlar que la máquina sigue bien ajustada, se escogen al azar cada día 100 piezas en la producción y se somete a un test. ¿Cuál es la probabilidad de que, si la máquina está bien ajustada, haya, en una de esas muestras, más del 2% de piezas defectuosas? Si un día, 3 piezas resultan defectuosas, ¿cuáles son las conclusiones que sacaríamos sobre el funcionamiento de la máquina?

Bajo el supuesto de que la máquina está bien ajustada, X (número de piezas defectuosas en la muestra) sigue una distribución **Binomial** con parámetros $n = 100$ y $p = 0.01$.

$$X \sim \mathcal{B}(n = 100, p = 0.01).$$

1) Probabilidad de obtener más del 2% de piezas defectuosas si la máquina está bien

El 2% de 100 piezas son 2 piezas. Por tanto, "más del 2%" significa que en la muestras salen **3 o más defectuosas**. Queremos

$$P(X > 2) = 1 - P(X \leq 2).$$

1.1) Aproximación con distribución Poisson

Dado que $p = 0.01$ es bastante pequeño y $n = 100$ moderado, podemos usar la **aproximación Poisson** con $\lambda = np = 1$.

Entonces $X \approx \text{Poisson}(\lambda = 1)$. Se calcula

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2),$$

donde, para $\text{Poisson}(\lambda = 1)$:

$$P(X = k) = \frac{e^{-1} 1^k}{k!}.$$

- $P(X = 0) = e^{-1} \approx 0.3679$
- $P(X = 1) = e^{-1} \cdot \frac{1}{1!} = 0.3679$
- $P(X = 2) = e^{-1} \cdot \frac{1^2}{2!} = 0.1839$

Sumamos:

$$P(X \leq 2) = 0.3679 + 0.3679 + 0.1839 = 0.9197$$

Por tanto,

$$P(X > 2) = 1 - 0.9197 = 0.0803 \approx 0.08.$$

Esto indica que, si la máquina está bien, hay alrededor de un 8% de probabilidad de observar 3 o más piezas defectuosas en una muestra de 100

2) Interpretación en un día con 3 defectuosas observadas

Si en la muestra diaria (de 100 piezas) aparecen **3 defectuosas**, podemos preguntarnos si esto sugiere que la máquina se ha desajustado.

- Hemos visto que, si realmente la máquina sigue produciendo al 1% de defectos, la probabilidad de ver 3 o más defectuosas es alrededor de un 8%.
- Un 8% no es un suceso extremadamente raro. En un test de hipótesis con un nivel de significación típico del 5%, podríamos decir que la **p-value** ≈ 0.08 es **mayor** que 0.05, de modo que no **no** rechazaríamos la hipótesis de que la máquina está bien.

En otras palabras, **3 piezas defectuosas no es una evidencia suficientemente fuerte** para concluir que la máquina está mal ajustada.

Tema 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 una 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 parámetros. **Estimación paramétrica**.
- No limitados la familia de distribuciones a la que pertenece nuestro modelo.
→ Buscamos información sobre la distribución en sí (función de distribución, de densidad o función puntual 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

Sondeo sobre intención de participación en una 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.

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

Ingredientes del modelo

- Experimento aleatoria
- 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

Cualquier estadístico diseñado para aproximar el valor de un parámetro θ del modelo, se llama **estimador puntual** del parámetro θ .

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 una 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, \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_θ , $\mu_k = E[X^k]$. Resolvemos:

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

Ejemplos

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

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

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

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

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

$$\theta = p$$

$$\mu(\theta) = E[X] = \bar{x} = p$$

$$\hat{p} = \bar{x} \text{ (proporción muestral)}$$

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

- 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 simpl (X_1, \dots, X_n) , que denotamos por (x_1, \dots, x_n) , consideramos la función de θ :

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

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

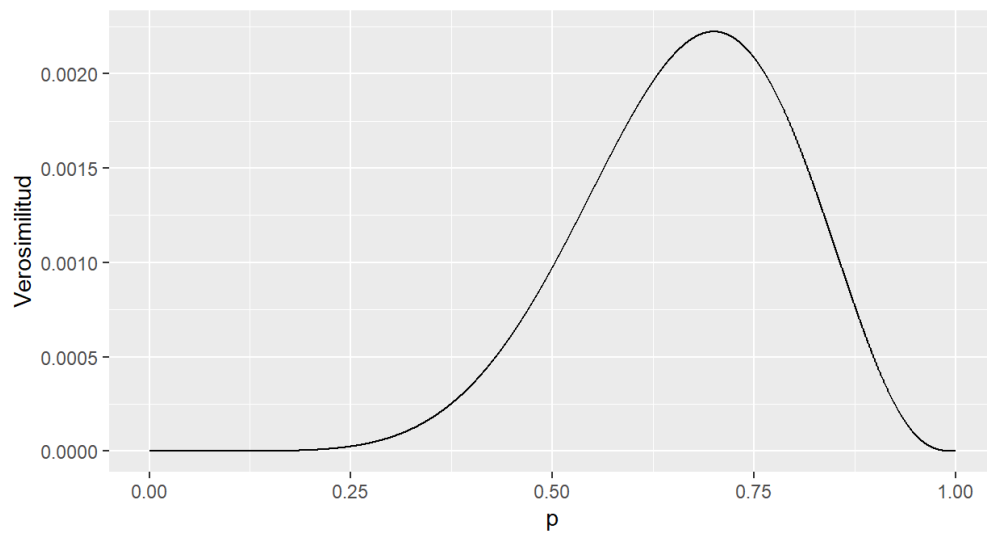
Ejemplo de cálculo de la verosimilitud

- 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

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

$$\hat{\theta}_{\arg\max_{\theta \in \Theta} L_n(\theta)}.$$

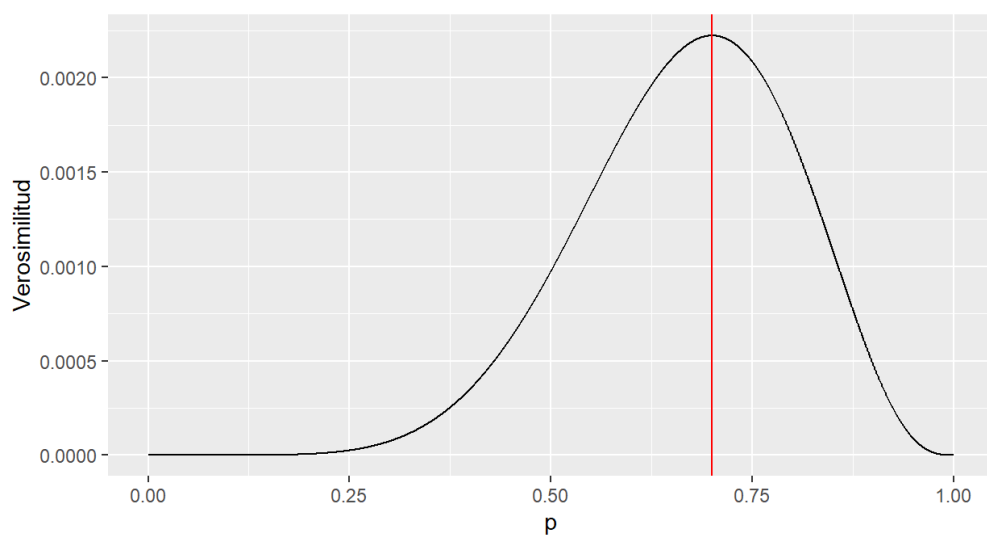
Nota

- 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:

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



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 .

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

(X_1, \dots, X_n) m.a.s de tamaño n

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

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

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

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

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

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

La función de 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:

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

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

2.8) Métodos para evaluar un estimador

Recordad

- Un estimador es una variable aleatoria.
- Es valioso disponer de conocimiento sobre la distribución del estimador (su **distribución en el muestreo**) \longrightarrow permite manejar el riesgo y el error que podemos cometer al aproximar 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 $n \rightarrow \infty$.

2.9) Sesgo

Definición

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 **inesgado**.

2.10) Error cuadrático medio

Para medir la variabilidad en el muestreo de un estimador.

Definición

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

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

desconocidos

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

$$E[\bar{X}] = \mu \longrightarrow \hat{\mu} = \bar{X} \text{ es un estimador inesgado para } \mu$$

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

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

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

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

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

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

Nota

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:

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

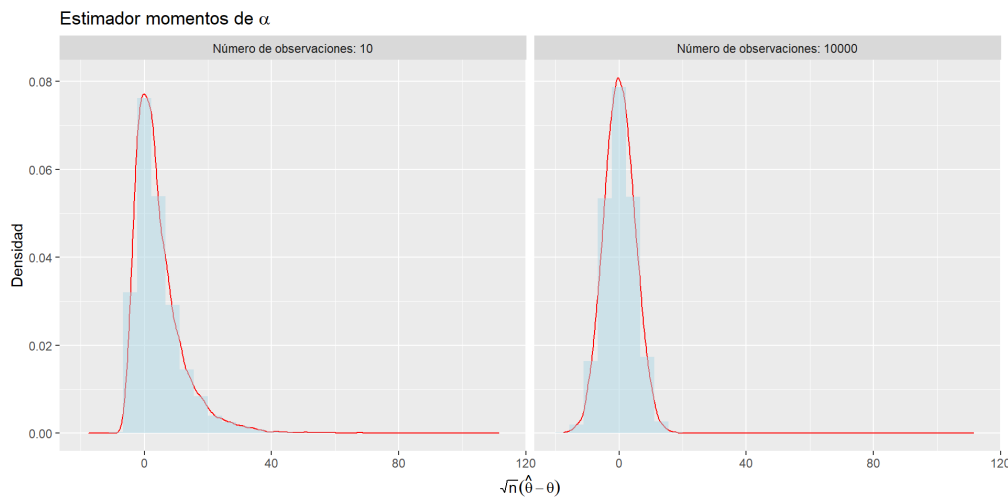
(la distribución de $\sqrt{n}(\hat{\theta}_n - \theta)$ se aproxima a la distribución $\mathcal{N}(0, \Lambda)$).

Ejemplo: distribución gamma(α, β)

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

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

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



2.15) Estimación no paramétrica

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 paramétricos** 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**.

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**.

Estimación de la función puntual de probabilidad

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

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

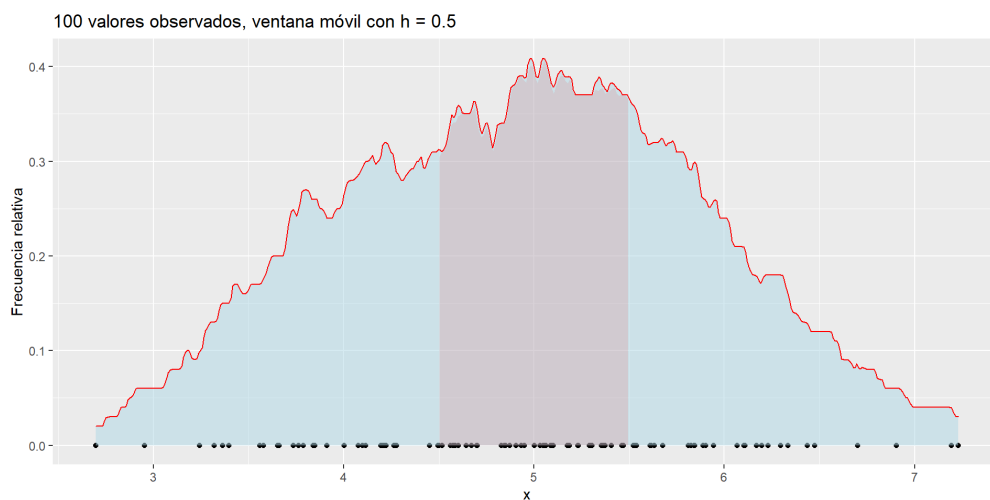
Primera opción: nos basamos en frecuencias relativas

- 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.



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

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

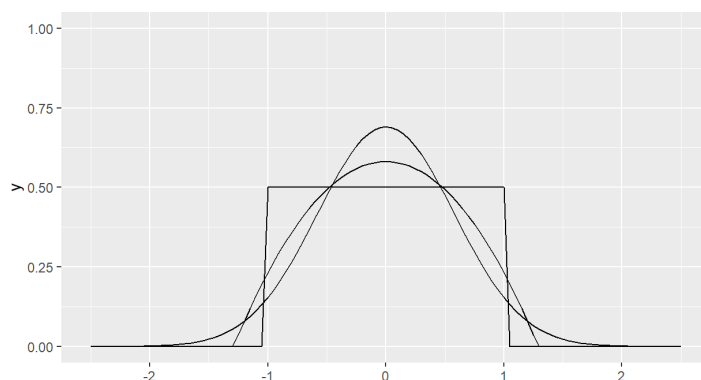
- En la expresión anterior hay dos elementos principales, la función K , que en el ejemplo es la función de densidad de una distribución uniforme, y el parámetro h , que hemos fijado en 0.5 en el ejemplo, y en función de su valor tiene en cuenta valores de la muestra cercanos o alejados al valor x .
- En el primer caso hablamos sobre la [función núcleo \(kernel\)](#) y en el segundo caso del [ancho de banda \(bandwidth\)](#).
- 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

Función núcleo

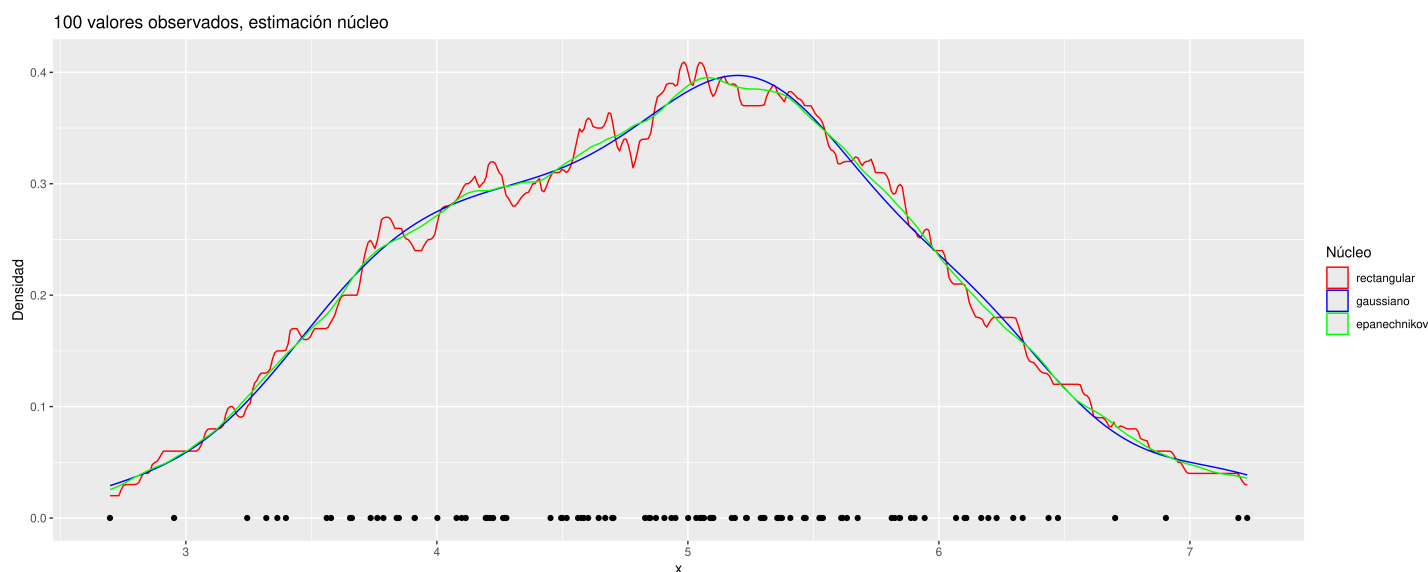
- En la literatura hay distintas propuestas de la función K , **función núcleo**.
- Cualquier estimador de la densidad en la forma anterior, con una función núcleo K , recibe el nombre de **estimador tipo núcleo**.
- Los casos más usados de función núcleo son los siguientes:
 - **Núcleo normal**: K es 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 a valor x en el intervalo $x \pm h$.



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

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



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 asociada 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 tienda 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 $nh_n \rightarrow \infty$.

Resultado

Para todo x ,

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

- Este resultado junto con las suposiciones que hemos hecho anteriormente nos lleva a concluir a 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

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 $R(f'')$

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 muestral S :

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

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)

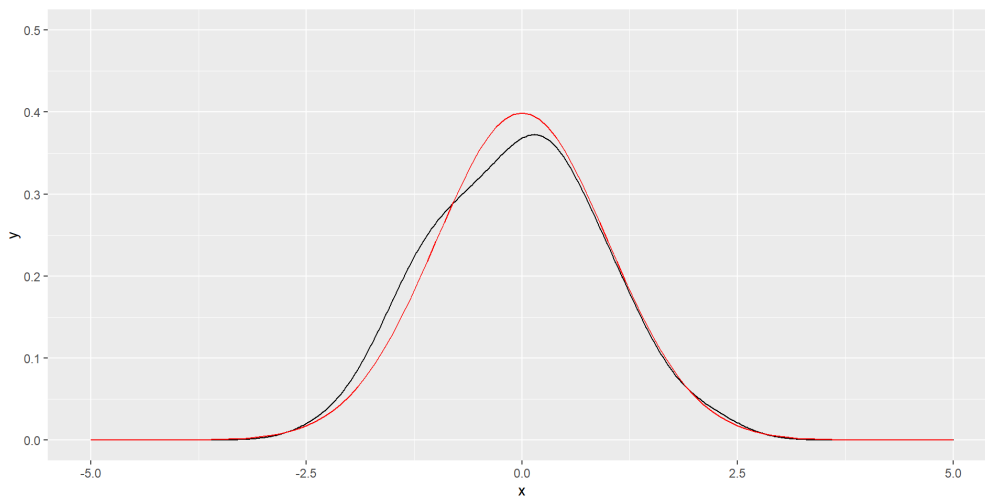
La forma más complicada es dar una estimación no paramétrica de $R(f'')$. La metodología va más allá de los contenidos de esta asignatura y no veremos los detalles, solo indicar que podemos seleccionarla con [R](#) y lo veremos en prácticas.

Implementación en R

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

```
## [1] 0.4047312
```

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



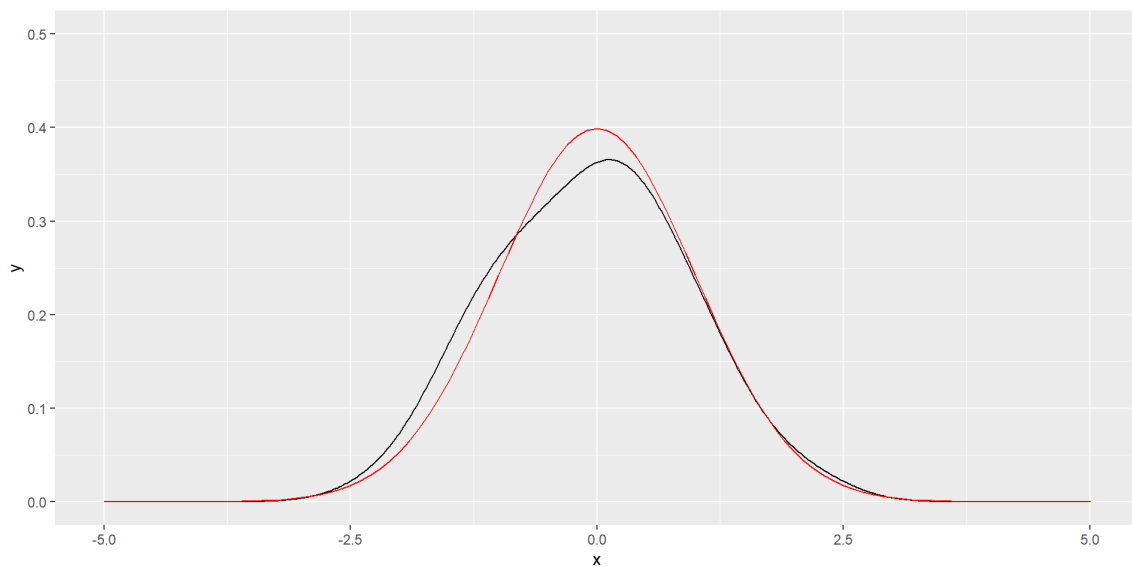
2.17.1.2) Aproximación no paramétrica para el cálculo de $R(f'')$, implementación en R

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

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

```
## [1] 0.442874
```

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



2.18) Introducción al Bootstrap

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, submuestras 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 del 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 bootstraps"*, que hace referencia a situaciones en la 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 .

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 aproxima (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.

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*:

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

Tened en cuenta

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

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} asocia a cualquier evento A , la probabilidad empírica: $\text{Prob}(A) = \frac{\#\{x_i \in A\}}{n}$.
- Podemos calcular la esperanza de cualquier función g :

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

- Por ejemplo:

$$\text{La esperanza de } \hat{F} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$\text{La varianza de } \hat{F} = \frac{(n-1)}{n} s^2.$$

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.

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.

2.18.1) El error estándar Bootstrap

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 $\frac{S}{\sqrt{n}}$.

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 son 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{se} , como la desviación típica muestral de $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$, esto es, como

$$\hat{se} = \sqrt{\sum_{b=1}^B \frac{\left(\hat{\theta}^{*b} - \sum_{i=1}^B \frac{\hat{\theta}^{*i}}{B} \right)^2}{B-1}}$$

Hoja de ejercicios Tema 2: Estimación

1) Sea X variable aleatoria con distribución Bernoulli, de parámetro $p(X \sim b(p))$ y sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s) de X

a) Obtener el estimador de los momentos y el estimador de máxima verosimilitud de p .

1) Estimador de momentos

Sea $X \sim \text{Bernoulli}(p)$, entonces

$$E[X] = p$$

Si X_1, X_2, \dots, X_n es una muestra aleatoria simple, el estimador de momentos \hat{p} se obtiene igualando $E[X]$ a la media muestral:

$$\hat{p} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

2) Estimador de máxima verosimilitud

La función de verosimilitud para una muestra de tamaño n es:

$$L(p) = \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i}$$

Tomando logaritmos:

$$\ell(p) = \ln(L(p)) = \sum_{i=1}^n [X_i \ln(p) + (1-X_i) \ln(1-p)]$$

Derivamos respecto a p e igualamos a cero:

$$\frac{\partial \ell(p)}{\partial p} = \frac{\sum_{i=1}^n X_i}{p} - \frac{n - \sum_{i=1}^n X_i}{1-p} = 0$$

Simplificamos:

$$\begin{aligned} \sum_{i=1}^n x_i(1-p) &= \sum_{i=1}^n p \\ \sum_{i=1}^n x_i - p \sum_{i=1}^n x_i &= p \sum_{i=1}^n (1-x_i) \\ \sum_{i=1}^n x_i &= p \sum_{i=1}^n x_i + p \sum_{i=1}^n (1-x_i) \\ \sum_{i=1}^n x_i &= p \sum_{i=1}^n 1 \\ p &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}. \end{aligned}$$

b) Calcular el sesgo y el error cuadrático medio de ambos estimadores.

1) Sesgo

El sesgo de un estimador \hat{p} es:

$$\text{Sesgo}(\hat{p}) = E[\hat{p}] - p = p - p = 0$$

Dado que $\bar{X} \sim \mathcal{B}(n, p)/n$:

$$E[\bar{X}] = E\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \cdot n \cdot p = p$$

Por lo tanto, tanto el estimador de momentos como el estimador de máxima verosimilitud son insesgados.

2) Error Cuadrático Medio (ECM)

El ECM de un estimador \hat{p} es:

$$\text{ECM}(\hat{p}) = E[(\hat{p} - p)^2]$$

Usando la relación $\text{ECM}(\hat{p}) = \text{Var}(\hat{p}) + (\text{Sesgo}(\hat{p}))^2$, y como los estimadores son insesgados, el ECM coincide con la varianza:

$$\text{ECM}(\hat{p}) = \text{Var}(\hat{p})$$

La varianza de \bar{X} es:

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i)$$

Dado que $X_i \sim \text{Bernoulli}(p)$, la varianza de X_i es $\text{Var}(X_i) = p(1-p)$. Por lo tanto:

$$\text{Var}(\bar{X}) = \frac{1}{n^2} (n \cdot p(1-p)) = \frac{p(1-p)}{n}$$

Por lo tanto, el ECM de ambos estimadores es:

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

2) Sea X variable aleatoria con distribución exponencial, de parámetro $\lambda (X \sim P(\lambda))$ y sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s) de X . Obtener el estimador de los momentos y el estimador de máxima verosimilitud de λ .

Dado que $X \sim \text{Exp}(\lambda)$, su función de densidad de probabilidad es:

$$f_X(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Donde $\lambda > 0$ es el parámetro de la distribución.

1) Estimador de los momentos

Para el método de los momentos, igualamos el momento teórico con el momento muestral. El primero momento de X , es decir, la media esperada, es:

$$E[X] = \frac{1}{\lambda}.$$

La media muestra \bar{X} es:

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

Igualemos la media teórica con la media muestral para estimar λ :

$$E[X] = \bar{X} \longrightarrow \frac{1}{\lambda} = \bar{X}$$

Por lo tanto, el **estimador de los momentos** para λ es:

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

2) Estimador de máxima verosimilitud

La función de verosimilitud para la muestra X_1, X_2, \dots, X_n es el producto de las densidades:

$$L(\lambda) = \prod_{i=1}^n f_X(X_i; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda X_i}.$$

Romando logaritmos naturales para simplificar:

$$\ell(\lambda) = \ln L(\lambda) = \sum_{i=1}^n \ln(\lambda) - \lambda \sum_{i=1}^n X_i.$$

Derivamos respecto a λ e igualamos a cero para encontrar el máximo:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i = 0 \longrightarrow \lambda = \frac{n}{\sum_{i=1}^n X_i}.$$

Dado que $\sum_{i=1}^n X_i = n\bar{X}$, esto puede reescribirse como:

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

3) Demostrar que, si $\theta \mapsto L_n(\theta)$ es una verosimilitud, maximizar $\theta \mapsto L_n(\theta)$ es equivalente a maximizar $\theta \mapsto \log(L_n(\theta))$.

Para demostrar que maximizar $\theta \mapsto L_n(\theta)$ es equivalente a maximizar $\theta \mapsto \log(L_n(\theta))$, analizaremos la relación entre ambas funciones.

1) Propiedad de la función logaritmo

La función logaritmo natural, $\log(x)$, es estrictamente creciente para $x > 0$. Esto significa que si $x_1 > x_2 > 0$, entonces $\log(x_1) > \log(x_2)$. Por lo tanto, maximizar una función positiva $L_n(\theta) > 0$ es equivalente a maximizar su logaritmo natural $\log(L_n(\theta))$.

2) Relación entre $L_n(\theta)$ y $\log(L_n(\theta))$

Supongamos que θ^* maximiza $L_n(\theta)$. Esto implica que para todo θ :

$$L_n(\theta^*) \geq L_n(\theta).$$

Dado que $\log(x)$ es estrictamente creciente, tomar logaritmos no altera la relación:

$$\log(L_n(\theta^*)) \geq \log(L_n(\theta)).$$

Esto muestra que θ^* también maximiza $\log(L_n(\theta))$.

4) Sea X variable aleatoria con distribución de Poisson, de parámetro λ ($X \sim P(\lambda)$) y sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s) de X . Obtener el estimador de los momentos y el estimador de máxima verosimilitud de λ .

Para estimar el parámetro λ de una distribución de Poisson, analizamos los dos métodos de estimación: **método de los momentos** y **máxima verosimilitud**.

1) Función de probabilidad de Poisson

La función de probabilidad de una variable aleatoria $X \sim \text{Poisson}(\lambda)$ es:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

2) Estimador de los momentos

El momento teórico de primer orden de X es su esperanza:

$$E[X] = \lambda.$$

Por el método de los momentos, igualamos la media teórica con la media muestral:

$$E[X] = \bar{X} \longrightarrow \lambda = \bar{X}.$$

Por lo tanto, el **estimador de los momentos** para λ es:

$$\hat{\lambda} = \bar{X}.$$

3) Estimador de máxima verosimilitud

La función de verosimilitud para una muestra aleatoria simple X_1, X_2, \dots, X_n es el producto de las probabilidades individuales:

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

Tomamos el logaritmo natural para facilitar el cálculo:

$$\ell(\lambda) = \ln L(\lambda) = \left(\sum_{i=1}^n X_i \right) \ln(\lambda) - n\lambda - \sum_{i=1}^n \ln(X_i!).$$

Derivamos respecto a λ e igualamos a cero para encontrar el máximo:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{\sum_{i=1}^n X_i}{\lambda} - n = 0 \longrightarrow \lambda = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Por lo tanto, el **estimador de máxima verosimilitud** para λ es:

$$\hat{\lambda} = \bar{X}.$$

5) Sea X variable aleatoria con distribución normal, de parámetros μ y σ^2 ($X \sim \mathcal{N}(\mu, \sigma^2)$) y sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s.) de X . Obtener el estimador de máxima verosimilitud de (μ, σ^2) . ¿Cuál es el sesgo y el error cuadrático medio para el estimador de máxima verosimilitud de μ ?

1) Planteamiento del problema

Dado que $X \sim \mathcal{N}(\mu, \sigma^2)$, la función de densidad de probabilidad es:

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Para una muestra aleatoria simple X_1, X_2, \dots, X_n necesitamos encontrar los estimadores de máxima verosimilitud para μ y σ^2 .

2) Función de verosimilitud

La función de verosimilitud es:

$$L(\mu, \sigma^2) = \prod_{i=1}^n f_X(X_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i-\mu)^2}.$$

3) Logaritmo de la verosimilitud

Tomamos el logaritmo natural para facilitar los cálculos:

$$\ell(\mu, \sigma^2) = \ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2.$$

4) Derivadas y estimadores

a) Para μ :

Derivamos respecto a μ e igualamos a cero:

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu) = 0 \longrightarrow \sum_{i=1}^n (X_i - \mu) = 0 \longrightarrow \sum_{i=1}^n X_i = n\mu \longrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

Por tanto, el **estimador de máxima verosimilitud de μ** es la **media muestral**, \bar{X} .

b) Para σ^2 :

Derivamos respecto a σ^2 e igualamos a cero:

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

Este estimador de máxima verosimilitud de σ^2 es **no** insesgado.

5) Sesgo y ECM del estimador de μ

a) Sesgo

El sesgo de un estimador $\hat{\mu}$ es:

$$\text{Sesgo}(\hat{\mu}) = E[\hat{\mu}] - \mu.$$

Como $\hat{\mu} = \bar{X}$ y sabemos que $E[\bar{X}] = \mu$, se tiene:

$$\text{Sesgo}(\hat{\mu}) = \mu - \mu = 0$$

El estimador de μ es **insesgado**.

b) Error cuadrático medio (ECM)

El ECM de $\hat{\mu}$ es:

$$\text{ECM}(\hat{\mu}) = \text{Var}(\hat{\mu}) + (\text{Sesgo}(\hat{\mu}))^2 = \text{Var}(\hat{\mu}) = \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Por lo tanto:

$$\text{ECM}(\hat{\mu}) = \frac{\sigma^2}{n}.$$

6) Sea X una variable aleatoria con función de densidad

$$f(x, \theta) = \frac{2x}{\theta} \exp\left(-\frac{x^2}{\theta}\right) \chi_{(0, +\infty)}(x).$$

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s) de X . Obtener el estimador de máxima verosimilitud de θ .

1) Función de verosimilitud

La función de verosimilitud para la muestra X_1, X_2, \dots, X_n es el producto de las densidades individuales:

$$L(\theta) = \prod_{i=1}^n f(X_i, \theta) = \prod_{i=1}^n \frac{2X_i}{\theta} \exp\left(-\frac{X_i^2}{\theta}\right) = \left(\frac{2}{\theta}\right)^n \prod_{i=1}^n X_i \cdot \exp\left(-\frac{1}{\theta} \sum_{i=1}^n X_i^2\right).$$

2) Logaritmo de la función de verosimilitud

$$\ell(\theta) = \ln L(\theta) = n \ln 2 - n \ln \theta + \sum_{i=1}^n \ln X_i - \frac{1}{\theta} \sum_{i=1}^n X_i^2.$$

3) Derivada y condición de máximo

Para maximizar $\ell(\theta)$, derivamos respecto a θ e igualamos a cero:

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n X_i^2 = 0 \longrightarrow -n\theta + \sum_{i=1}^n X_i^2 = 0 \longrightarrow \theta = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

4) Estimador de máxima verosimilitud

El estimador de máxima verosimilitud para θ es:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2,$$

es decir, el promedio de los cuadrado de las observaciones.

7) Sea X una variable aleatoria con función de densidad

$$f(x, \theta) = \frac{\theta}{(1+x)^{1+\theta}} \chi_{(0,+\infty)}(x).$$

Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s) de X . Obtener el estimador de máxima verosimilitud de θ .

1) Función de verosimilitud

La función de verosimilitud para una muestra X_1, X_2, \dots, X_n es:

$$L(\theta) = \prod_{i=1}^n f(X_i, \theta) = \prod_{i=1}^n \frac{\theta}{(1+X_i)^{1+\theta}} = \theta^n \prod_{i=1}^n (1+X_i)^{-(1+\theta)} = \theta^n \prod_{i=1}^n (1+X_i)^{-1} \prod_{i=1}^n (1+X_i)^{-\theta}.$$

2) Logaritmo de la función de verosimilitud

Tomamos el logaritmo natural para facilitar el cálculo:

$$\ell(\theta) = \ln L(\theta) = n \ln \theta - (1+\theta) \sum_{i=1}^n \ln(1+X_i)$$

3) Derivada e igualación a cero

Para encontrar el valor de θ que maximiza la función de verosimilitud, derivamos $\ell(\theta)$ respecto a θ e igualamos a cero:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n \ln(1+X_i) = 0 \longrightarrow \frac{n}{\theta} = \sum_{i=1}^n \ln(1+X_i) \longrightarrow \theta = \frac{n}{\sum_{i=1}^n \ln(1+X_i)}.$$

4) Estimador de máxima verosimilitud

El **estimador de máxima verosimilitud** para θ es:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^n \ln(1+X_i)}.$$

8) Sea X una variable aleatoria con distribución uniforme en el intervalo $(a, 5)$ con $a < 5$ ($X \sim U(a, 5)$). Sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s.) de X . Obtener el estimador de máxima verosimilitud de a . Estudiar su distribución en el muestreo.

1) Planteamiento

Dada la variable aleatoria $X \sim U(a, 5)$ con $a < 5$, la función de densidad de probabilidad es:

$$f(x; a) = \begin{cases} \frac{1}{5-a} & \text{si } a < x < 5 \\ 0 & \text{en otro caso} \end{cases}$$

2) Función de verosimilitud

La función de verosimilitud para la muestra es el producto de las densidades:

$$L(a) = \prod_{i=1}^n f(X_i; a).$$

Sustituyendo $f(x; a)$:

$$L(a) = \begin{cases} \left(\frac{1}{5-a}\right)^n & \text{si } a < \min(X_1, X_2, \dots, X_n) \\ 0 & \text{en otro caso} \end{cases}$$

Para que $f(x; a) > 0$, se requiere que $a < \min(X_i)$. De lo contrario, $L(a) = 0$.

Por lo tanto, la función de verosimilitud se expresa como:

$$L(a) = \left(\frac{1}{5-a}\right)^n, \quad \text{para } a < \min(X_i).$$

3) Estimador de máxima verosimilitud

El estimador de máxima verosimilitud es el valor de a que maximiza $L(a)$. Observamos que $L(a)$ decrece conforme a aumenta. Por lo tanto, para maximizar $L(a)$, a debe ser lo más grande posible mientras satisface $a < \min(X_i)$.

Así, el estimador de máxima verosimilitud es:

$$\hat{a} = \min(X_1, X_2, \dots, X_n).$$

4) Distribución del estimador en el muestreo

Sabemos que para una muestra aleatoria simple X_1, X_2, \dots, X_n de una distribución uniforme $U(a, 5)$, la función de distribución acumulada de X_i es:

$$F_{X_i}(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{5-a} & \text{si } a \leq x \leq 5 \\ 1 & \text{si } x > 5 \end{cases}$$

El estimador $\hat{a} = \min(X_1, X_2, \dots, X_n)$ es el mínimo de n observaciones independientes, cuya función de distribución acumulada es:

$$F_{\hat{a}}(x) = P(\hat{a} \leq x) = 1 - P(\hat{a} > x) = 1 - P(X_1 > x, X_2 > x, \dots, X_n > x).$$

Dado que las observaciones son independientes:

$$P(\hat{a} > x) = P(X_1 > x)^n = (1 - F_{X_i}(x))^n.$$

Sustituyendo $F_{X_i}(x)$ para $a \leq x \leq 5$:

$$1 - F_{X_i}(x) = 1 - \frac{x-a}{5-a} = \frac{5-x}{5-a}.$$

Por lo tanto:

$$F_{\hat{a}}(x) = \begin{cases} 0 & \text{si } x < a \\ 1 - \left(\frac{5-x}{5-a}\right)^n & \text{si } a \leq x \leq 5 \\ 1 & \text{si } x > 5 \end{cases}$$

La densidad \hat{a} se obtiene derivando la función de distribución acumulada:

$$f_{\hat{a}}(x) = \begin{cases} n \frac{(5-x)^{n-1}}{(5-a)^{n-1}} & \text{si } a \leq x \leq 5 \\ 0 & \text{en otro caso} \end{cases}$$

9) Sea X variable aleatoria con distribución $\text{Gamma}(\alpha, \beta)$, de parámetros α y β ($X \sim \text{Gamma}(\alpha, \beta)$) y sea X_1, X_2, \dots, X_n una muestra aleatoria simple (m.a.s.) de X .

a) Obtener los estimadores de α y β usando el método de los momentos.

La distribución $X \sim \text{Gamma}(\alpha, \beta)$ tiene los siguientes momentos teóricos:

1) La media es:

$$E[X] = \frac{\alpha}{\beta}.$$

2) La varianza es:

$$\text{Var}(X) = \frac{\alpha}{\beta^2}$$

Para el método de los momentos, igualamos estos momentos teóricos a los momentos muestrales:

• **Media muestral:**

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

• **Varianza muestral:**

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

Igualemos los momentos:

1) De la media:

$$E[X] = \frac{\alpha}{\beta} \longrightarrow \frac{\alpha}{\beta} = \bar{X}$$

2) De la varianza:

$$\text{Var}(X) = \frac{\alpha}{\beta^2} \longrightarrow \frac{\alpha}{\beta^2} = S^2.$$

De la primera ecuación:

$$\alpha = \beta \bar{X}.$$

Sustituimos en la segunda ecuación de la varianza:

$$\frac{\beta \bar{X}}{\beta^2} = S^2 \longrightarrow \frac{\bar{X}}{\beta} = S^2 \longrightarrow \beta = \frac{\bar{X}}{S^2}.$$

Finalmente, sustituimos este valor de β en $\alpha = \beta \bar{X}$:

$$\alpha = \frac{\bar{X}}{S^2} \cdot \bar{X} = \frac{\bar{X}^2}{S^2}.$$

Estimadores de los momentos:

$$\hat{\alpha} = \frac{\overline{X^2}}{S^2}, \quad \hat{\beta} = \frac{\overline{X}}{S^2}.$$

b) Obtener la log-verosimilitud como función de (α, β)

La función de densidad de la distribución gamma es:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

donde $\Gamma(\alpha)$ es la función gamma. Para una muestra aleatoria simple X_1, X_2, \dots, X_n , la función de verosimilitud es:

$$L(\alpha, \beta) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} X_i^{\alpha-1} e^{-\beta X_i} = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} \prod_{i=1}^n X_i^{\alpha-1} e^{-\beta \sum_{i=1}^n X_i}.$$

La log-verosimilitud es el logaritmo de esta función:

$$\ell(\alpha, \beta) = \ln L(\alpha, \beta) = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \ln X_i - \beta \sum_{i=1}^n X_i.$$

c) Demostrar que, para un valor α fijado, el valor de β que maximiza la log-verosimilitud es:

$$\hat{\beta} = \frac{\overline{x}}{\alpha}.$$

Paso 1: Derivada respecto a β

Derivamos $\ell(\alpha, \beta)$ respecto a β :

$$\frac{\partial \ell(\alpha, \beta)}{\partial \beta} = -n\alpha \frac{1}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n X_i = -n\alpha\beta + \sum_{i=1}^n X_i = 0.$$

Paso 2: Igualamos la derivada a cero

Para maximizar $\ell(\alpha, \beta)$, igualamos la derivada a cero:

$$n\alpha\beta = \sum_{i=1}^n X_i \longrightarrow \beta = \frac{\sum_{i=1}^n X_i}{n\alpha}.$$

Paso 3: Relación con la media muestral

Recordemos que la media muestral es:

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

Sustituyendo $\sum_{i=1}^n X_i = n\overline{X}$ en la ecuación para β :

$$\beta = \frac{n\overline{X}}{n\alpha} = \frac{\overline{X}}{\alpha}$$

Por lo tanto, el valor de β que maximiza la log-verosimilitud para un α fijo es:

$$\hat{\beta} = \frac{\overline{X}}{\alpha}$$

Tema 3: Estimación por intervalo

3.1) Introducción

Estimación paramétrica

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

Bootstrap

También construiremos intervalos de confianza basados en percentiles Bootstrap.

Idea básica

Supongamos:

- $X \sim \mathcal{N}(\mu, 4)$.
- Consideramos una muestra aleatoria simple X_1, \dots, X_4 , estimamos 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 dónde 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.

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

- $\bar{X} \pm 2$ 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

- (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 $D(X_1, \dots, X_n)$ (extremos 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

- 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.

Procedimiento

- Nos fijamos el "nivel de riesgo", $0 < \alpha < 1$, por ejemplo, 0.1, 0.05, o 0.01.
- Buscamos $T(X_1, \dots, X_n)$ que se **pivotal**, 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 \rightarrow el nivel de cobertura $1 - \alpha$ de $[I(X_1, \dots, X_n), D(X_1, \dots, X_n)]$ 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

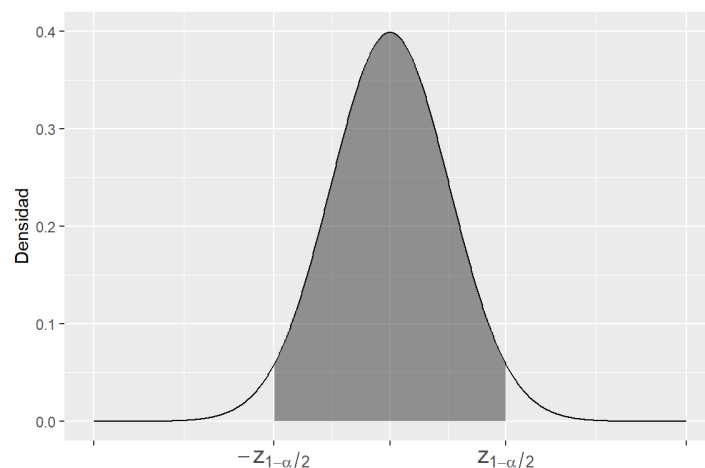
Consideramos una muestra aleatoria simple de una distribución $X \sim \mathcal{N}(\mu, \sigma^2)$. El valor de σ^2 es conocido

- 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.



Cuantil

Para $0 \leq u \leq 1$, z_u es el **cuantil** u de una $\mathcal{N}(0, 1)$, es decir, el valor que cumple $\mathbb{P}(Z \leq z_u) = u$.

$$\mathbb{P}\left[-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right] = 1 - \alpha.$$

Despejamos μ :

- $\mathbb{P}\left[\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$
- El **intervalo de confianza** al $100(1 - \alpha)\%$ para μ es:

$$\mu \in \left(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right).$$

- De forma equivalente: $\mu \in \bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$.

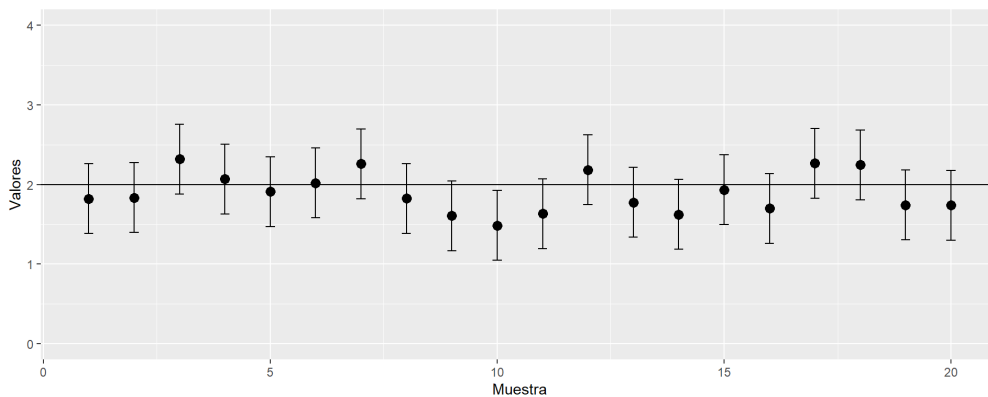
Margen de error

$z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ se llama **término de error**.

3.4.1) Interpretación

Importante

- El intervalo de confianza $\left[\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$ es un intervalo **aleatorio**.
- Al extraer una muestra, tengo una probabilidad α de que, al afirmar que μ se encuentra en $\left[\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$, me equivoque.

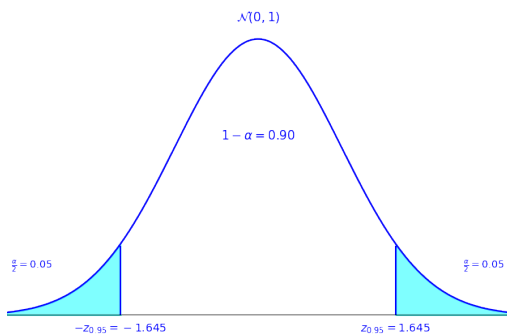


Ejemplo:

- Queremos estimar la longitud media de un artículo producido por una máquina.
- Por experiencia, sabemos que es razonable modelizar la distribución de los valores de la longitud de los artículos producidos por una distribución Normal con media μ y desviación típica a 0.05.
- Para estimar μ extraemos una muestra de 5 artículos:

20.1, 20.05, 19.95, 19.99.

- Construir un intervalo de confianza al 90%.



$$\bar{x} = 20.02$$

$$1 - \alpha = 0.90$$

$$\alpha = 0.1$$

$$\frac{\alpha}{2} = 0.05$$

$$\left(20.2 \pm 1.645 \frac{0.05}{\sqrt{5}} \right) = (20.02 \pm 0.0367) = (19.9833, 20.0588)$$

3.5) Comentarios importantes

Si X no es Normal

- Hemos trabajado con la hipótesis de que X es Normal para encontrar el estadístico pivotal $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.
- Si X no es Normal, no podemos garantizar la confianza especificada.
- Sin embargo, si n grande, tenemos por el Teorema Central del Límite

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

entonces la confianza especificada no será exacta pero casi...

3.5.1) Factores que afectan a la precisión de la estimación

El margen de error es $z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$.

- $n \uparrow \rightarrow$ precisión \uparrow
- $\sigma \uparrow \rightarrow$ precisión \downarrow
- Confianza $\uparrow \rightarrow$ precisión \downarrow

3.5.2) Determinación del tamaño muestral

Contexto

Antes de extraer la muestra:

- Tenemos decidido el valor de σ .
- Tenemos decidido la confianza con la que trabajamos.
- Tenemos decidido el margen de error máximo max que estamos dispuestos a cometer.

¿Qué tamaño de la muestra debemos escoger?

Margen de error: $z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq max. \rightarrow$ Despejamos n

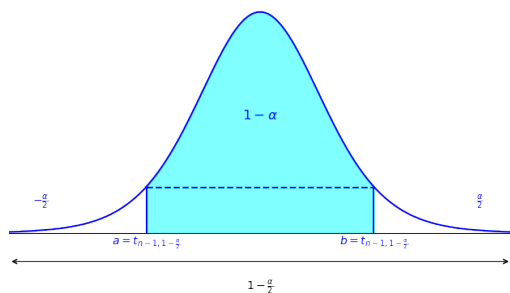
3.5.3) Otros modelos, estadísticos pivotaes

$X \sim \mathcal{N}(\mu, \sigma^2)$, estimamos μ, σ desconocida

Estadístico pivotal:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Debemos encontrar valores a y b tales que:



$$P \left[a \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq b \right]$$

$$P \left[-t_{n-1, 1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1, 1-\frac{\alpha}{2}} \right]$$

$$P \left[X - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq X + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right] = 1 - \alpha$$

Obteniendo como I.C aleatorio a nivel $100(1 - \alpha)\%$

$$X - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

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

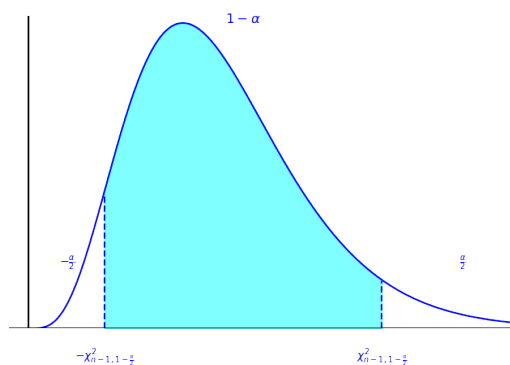
Estadístico pivotal:

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

I.C para σ^2 al nivel de confianza $(1 - \alpha)100\%$

$$T = \frac{(n-1)s^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

Debemos encontrar valores a y b tales que: $P \left[a \leq \frac{(n-1)s^2}{\sigma^2} \leq b \right] = 1 - \alpha$



$$P \left[-\chi_{n-1, \frac{1-\alpha}{2}}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2 \right] = 1 - \alpha$$

$$P \left[-\frac{1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \right] = 1 - \alpha$$

$$P \left[-\frac{(n-1)s^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \right] = 1 - \alpha$$

$X \sim \text{Bernoulli}$, estimamos p

Estadístico pivotal aproximado:

$$\frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}} \sim \mathcal{N}(0, 1) \quad \text{aprox.}$$

Diferencias de dos variables aleatorias normales

Contexto:

- $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, independientes.
- Dos muestras aleatorias simples X_1, \dots, X_{n_X} , de X e Y_1, \dots, Y_{n_Y} , de Y .
- Suponemos que conocemos σ_X^2 y σ_Y^2 .

Estadístico pivotal:

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim \mathcal{N}(0, 1).$$

Diferencias de dos variables aleatorias normales

Contexto:

- $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, independientes.
- Dos muestras aleatorias simples X_1, \dots, X_{n_X} , de X e Y_1, \dots, Y_{n_Y} , de Y .
- Desconocemos σ_X^2 y σ_Y^2 , pero las suponemos iguales, $\sigma_X^2 = \sigma_Y^2$.

Estadístico pivotal:

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(n_X-1)S_X^2 + (n_Y-1)S_Y^2}{n_X+n_Y-2}} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t_{n_X+n_Y-2}.$$

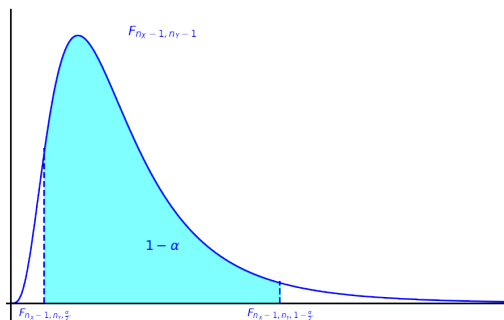
Cociente de varianzas, para dos variables aleatorias normales

Contexto:

- $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ y $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, independientes.
- Dos muestras aleatorias simples X_1, \dots, X_{n_X} , de X e Y_1, \dots, Y_{n_Y} , de Y .

Estadístico pivotal:

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



$$P \left[a \leq \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \leq b \right] = 1 - \alpha$$

$$P \left[F_{n_X-1, n_Y-1, \frac{\alpha}{2}} \leq \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \leq F_{n_X-1, n_Y-1, 1-\frac{\alpha}{2}} \right]$$

$$P \left[\frac{S_Y^2}{S_X^2} F_{n_X-1, n_Y-1, \frac{\alpha}{2}} \leq \frac{\sigma_Y^2}{\sigma_X^2} \leq \frac{S_Y^2}{S_X^2} F_{n_X-1, n_Y-1, 1-\frac{\alpha}{2}} \right] = 1 - \alpha$$

3.6) Intervalos de confianza basados en el Bootstrap

Una primera aproximación

- Podemos usar los cuantiles de una distribución normal, al igual que lo hacemos para construir un IC para la media de una normal.
- Se suele usar la aproximación si n es grande: $\frac{\hat{\theta}_n - \theta}{\hat{se}} \sim \mathcal{N}(0, 1)$.
- Por ejemplo al 95%: $\hat{\theta}_n \pm 1.96 \cdot \hat{se}$, o aproximadamente $\hat{\theta}_n \pm 2 \cdot \hat{se}$.

```
1 ingresos <- c(22010, 43950, 31416, 143200, 19298, 21521, 60025, 24320, 18221, 29528)
```

```
1 library(tidyverse)
2 B <- 200
3 lista_boot_muestras <- replicate(
4   B,
5   sample(ingresos, size = length(ingresos), replace = TRUE),
6   simplify = FALSE
7 )
8 valores_medianas <- map_dbl(lista_boot_muestras, median)
9 error_estandar_estimacion <- sd(valores_medianas)
```

En el ejemplo de los ingresos, obtenemos para la mediana: 26924 ± 12311 .

La construcción anterior es una aproximación

- Es posible, en particular si n no es grande o si la distribución es asimétrica, que el nivel de confianza no sea correcto.
- Siempre produce intervalos simétricos, lo que puede ser una limitación.
- Para mejorar la veracidad del intervalo construido, vamos a ver el intervalo de confianza basado en percentiles Bootstrap.

Procedimiento

- Habiendo observado una muestra \mathbf{x} , queremos estimar θ , usando un estimador $\hat{\theta} = T(\mathbf{x})$.
- Simulamos un gran número B de muestras Bootstrap a partir de \mathbf{x} , obteniendo $\mathbf{x}^{*1}, \dots, \mathbf{x}^{*n}$.
- Calculamos para cada muestra Bootstrap, $\hat{\theta}^{*b} = T(\mathbf{x}^{*b})$.
- Si $1 - \alpha$ es el nivel de confianza, consideramos el conjunto $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$ y calculamos:
 - $\hat{\theta}^{*(\alpha/2)}$ el percentil $100 \cdot \frac{\alpha}{2}$.
 - $\hat{\theta}^{*(1-\alpha/2)}$ el percentil $100 \cdot \left(1 - \frac{\alpha}{2}\right)$.
- Nuestro intervalo al $100(1 - \alpha)$ es $[\hat{\theta}^{*(\alpha/2)}, \hat{\theta}^{*(1-\alpha/2)}]$.

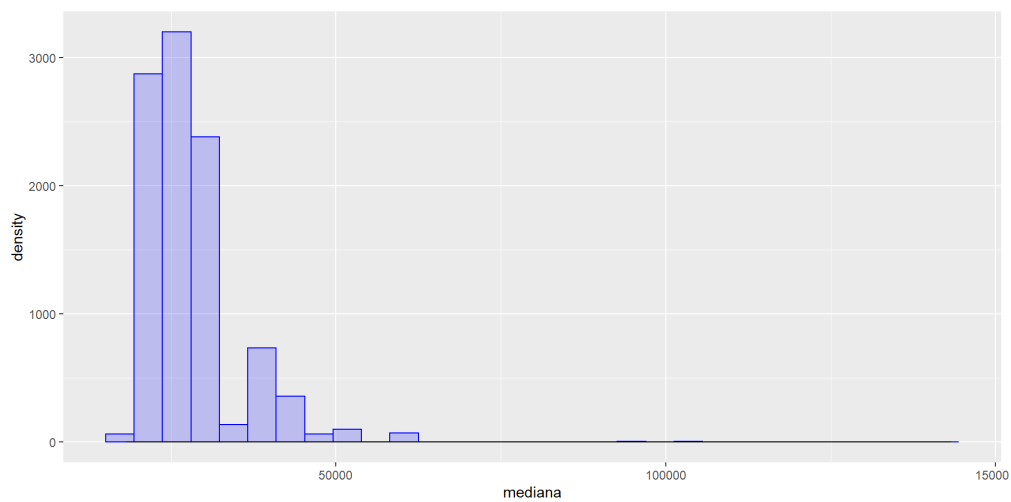
¿Cómo se calcula el percentil $100 \cdot p$?

Si tenemos un conjunto $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*B}$, B grande, para calcular el percentil $100 \cdot p$ para $0 \leq p \leq 1$, podemos seguir estos pasos:

- Ordenamos los datos de menor a mayor.
- Calculamos un índice como $i = p \times (B + 1)$, donde B es el número de observaciones.
- Si i es un número entero, podemos interpolar entre los puntos de datos adyacentes.
 - Existen distintas posibilidades para la interpolación.
 - En R: Usamos la función `quantile`, que admite hasta 9 tipos de interpolación con el argumento `type`.

3.6.1) Aplicación con el ejemplo de los ingresos

Simulamos $B = 10000$ muestras Bootstrap, calculamos las 10000 medianas $\hat{\theta}^{*1}, \dots, \hat{\theta}^{*10000}$ y hacemos un histograma:



Calculamos los percentiles 2.5% y 97.5%

```
1 str_glue("El intervalo basado en percentiles Bootstrap para la mediana es: [{formatC(
  quantile(valores_medianas, 0.025), format = 'd')}, {formatC(quantile(valores_medianas,
    0.975), format = 'd')}]")
```

El intervalo basado en percentiles Bootstrap para la mediana es: [20654, 45720]

Recordad que, con la aproximación $\hat{\theta}_n + 1 - 96 \cdot \hat{se}$,

habíamos obtenido [14612,39235]

```
1 ingresos
```

```
## [1] 22010 43950 31416 143200 19298 21521 60025 24320 18221 29528
```

Hoja de ejercicios Tema 3: Intervalos de confianza

- 1) Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ y se considera una m.a.s. de tamaño n de X .
 - a) Suponiendo σ^2 conocida obtener el intervalo de confianza para μ con nivel de confianza $1 - \alpha$. Determinar el tamaño de muestra necesario para estimar μ con una precisión $\pm \delta$.
 - b) Suponiendo σ^2 desconocida, obtener un intervalo de confianza óptimo para μ con nivel de confianza $1 - \alpha$. Determinar el tamaño de muestra necesario para estimar μ con una precisión $\pm \delta$.
- 2) Sean $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ e $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ independientes. Se considera una m.a.s. de tamaño n_1 de X y una m.a.s. de tamaño n_2 de Y .
 - a) Suponiendo σ_1^2 y σ_2^2 conocidas obtener un intervalo de confianza para $\mu_1 - \mu_2$ con nivel de confianza $1 - \alpha$.
 - b) Suponiendo σ_1^2 y σ_2^2 desconocidas pero iguales, obtener un intervalo de confianza para $\mu_1 - \mu_2$ con nivel de confianza $1 - \alpha$.
- 3) Sea $X \sim \mathcal{N}(\mu, \sigma^2)$. Se considera una muestra aleatoria simple de tamaño n de X . Suponiendo μ desconocida obtener un intervalo de confianza para σ^2 con nivel de confianza $1 - \alpha$.
- 4) Sea X una población con función de densidad

$$f(x, \theta) = \frac{2x}{\theta^2} \text{ para } x \in (0, \theta),$$

siendo θ un parámetro positivo desconocido. Se considera una muestra aleatoria simple de tamaño n de X . Obtener el intervalo de confianza óptimo para θ con nivel de confianza $1 - \alpha$.

- 5) Sea X una población con función de densidad

$$f(x, \theta) = \exp(-(x - \theta)), \text{ para } x \in (\theta, +\infty),$$

siendo θ un parámetro positivo desconocido. Se considera una muestra aleatoria simple de tamaño n de X . Obtener el intervalo de confianza óptimo para θ con nivel de confianza $1 - \alpha$.

- 6) En un laboratorio, se estudia la velocidad de combustión de dos tipos de combustibles sólidos A y B. Se pretende determinar qué combustibles presentan una velocidad de combustión en promedio mayor de 50 cm/s. Los resultados obtenidos fueron:

Combustible	Tamaño muestral	Media muestral	Varianza muestral
A	25	51.3	3.9
B	30	51.5	3.6

Suponiendo que la distribución de las variables es normal, se pide:

- a) Obtener un intervalo de confianza al 95% para la velocidad esperada para el combustible A y otro para el combustible B.
- b) Obtener un intervalo de confianza al 95% para la varianza de la velocidad de combustión del combustible A.
- c) Obtener un intervalo de confianza al 95% para el cociente de las varianzas σ_A^2/σ_B^2 .
- d) Suponiendo que las varianzas σ_A^2 y σ_B^2 son iguales, construir un intervalo de confianza para la diferencia entre la velocidad esperada para el combustible A y la velocidad esperada para el combustible B.

Tema 4: Contrastes de hipótesis

4.1) Introducción

Basándonos en una muestra, queremos decidir entre dos opciones o hipótesis acerca de la distribución de interés en la población.

Las hipótesis pueden verter sobre

- La forma de la distribución: ¿Es una Poisson? ¿Es una Normal? [Contrastes no paramétricos](#).
- Los parámetros de la distribución, si suponemos dada la familia paramétrica. [Contrastes paramétricos](#).
- Si consideramos dos variables, ¿existe una relación de dependencia? [Contrastes no paramétricos](#).

4.2) Contrastes de hipótesis paramétricos

4.2.1) Elementos principales

Idea básica

Un **test de hipótesis** es una regla que nos lleva a decidir si lo observado en la muestra es compatible con una hipótesis planteada sobre los parámetros del modelo.

Ejemplo

Un nuevo medicamento pretende mejorar un medicamento del mercado, que tiene probabilidad $\theta_0 = 0.7$ de curar un paciente.

- Planteamos $X \sim \text{Bernoulli}(\theta)$, $X = 1 \iff$ el paciente se cura con el medicamento nuevo.
- Lanzamos el nuevo medicamento si $\theta > \theta_0 + 0.15$, es decir, que debemos decidir entre

$$H_0 : \theta \leq \theta_0 + 0.15$$

$$H_1 : \theta > \theta_0 + 0.15$$

- Consideramos una m.a.s. de tamaño 100, el estadístico $T(X_1, \dots, X_{100}) = \sum_{i=1}^{100} X_i$, resumen de la información muestral.
- Posible test:

$$\rightarrow \text{Aceptamos } H_0 \text{ si } \sum_{i=1}^{100} X_i \leq 85.$$

$$\rightarrow \text{Rechazamos } H_0 \text{ si } \sum_{i=1}^{100} X_i > 85.$$

Si queremos ser más conservadores: aceptamos H_0 si $\sum_{i=1}^{100} X_i \leq 90$, y rechazamos H_0 si $\sum_{i=1}^{100} X_i > 90$.

Podemos cometer dos tipos de errores:

- Rechazar una hipótesis verdadera.
- Aceptar una hipótesis falsa.

4.2.2) Formulación general del contraste de hipótesis

- Contexto

- $X \sim F(x, \theta)$, con $\theta \in \Theta \subseteq \mathbb{R}^p$.
- $\mathbf{X} = (X_1, X_2, \dots, X_n)$ una m.a.s. de X .
- Θ_0 y Θ_1 dos subconjuntos del espacio paramétrico Θ tales que: $\Theta = \Theta_0 \cup \Theta_1$ y $\Theta_0 \cap \Theta_1 = \emptyset$
- **Definición**

Un **test** o **contraste** es una regla basada en \mathbf{X} para decidir entre las dos hipótesis:

$$H_0 : \theta \in \Theta_0$$

frente a

$$H_1 : \theta \in \Theta_1.$$

Llamamos

- H_0 : hipótesis nula.
- H_1 : hipótesis alternativa.

Definición

Se llama **test (no aleatorizado)** a una función $\delta : \text{sop}(\mathbf{X}) \rightarrow \{0, 1\}$ tal que:

- si $\delta(\mathbf{x}) = 0$ se acepta la hipótesis nula H_0 .
- si $\delta(\mathbf{x}) = 1$ se rechaza la hipótesis H_0 .

Notación

- $\text{sop}(\mathbf{X})$ es el soporte de \mathbf{X} , que es el conjunto donde $f(x)$, la función puntual de probabilidad (caso discreto) o la función densidad (caso continuo) de X es > 0 .
- Denotaremos por T el conjunto de tests definidos sobre el $\text{sop}(\mathbf{X})$.

Dado un test $\delta \in T$ se definen los siguientes conjuntos:

Definición

- **Región de aceptación:**

$$S_0 = \{\mathbf{x} \in \text{sop}(\mathbf{X}) : \delta(\mathbf{x}) = 0\}.$$

- **Región de rechazo:**

$$S_1 = \{\mathbf{x} \in \text{sop}(\mathbf{X}) : \delta(\mathbf{x}) = 1\}.$$

4.2.2.1) Errores

- Definimos los errores asociados

- Error de tipo I: rechazar H_0 cuando es cierta.
- Error de tipo II: aceptar H_0 cuando es falsa.

Resumen

	$\mathbf{X} \in S_1$	$\mathbf{X} \in S_0$
$\theta \in \Theta_0$	Error tipo I	Correcto
$\theta \in \Theta_1$	Correcto	Error tipo II

Recordad, S_0 región de aceptación, S_1 región de rechazo.

- Definimos las probabilidades asociadas, para un test δ dado,

- La probabilidad de cometer un error de tipo I, si el valor del parámetro es θ , se denota por:

$$\text{Si } \theta \in \Theta_0 : \quad P_{I,\delta}(\theta) = P_\theta(\mathbf{X} \in S_1).$$

- La probabilidad de cometer un error de tipo II, si el valor del parámetro es θ , se denota por:

$$\text{Si } \theta \in \Theta_1 : \quad P_{II,\delta}(\theta) = P_\theta(\mathbf{X} \in S_0).$$

¿Qué buscamos?

Queremos encontrar un test δ con $P_{I,\delta}(\theta)$ y $P_{II,\delta}(\theta)$ lo más pequeñas posibles [simultáneamente](#).

Ejemplo de test: control de calidad

Contexto

- En la producción de un artículo electrónico, se cuida que la tensión que circula por un componente es 10mV. Hay variabilidad en la tensión conseguida en los artículos producidos, que pensamos modelizar por una Normal con desviación típica 0.05.
- Para controlar la calidad de los artículos producidos, se escogen al azar cada día 20 artículos y se mide su tensión.
- ¿Qué regla podríamos fijar que describa cuándo debe saltar una alarma si parece que la tensión se aleja de 10mV?

Traducimos el contexto

- X : tensión medida en un artículo escogido al azar en la producción. $X \sim \mathcal{N}(\mu, 0.025)$.
- $\mathbf{X} = (X_1, X_2, \dots, X_{20})$ una m.a.s. de X de tamaño 20.
- Queremos controlar si $\mu = 10$.

- Planteamos las hipótesis:

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

- Nuestra regla, es decir nuestro test δ , se basará en $\bar{\mathbf{X}}$:

→ si $\bar{\mathbf{X}}$ está "alejado" de 10, rechazaremos H_0 .

→ si $\bar{\mathbf{X}}$ está "cerca" de 10, aceptaremos H_0 .

Nuestra regla

Nos fijamos un umbral u ,

- si $\bar{\mathbf{X}} < 10 - u$ ó $\bar{\mathbf{X}} > 10 + u$, rechazaremos H_0 .
- si $10 - u \leq \bar{\mathbf{X}} \leq 10 + u$, aceptaremos H_0 .
- $\delta(\mathbf{x}) = 1$, si $|\bar{\mathbf{X}} - 10| > u$.
- $\delta(\mathbf{x}) = 0$, si $|\bar{\mathbf{X}} - 10| \leq u$.

Calculamos las probabilidades de error

Si $\mu = 10$,

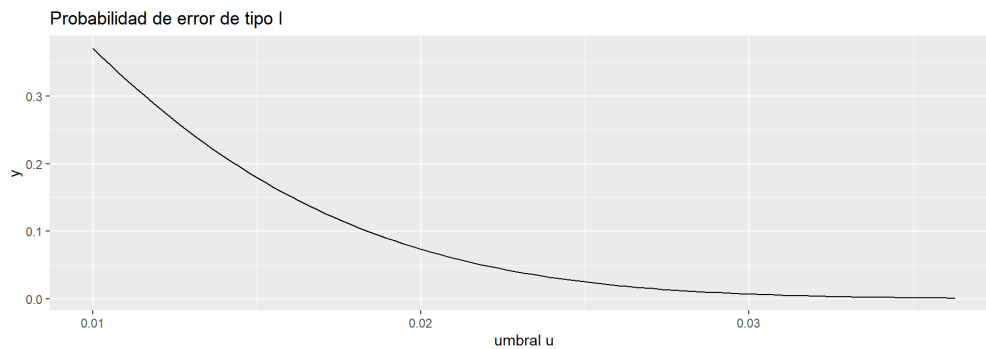
$$P_{\mu}(\text{rechazar } H_0) = P_{\mu}(|\bar{X} - 10| > u) = 2P\left(Z > \frac{u}{0.05/\sqrt{20}}\right)$$

Si $\mu \neq 10$,

$$P_{\mu}(\text{aceptar } H_0) = P_{\mu}(|\bar{X} - 10| \leq u) = P\left(\frac{10 - u - \mu}{0.05/\sqrt{20}} \leq Z \leq \frac{10 + u - \mu}{0.05/\sqrt{20}}\right)$$

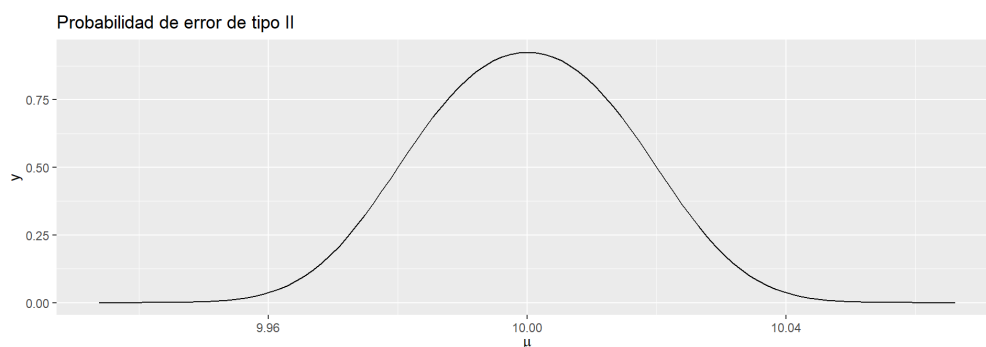
Por ejemplo, si fijamos $u = 0.025$, P_{μ}

Podemos variar u :

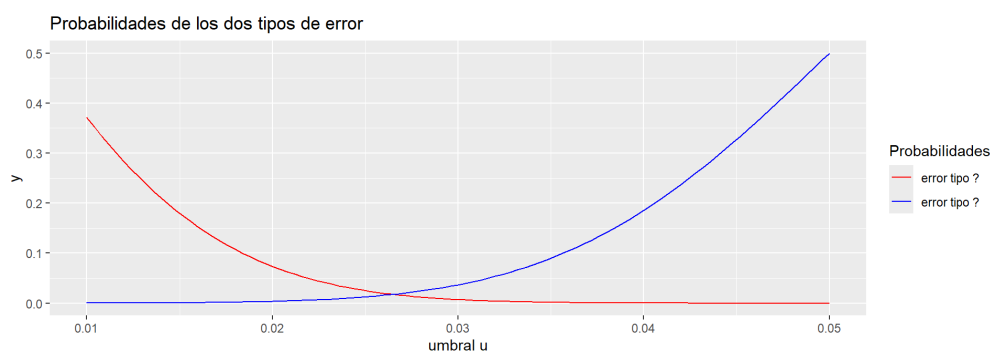


Si fijamos $u = 0.02$, tenemos una probabilidad de error de tipo I de 0.05.

¿Qué pasa con la probabilidad de error de tipo II?



Queremos detectar $\mu = 10.05$ con una probabilidad suficiente pero a la vez, no queremos rechazar H_0 erróneamente con demasiada facilidad.



4.3) Métodos útiles para construir tests

4.3.1) Un primer procedimiento general

Para llevar a cabo un contraste de hipótesis, podemos

- Formular las hipótesis H_0 y H_1 .
- Fijarnos la probabilidad de error de tipo I, α .
- Para construir nuestra regla δ , escogemos el estadístico de prueba $T(X_1, \dots, X_n)$ basado generalmente en un estimador del parámetro. Consideramos su distribución muestral suponiendo H_0 cierta.
- Determinamos la región de rechazo S_1 , teniendo en cuenta la probabilidad de error I, es decir,

$$P_{H_0}(T(X_1, \dots, X_n) \in S_1) = \alpha.$$

- Para nuestra muestra, calculamos $T(x_1, \dots, x_n)$ y decidimos si aceptamos o rechazamos H_0 , según si su valor está en S_0 o en S_1 .

4.4) Contraste de hipótesis para la media μ

Primer caso

Empezamos por considerar la construcción de un contraste para la media μ de una población Normal con varianza conocida.

Contexto

- Consideramos una v.a. X .
- Hemos decidido modelar: $X \sim \mathcal{N}(\mu, \sigma^2)$.
Tenemos fijado el valor de σ (valor poblacional, por ejemplo gracias a datos históricos).
- Queremos llevar a cabo un contraste sobre μ .
- Para ello, extraeremos una muestra de tamaño n de la distribución de X .

4.5) Procedimiento, hipótesis bilateral

- Formulamos las hipótesis:

$$\begin{cases} H_0 : \mu = \mu_0, \\ H_1 : \mu \neq \mu_0, \end{cases}$$

donde μ_0 representa el valor concreto con el que queremos comparar μ .

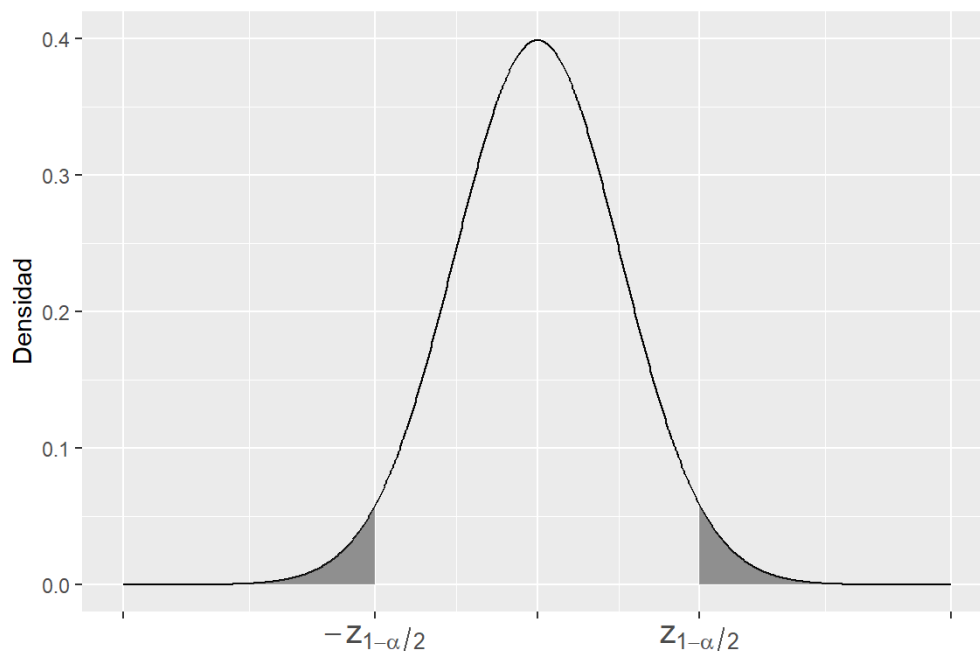
- Fijamos el valor de α .
- El estadístico de prueba es la versión tipificada de \bar{X} :

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

- Si H_0 es cierto, $\mu = \mu_0$

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

- Establecemos la región de rechazo:



La región S_1 está formada por los valores menores que $-z_{1-\frac{\alpha}{2}}$ o mayores que $z_{1-\frac{\alpha}{2}}$.

- Nos queda calcular, para nuestra muestra, el valor concreto del estadístico de prueba z_0 :
 - Si pertenece a S_1 , rechazaremos H_0 y afirmaremos H_1 .
 - Si no pertenece a S_1 , admitiremos H_0 .

Ejemplo

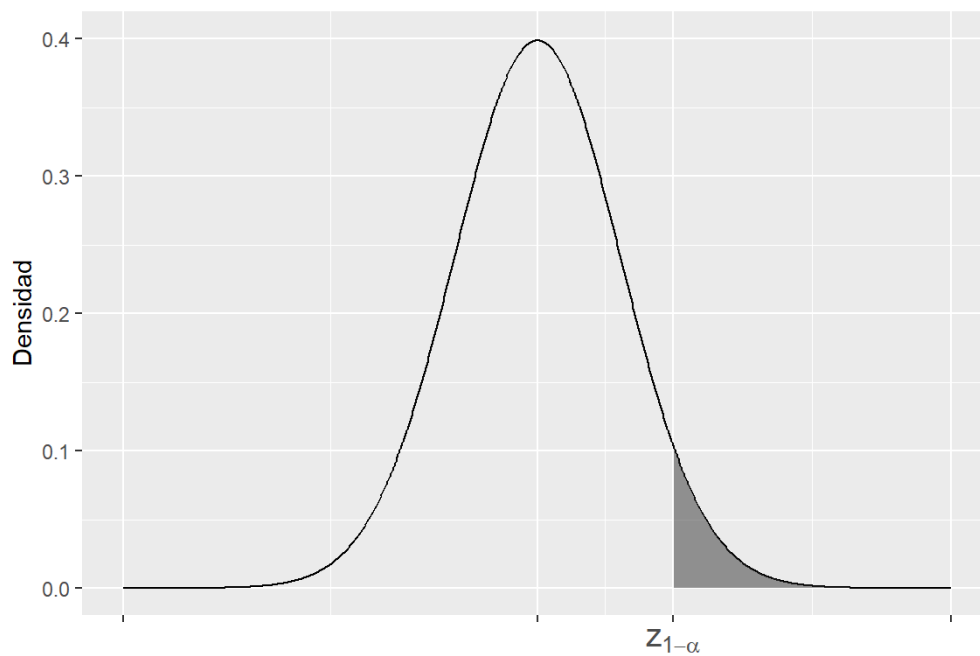
En un proceso de producción

- La longitud de los artículos producidos se modeliza a través de una distribución Normal con media μ .
- Por experiencia acerca del proceso, se cuantifica su desviación típica en $\sigma = 1mm$.
- En condiciones de funcionamiento correcto, se espera que la longitud media de los artículos sea 50 mm.
- Para comprobar la calidad se decide tomar una muestra de 10 artículos que resultan tener la longitud media \bar{x} igual a 51 mm.
- Basándonos en esta muestra, ¿qué podemos decir acerca del funcionamiento del proceso?

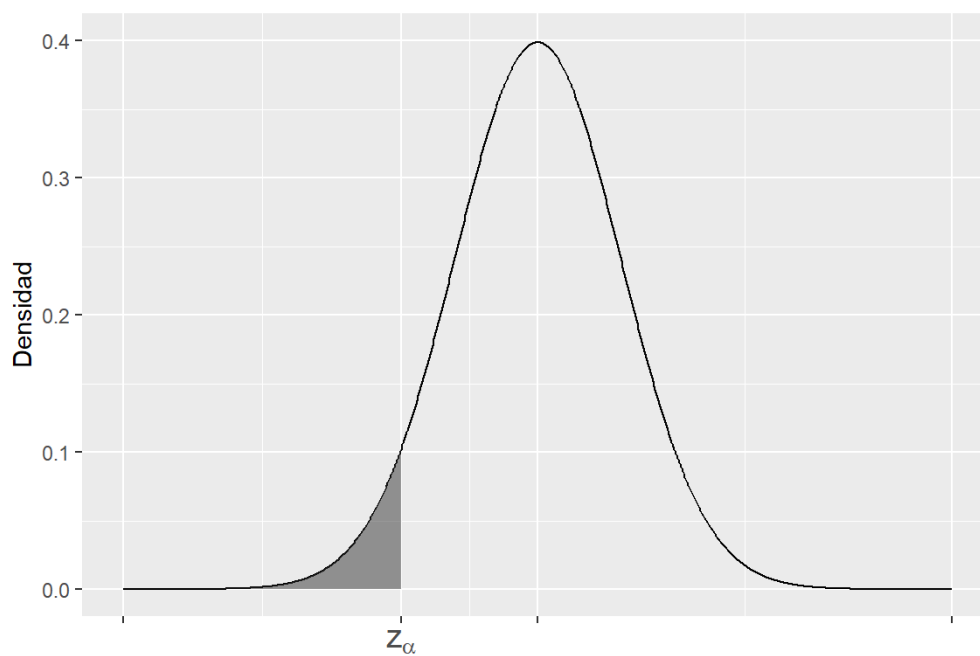
4.6) Procedimiento, hipótesis unilateral

Todo igual, excepto S_1 .

- Si la hipótesis es $H_1 : \mu > \mu_0$, la región de rechazo será



- Si la hipótesis es $H_1 : \mu < \mu_0$, la región de rechazo será



4.7) El p-valor

Definición

Consideramos un contraste, un test, y hemos observado una muestra concreta \mathbf{x} .

- El **p-valor** es el nivel de significación más pequeño que nos lleve, para esa muestra \mathbf{x} , a rechazar la hipótesis nula.
- El p-valor está asociado a la región S_1 más pequeñas que nos lleve, para esa muestra, a rechazar H_0 .

Ejemplo

Volvamos al ejemplo de la longitud de los artículos producido. Para el contraste

$$H_0 : \mu = 50,$$

$$H_1 : \mu \neq 50.$$

Supongamos que hemos encontrado un estadístico de prueba $z_0 = 1.75$ y rechazamos H_0 al 95% de confianza. ¿Cuál habría sido nuestra decisión al 90%? ¿Y al 99% de confianza?

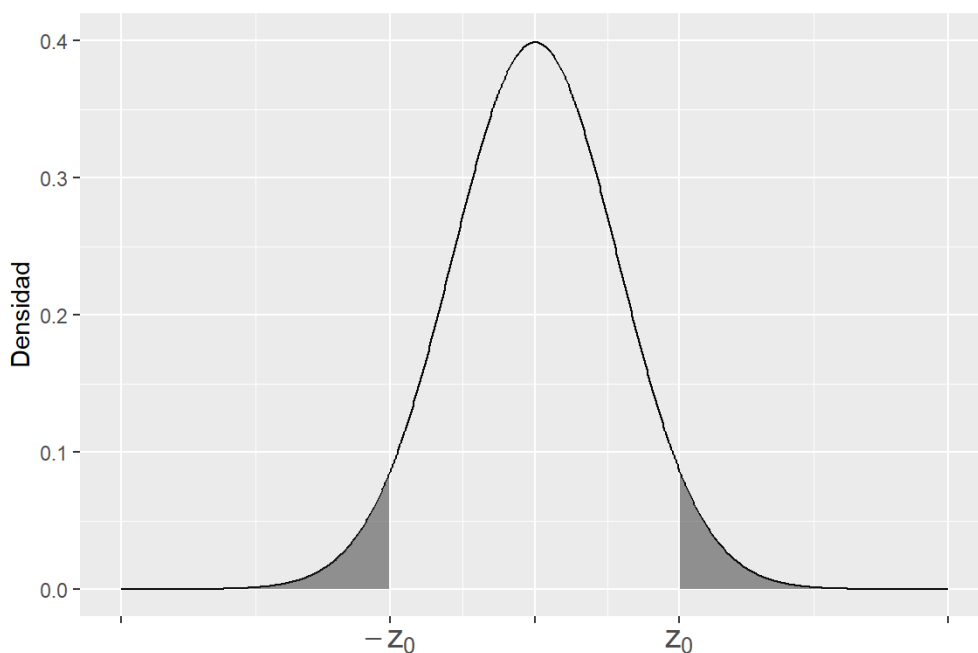
Importante

- Si rechazamos H_0 a un nivel de confianza dado, también la rechazamos a cualquier nivel de confianza menor.
- Para saber si la rechazamos a un nivel de confianza mayor, hay que probar.

Para calcularlo:

- Dibujamos la densidad de la distribución muestral del estadístico de prueba.
- Colocamos el valor de z_0 calculado para mi muestra concreta.
- Dibujamos una región de rechazo de manera que z_0 coincida con una de las fronteras.
- El p-valor α_0 es el valor de α correspondiente a esta región de rechazo.

Para el ejemplo de las longitudes, sabemos $z_0 = 1.75$, lo colocamos:



Por lo que $\frac{\alpha_0}{2} = \mathbb{P}(Z \geq 1.75)$, es decir que $\alpha_0 = 2(1 - \phi(1.75))$, lo que lleva a $\alpha_0 \simeq 0.08$.

Para nuestro ejemplo

- Encontramos un p-valor igual a 0.08, por lo que deducimos que la confianza máxima con la que podríamos haber rechazado H_0 es del 92%.
- Deducimos que habríamos en particular rechazado H_0 al 90% de confianza, pero, por ejemplo, no al 95% de confianza.

Cuidado con la interpretación del p-valor

El p-valor se puede interpretar como una medida de compatibilidad de los datos con la hipótesis nula.

- No debemos usarlo como el único criterio para basar nuestra decisión.
- Es bueno hacer gráficas, considerar el contexto, ver si el efecto es relevante y no solamente significativo.
- Se debe evitar considerar $p\text{-valor} < 0.05$ como un valor mágico que es sinónimo de efecto significativo.

4.8) Test óptimo

Empezamos por considerar los tests con probabilidad de error tipo I menor o igual que α :

- Dado α , sea

$$T_\alpha = \{\delta \in T \mid P_{i,\delta}(\theta) \leq \alpha \text{ para todo } \theta \in \Theta_0\},$$

llamaremos α el nivel de significación del test.

- Para un test δ , se llama tamaño o extensión de λ a: $\sup_{\theta \in \Theta_0} P_{I,\delta}(\theta)$.

Definición

Dado $\delta^* \in T_\alpha$ decimos que es óptimo si

$$P_{II,\delta^*}(\theta) \leq P_{II,\delta}(\theta) \text{ para todo } \delta \in T_\alpha \text{ y para todo } \theta \in \Theta_1.$$

4.9) Potencia de un test

Definición

La función potencia de un test δ asocia a cada θ la probabilidad de rechazar la hipótesis nula:

$$\theta \mapsto \pi_\delta(\theta) = P_\theta(X \in S_1), \quad \text{con } \theta \in \Theta.$$

Relación de la potencia con los errores

- Si $\theta \in \Theta_0$ entonces $\pi_\delta(\theta) = P_{I,\delta}(\theta)$.
- Si $\theta \in \Theta_1$ entonces

$$\pi_\delta(\theta) = P_\theta(X \in S_1) = 1 - P_\theta(X \in S_0) = 1 - P_{II,\delta}(\theta).$$

4.10) Test óptimo como el de máxima potencia

Sea $\delta^* \in T_\alpha$, diremos que es el test uniforme de máxima potencia para el contraste

$$H_0 : \theta \in \Theta$$

frente a

$$H_1 : \theta \in \Theta_1$$

si se verifica

$$\pi_{\delta^*}(\theta) \geq \pi_\delta(\theta) \text{ para todo } \delta \in T_\alpha \text{ y para todo } \theta \in \Theta_1.$$

4.11) Procedimientos basados en la verosimilitud

Un procedimiento natural

Puesto que la verosimilitud expresa la compatibilidad de los datos con el modelo considerado, es natural construir un test basado en la comparación de la verosimilitud para los valores de Θ_0 y para los valores de Θ_1 .

Constrastes de hipótesis simples

En este tipo de contraste los subconjunto del espacio paramétrico solo tienen un elemento, es decir $\Theta_0 = \{\theta_0\}$ y $\Theta_1 = \{\theta_1\}$ y por lo tanto los errores de ambos tipos son cantidades fijas. Implícitamente, estamos asumiendo que el parámetro es unidimensional.

4.12) Constrastes de hipótesis simples

Lema de Neyman-Pearson

Sea $X \sim F(x, \theta)$, $\Theta = \{\theta_0, \theta_1\}$, $\mathbf{X} = (X_1, X_2, \dots, X_n)$, una muestra aleatoria simple de la población X , α fijo con $0 < \alpha < 1$. Para contrastar la hipótesis nula $H_0 : \theta = \theta_0$ frente a la hipótesis alternativa $H_1 : \theta = \theta_1$, el test de región de rechazo:

$$S_1 = \left\{ \mathbf{x} \in \text{Sop}(\mathbf{X}) \mid \frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)} \geq k \right\}$$

con $k > 0$, que además tenga tamaño α , es decir $\alpha = P(\mathbf{X} \in S_1 | \theta = \theta_0)$, cumple que es el test de máxima potencia en la clase de los tests T_α .

• Observaciones

- Si la distribución es de tipo continuo siempre se cumple que el test alcanza el tamaño obteniéndose el test de extensión de α .
- Si la distribución es de tipo discreto, en general, el test de máxima potencia será de extensión menor que α .

• Demostración

Sea δ^* el test definido anteriormente.

- La condición

$$\alpha = P(\mathbf{X} \in S_1 | \theta = \theta_0)$$

nos asegura que $\delta^* \in T_\alpha$. Sea ahora $\delta \in T_\alpha$ cualquiera. Tenemos que probar que:

$$\pi_{\delta^*}(\theta_1) \geq \pi_\delta(\theta_1)$$

supondremos X variable aleatoria continua (en el caso discreto se razonará de forma análoga).

- Sea $\mathbf{X} \in \text{sop}(X)$ se verifica que:

$$(\delta^*(\mathbf{X}) - \delta(\mathbf{X}))(L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0)) \geq 0.$$

- Si $\delta(\mathbf{X})^* = 1$ entonces:

$$\frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)} \geq k \iff L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0) \geq 0$$

y $\delta^*(\mathbf{X}) - \delta(\mathbf{X}) \geq 0$ ($\delta(\mathbf{X})$ sólo puede ser 0 o 1) y el producto es no negativo.

- Si $\delta(\mathbf{X})^* = 0$ entonces:

$$\frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta_0)} < k \iff L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0) < 0$$

y $\delta^*(\mathbf{X}) - \delta(\mathbf{X}) \leq 0$ ($\delta(\mathbf{X})$ sólo puede ser 0 o 1) y el producto es no negativo.

- La integral de la expresión anterior sobre el $\text{sop}(\mathbf{X})$ será no negativa:

$$\int_{\text{sop}(\mathbf{X})} (\delta^*(\mathbf{X}) - \delta(\mathbf{X}))(L(\mathbf{x}, \theta_1) - kL(\mathbf{x}, \theta_0))d\mathbf{x} \geq 0$$

$$\int_{\text{sop}(\mathbf{X})} \delta^*(\mathbf{X})L(\mathbf{x}, \theta_1)d\mathbf{x} - \int_{\text{sop}(\mathbf{X})} \delta(\mathbf{X})L(\mathbf{x}, \theta_1)d\mathbf{x} - k \int_{\text{sop}(\mathbf{X})} \delta^*(\mathbf{X})L(\mathbf{x}, \theta_0)d\mathbf{x} - \int_{\text{sop}(\mathbf{X})} \delta(\mathbf{X})L(\mathbf{x}, \theta_0)d\mathbf{x} \geq 0$$

notar que ambos contrastes solo toman el valor 1 en sus respectivas regiones de rechazo (y cero en las regiones de aceptación) con lo que al desarrollar los terminos de la integral tenemos:

$$\pi_{\delta^*} - \pi_{\delta}(\theta_1) - k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \geq 0.$$

- Como $\delta \in T_{\alpha}$ tenemos que $\pi_{\delta}(\theta_0) \leq \alpha$ con lo que:

$$0 \leq \alpha - \pi_{\delta}(\theta_0) = \pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)$$

y por lo tanto:

$$k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \geq 0$$

con lo que:

$$0 \leq \pi_{\delta^*}(\theta_1) - \pi_{\delta}(\theta_1) - k(\pi_{\delta^*}(\theta_0) - \pi_{\delta}(\theta_0)) \leq \pi_{\delta^*}(\theta_1) - \pi_{\delta}(\theta_1)$$

de donde se deduce:

$$\pi_{\delta^*}(\theta_1) \geq \pi_{\delta}(\theta_1)$$

que era lo que queríamos probar.

4.13) Hipótesis compuestas

Planteamiento

En las aplicaciones a problemas reales no suelen plantearse, en general, contrastes de hipótesis simple y alternativa simple, pues las hipótesis alternativas no suelen ser tan precisas para que estén definidas por un único valor del parámetro. Consideremos ahora contrastes de hipótesis donde al menos una de ellas es compuesta.

Casos

a)	$H_0 : \theta = \theta_0$ $H_1 : \theta > \theta_0$	d)	$H_0 : \theta \geq \theta_0$ $H_1 : \theta < \theta_0$
b)	$H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$	e)	$H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$
c)	$H_0 : \theta \leq \theta_0$ $H_1 : \theta > \theta_0$	f)	$H_0 : \theta \in [\theta_1, \theta_2]$ $H_1 : \theta \notin [\theta_1, \theta_2]$

En los casos a)-d) anteriores es posible obtener tests uniformes de máxima potencia para casos en los que la variable pertenece a la familia exponencial uniparamétrica.

4.14) Cociente de verosimilitudes monótono

Resultado

Sea $X \sim F(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$ y X_1, \dots, X_n una muestra aleatoria simple de X . Decimos que X o F tienen la propiedad de cociente de verosimilitud monótono en el estadístico $R = R(X_1, \dots, X_n)$, si para todo $\theta_0, \theta_1 \in \Theta$ con $\theta_0 < \theta_1$ se verifica que el cociente de verosimilitudes $\frac{L(\mathbf{x}, \theta_1)}{L(\mathbf{x}, \theta)}$, es una función monótona creciente de $R(\mathbf{x})$.

Nota

A los efectos de la definición si $L(\mathbf{x}, \theta_0) = 0$ y $L(\mathbf{x}, \theta_1) > 0$, el cociente anterior se considerará igual a infinito.

Familia exponencia uniparamétrica

Se dice que la variable $X \sim F(x, \theta)$ pertenece a la familia exponencia uniparamétrica si su función de verosimilitud se puede escribir en la forma:

$$L(\mathbf{x}, \theta) = \exp\{A(\theta)T(\mathbf{x}) + B(\theta) + h(\mathbf{x})\}$$

para toda muestra \mathbf{x} .

Si $A(\theta)$ es monótona creciente (decreciente) en θ , entonces X tiene cociente de verosimilitud monótono en el estadístico $T(X_1, \dots, X_n)(-T(X_1, \dots, X_n))$.

Cociente de verosimilitudes monótono

Sea $X \sim F(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$ y X_1, \dots, X_n una m.a.s de X . Supongamos que X tiene cociente de verosimilitud monótono en el estadístico $R = R(X_1, \dots, X_n)$ para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$, se sigue que el test

$$\delta_1(\mathbf{x}) = \begin{cases} 0 & \text{si } R(\mathbf{x}) < c \\ 1 & \text{si } R(\mathbf{x}) \geq c \end{cases}$$

que cumpla $P_{I, \delta_1}(\theta_0) = \alpha$, es el test uniforme de máxima potencia en la clase de tests $T_\alpha = \{\delta \in T | P_{I, \delta}(\theta_0) \leq \alpha\}$, con $0 < \alpha < 1$.

El teorema anterior se extiende al constraente $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$.

Cociente de verosimilitudes monótono

Sea $X \sim F(x, \theta)$, $\theta \in \Theta \subset \mathbb{R}$ y X_1, \dots, X_n una m.a.s de X . Supongamos que X tiene cociente de verosimilitud monótono en el estadístico $R = R(X_1, \dots, X_n)$ para contrastar $H_0 : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$, se sigue que el test

$$\delta_1(\mathbf{x}) = \begin{cases} 0 & \text{si } R(\mathbf{x}) > c \\ 1 & \text{si } R(\mathbf{x}) \leq c \end{cases}$$

que cumpla $P_{I, \delta_1}(\theta_0) = \alpha$, es el test uniforme de máxima potencia en la clase de tests $T_\alpha = \{\delta \in T | P_{I, \delta}(\theta_0) \leq \alpha\}$, con $0 < \alpha < 1$.

El teorema anterior se extiende al constraente $H_0 : \theta \geq \theta_0$ frente a $H_1 : \theta < \theta_0$.

4.15) Contrastes de hipótesis compuesta

La metodología anterior no se puede aplicar a los casos e) y f) y en los casos de parámetros que no sean unidimensionales. Por tanto, es importante para muchos problemas prácticos, disponer de métodos de construcción de contrastes, basados en principios razonables y que permitan diseñar el test según el problema planteado. El método más importante lo constituye el **test de la razón de verosimilitudes generalizado**.

Razón de verosimilitudes Generalizada (RVG)

Se llama razón de verosimilitudes generalizada para contrastar la hipótesis $H_0 : \theta \in \Theta$ frente $H_1 : \theta \in \Theta_1$ al cociente

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta_1} L(\mathbf{x}; \theta)}.$$

4.16) Procedimiento basado en la verosimilitud

Test de razón de verosimilitud generalizada

Un test de razón de verosimilitud generalizada (RVG) para las hipótesis anteriores es aquel que tiene región crítica de la forma

$$S_1 = \left\{ \mathbf{x} \in \text{Sop}(X_1, \dots, X_n) \mid \lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{x}; \theta)}{\sup_{\theta \in \Theta} L(\mathbf{x}; \theta)} < k \right\}$$

Determinamos la constante k , imponiendo que el test tenga extensión α , es decir:

$$\sup_{\theta \in \Theta_0} P((X_1, \dots, X_n) \in S_1 | \theta) = \alpha.$$

4.17) Test RVG para la media de una población normal con varianza conocida

Ejercicio

- $X \sim \mathcal{N}(\mu, \sigma^2)$, suponemos σ conocida.
- M.a.s X_1, \dots, X_n de X .
- Formulamos las hipótesis

$$H_0 : \mu = \mu_0$$

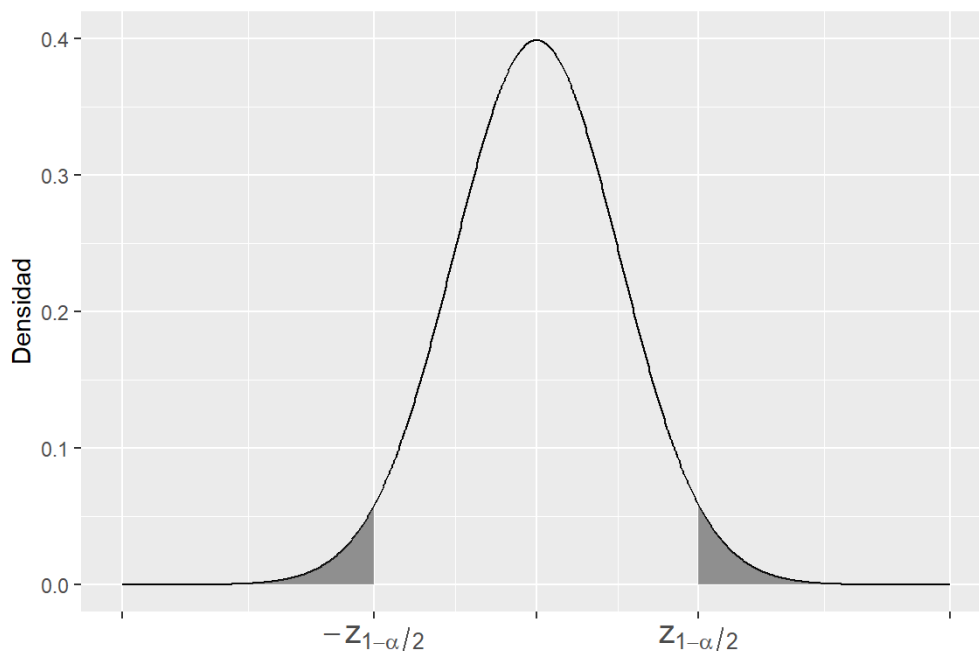
$$H_1 : \mu \neq \mu_0$$

- 1) Demostrar que el test que rechaza H_0 si $\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| > z_{1-\frac{\alpha}{2}}$ es de extensión α .
- 2) Demostrar que el test anterior es el test de RVG.

¿Es de extensión α ?

- Hemos de comprobar que: $\sup_{\theta \in \Theta_0} P_{i,\delta}(\theta) = \alpha$.
- En este caso H_0 es simple, si $\mu = \mu_0$ se cumple que $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$.

$$P\left(\left|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}\right| > z_{1-\frac{\alpha}{2}} \mid \mu_0\right) = \alpha.$$



¿Es el test de RVG?

$$\lambda(\mathbf{x}) = \frac{\sup_{\mu=\mu_0} L(\mathbf{x}; \mu)}{\sup_{\mu \in R} L(\mathbf{x}; \mu)}.$$

- Numerador:

$$L(\mathbf{X}; \mu_0) = \frac{1}{\sigma^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2 \right)$$

- El supremos en el denominador se alcanza en el EMV $\hat{\mu} = \bar{x}_n$.

$$\lambda(\mathbf{x}) = \exp \left(-\frac{n}{2\sigma^2} \sum_{i=1}^n (\bar{x}_n - \mu_0)^2 \right)$$

¿Es el test de RVG?

- Determinamos la constante k , imponiendo que el test tenga extensión α , es decir,

$$\begin{aligned} P(\lambda(\mathbf{x}) < k | \mu_0) &= \alpha \\ \alpha &= P \left(0 < \exp \left(-\frac{n}{2\sigma^2} \sum_{i=1}^n (\bar{x}_n - \mu_0)^2 \right) < k | \mu_0 \right) \\ &= P \left(-\infty < -\frac{n}{2\sigma^2} \sum_{i=1}^n (\bar{x}_n - \mu_0)^2 < \log(k) | \mu_0 \right) \\ &= P \left(\left(\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right)^2 > -2 \log(k) | \mu_0 \right) \\ &= P \left(\left| \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right| > \sqrt{-2 \log(k)} | \mu_0 \right) = P \left(\left| \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}} \right| > c | \mu_0 \right) \end{aligned}$$

tomamos $c \Leftarrow 1 - \frac{\alpha}{2}$.

4.18) Test de razón de verosimilitudes generalizada

Lo más difícil

Aunque el test RVG sea un test que suele tener buenas propiedades en cuanto a errores, para muchos modelos, no es sencillo deducir su distribución muestral.

Aproximación asintótica de la distribución muestral para el test RVG

Para el contraste

$$H_0 : \theta \in \Theta_0$$

frente a la alternativa

$$H_1 : \theta \in \Theta_1$$

si $\dim(\Theta) = q$ y $\dim(\Theta_0) = q'$ se verifica, bajo ciertas condiciones, que el estadístico $-2 \log \lambda(\mathbf{X})$ converge en distribución a una ji-cuadrado con $q - q'$ grados de libertad.

4.19) Test de bondad de ajuste

Es un contraste no paramétrico

- Su objetivo es comprobar hasta qué punto los datos observados parecen haber sido generados por una distribución de probabilidad dada.
- Se conoce como "contraste chi-cuadrado", porque es la distribución que usamos para construir nuestra regla de decisión.

Primer ejemplo, el caso más sencillo:

- Tenemos k clases posibles asociadas a un experimento aleatorio: A_1, \dots, A_k .
- Tenemos unos valores esperados para sus probabilidades

$$p_1 = P(A_1), \dots, p_k = P(A_k).$$

- Observamos n realizaciones del experimento y queremos decidir si lo observado es compatible con los valores esperados.

Caso sencillo bondad de ajuste

El caso más sencillo:

- Planteamos

$$H_0 : P(A_i) = p_i, i = 1, \dots, k$$

$$H_1 : \text{Existe } i \text{ tal que } P(A_i) \neq p_i.$$

- Tenemos k clases posibles asociadas a un experimento aleatorio: A_1, \dots, A_k .
- Tenemos unos valores esperados para sus probabilidades

$$p_1 = P(A_1), \dots, p_k = P(A_k).$$

- Observamos n realizaciones del experimento y queremos decidir si lo observado es compatible con los valores esperados.

Caso sencillo bondad de ajuste: grupo sanguíneo

Cuatro grupos sanguíneos

Experimento: escoger una persona

- $A_1 = \text{"Su grupo es O"}, A_2 = \text{"Su grupo es A"}, A_3 = \text{"Su grupo es B"}, A_4 = \text{"Su grupo es AB"}$.
- Los valores esperados para sus probabilidades son

$$p_1 = 43\%, p_2 = 40\%, p_3 = 12\%, p_4 = 5\%.$$

- En un grupo de 250 personas, observamos $f_1 = 110$ "O", $f_2 = 100$ "A", $f_3 = 30$ "B", $f_4 = 10$ "AB".
- En el grupo de 250, esperamos $\hat{f}_1 = 250 \times 43\% \simeq 107.5$ "O", $\hat{f}_2 = 250 \times 40\% \simeq 100$ "A", $\hat{f}_3 = 250 \times 12\% \simeq 30$ "B", $\hat{f}_4 = 250 \times 5\% \simeq 12.5$ "AB".
- Escribimos los datos en una tabla

Clases	A_1	A_2	A_3	A_4
Frecuencia observada	$f_1 = 110$	$f_2 = 100$	$f_3 = 30$	$f_4 = 10$
Frecuencia esperada	$\hat{f}_1 = 107.5$	$\hat{f}_2 = 100$	$\hat{f}_3 = 30$	$\hat{f}_4 = 12.5$

Construimos el estadístico para el contraste:

El estadístico chi-cuadrado de Pearson

Para contrastar

$$H_0 : P(A_i) = p_i, i = 1, \dots, k$$

$$H_1 : \text{Existe } i \text{ tal que } P(A_i) \neq p_i.$$

Karl Pearson propuso el estadístico

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}.$$

Si n es suficientemente grande,

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} \text{ es aprox. } \chi_{k-1}^2.$$

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i} = \frac{(100 - 107.5)^2}{107.5} + \frac{0^2}{100} + \frac{0^2}{30} + \frac{(10 - 12.5)^2}{12.5} = 1.02$$

Para calcular el p-valor, lo comparamos con los cuantiles de una χ_3^2 , valores extremos positivos nos llevan a rechazar H_0 .

p-valor:

Encontramos $p\text{-valor} = 0.79$, lo que nos lleva a afirmar que los datos apoyan la hipótesis nula, y son compatibles con los valores esperados $p_1 = 43\%$ "O", $p_2 = 40\%$ "A", $p_3 = 12\%$ "B", $p_4 = 5\%$ "AB".

Validez de la aproximación de la distribución de estadístico de Pearson por un χ^2_{k-1} .

n debe ser "suficientemente" grande.

- Se suele considerar que todas las frecuencias esperadas deben ser al menos 5, excepto una, y que esa sea mayor que 0.5.
- Si dos de ellas son pequeñas, entonces esas no deben ser menor que 1, y el resto debe ser al menos 5.
- Si no se cumple, se agrupan clases.

Segundo caso: debemos estimar parámetros para calcular las frecuencias esperadas.

Segundo ejemplo, un caso común:

- Tenemos k clases posibles asociadas a un experimento aleatorio A_1, \dots, A_k .
- Los valores esperados para las probabilidades $p_1 = P(A_1), \dots, p_k = P(A_k)$ dependen de unos parámetros desconocidos.
- Observamos n realizaciones del experimento y queremos decidir si lo observado es compatible con los valores esperados.
- Planteamos $H_0 : P(A_i) = p_i, i = 1, \dots, k$ frente a H_1 : Existe i tal que $P(A_i) \neq p_i$.

Procederemos igual que antes, pero:

- Estimaremos los parámetros desconocidos a partir de las observaciones.
- El estadístico de Pearson tendrá como distribución aproximada χ^2_{k-d-1} , donde d es el número de parámetros que debemos estimar para aproximar las frecuencias esperadas.

Ejemplo: daltonismo

La incidencia del daltonismo depende del sexo:

- Los genes responsables del daltonismo se encuentran en el cromosoma X .
- Los hombres tienen un solo cromosoma $X \implies$ una copia del gen defectuoso en su cromosoma X es suficiente para ser daltónico.
- Las mujeres tienen dos cromosomas $X \implies$ dos copias del gen defectuoso son necesarias para ser daltónicas.

Si llamamos q a la probabilidad de que el cromosoma X tenga el gen defectuoso y $p = 1 - q$.

Tenemos:

	Masculino	Femenino
Normales	$\frac{p}{2}$	$\frac{p^2}{2} + pq$
Daltónicos	$\frac{q}{2}$	$\frac{q^2}{2}$

Datos observados:

Para una muestra de 1000 personas:

	Masculino	Feminino
Normales	442	514
Daltónicos	38	6

Hoja de ejercicios Tema 4: Contrastes de hipótesis

- 1) Sea $X \sim \mathcal{N}(\mu, \sigma^2)$ y se considera una m.a.s. de tamaño n de X , donde σ^2 es conocida.
 - a) Consideremos las hipótesis $H_0 : \mu = \mu_0$ y $H_1 : \mu > \mu_0$. Demostrar que el test con región de rechazo $S_1 = \{\mathbf{x} \in \mathbf{X} | (\bar{x} - \mu_0)/(\sigma/\sqrt{n}) > z_{1-\alpha}\}$ es de extensión α para las hipótesis anteriores.
 - b) Consideremos las hipótesis $H_0 : \mu \leq \mu_0$ y $H_1 : \mu > \mu_0$. Demostrar que el test con región de rechazo $S_1 = \{\mathbf{x} \in \mathbf{X} | (\bar{x} - \mu_0)/(\sigma/\sqrt{n}) > z_{1-\alpha}\}$ es de extensión α para las hipótesis anteriores.
- 2) Sean $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ independientes. Se considera una m.a.s. de tamaño n_X de X y una m.a.s. de tamaño n_Y de Y .
 - a) Suponiendo σ_X^2 y σ_Y^2 conocidas, obtener un procedimiento para contrastar $H_0 : \mu_X = \mu_Y$ frente a $H_1 : \mu_X \neq \mu_Y$.
 - b) Suponiendo σ_X^2 y σ_Y^2 desconocidas pero iguales, obtener un procedimiento para contrastar $H_0 : \mu_X = \mu_Y$ frente a $H_1 : \mu_X \neq \mu_Y$.
- 3) Consideremos una m.a.s. X_1, \dots, X_n de una población que tiene una densidad $f(x; \theta) = e^{-(x-\theta)} \chi_{(\theta, +\infty)}$. Obtener el test RVG para $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$.
- 4) Sean $X \sim \mathcal{N}(\mu, \sigma^2)$ y se considera una m.a.s. de tamaño n de X . Consideramos las hipótesis $H_0 : \mu = \mu_0$ frente a $H_1 : \mu > \mu_0$. Demostrar que el test que rechaza H_0 si $\frac{\bar{X} - \mu_0}{S/\sqrt{n}} > t_{n-1, 1-\alpha}$ es de extensión α .
- 5) A continuación indicamos para distintas situaciones las hipótesis sobre la media poblacional de una variable normal para σ conocida que se plantearon, así como el valor del estadístico de prueba $z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$, que se obtuvo. Determinar los p-valores correspondientes e indicar cuál es la decisión acerca de H_0 .
 - a) $H_0 : \mu = 120; H_1 : \mu \neq 120$, estadístico de prueba: $z_0 = 2.32$.
 - b) $H_0 : \mu = 120; H_1 : \mu \neq 120$, estadístico de prueba: $z_0 = -1.88$.
 - c) $H_0 : \mu = 1000; H_1 : \mu > 1000$, estadístico de prueba: $z_0 = 1.6$.
 - d) $H_0 : \mu = 1000; H_1 : \mu < 1000$, estadístico de prueba: $z_0 = -1.24$.
- 6) En un laboratorio, se estudia la velocidad de combustión de dos tipos de combustibles sólidos A y B. Se pretende determinar qué combustibles presentan una velocidad de combustión en promedio mayor de 50 cm/s. Los resultados obtenidos fueron:

Combustible	Tamaño muestral	Media muestral	Varianza muestral
A	25	51.3	3.9
B	30	51.5	3.6

Suponiendo que la distribución de las variables es normal, plantear un test para contrastar si la varianza en la población de la velocidad de combustión para A es superior a la de B.

- 7) Sea X una población con función de densidad gamma:

$$f(x; \theta) = \theta^2 x \exp(-\theta x),$$

con $x > 0, \theta > 0$. Se considera una muestra de tamaño n de X .

- a) Estudiar el modelo anterior tiene cociente de verosimilitudes monótono en algún estadístico.
- b) En caso afirmativo, obtener el test uniforme de máxima potencia y extensión α para el contraste $H_0 : \theta = \theta_0$ frente a $H_1 : \theta > \theta_0$.
- c) Si $n = 20, \alpha = 0.05, \theta_0 = 1$ y $\bar{X} = 1.2$, ¿qué criterio de decisión de tomaría? Obtener el p-valor para dicha muestra.

- d) Obtener el test de razón de verosimilitudes generalizado y extensión α , para el contraste $H_0 : \theta = 2$ frente a la hipótesis $H_1 : \theta \neq 2$.

Tema 5: Inferencia Bayesiana

5.1) Introducción

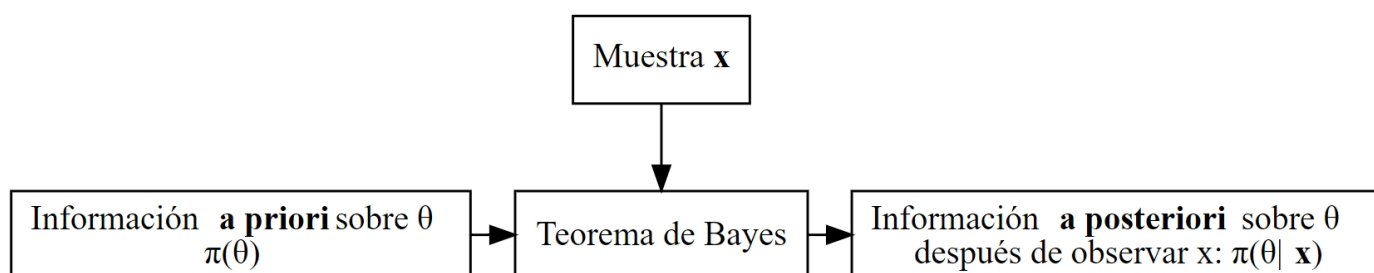
Lo que hemos visto hasta el momento:

- Se llama el enfoque frecuentista, o clásico de la inferencia.
- Hay otro enfoque, llamado la **inferencia bayesiana**.

La inferencia bayesiana

- Asume que modelizamos la información que podemos tener sobre el parámetro a través de una distribución de probabilidad.
- Formulamos como parte del modelo una **distribución a priori** sobre θ . Su f.p.p o función de densidad se denota por $\pi(\theta)$.
- La inferencia bayesiana actualiza la información a priori, usando para ello la muestra \mathbf{x} observada, para obtener la **distribución a posteriori** $\pi(\theta|\mathbf{x})$.

5.2) La inferencia bayesiana



Fórmula de Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Para distribuciones condicionadas:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{m(\mathbf{x})},$$

- $m(\mathbf{x})$ es la marginal de \mathbf{x} .
- Si π es una distribución continua: $m(\mathbf{x}) = \int f(\mathbf{x}|u)\pi(u) du$, la densidad marginal de \mathbf{x} calculada a partir de la conjunta.
- Esta integral se realiza sobre el conjunto de valores posibles de θ .

5.3) La distribución a posteriori

En la expresión

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{m(\mathbf{x})},$$

\mathbf{x} es el input, una constante. Nos interesa la variable θ , y el denominador solo depende de \mathbf{x} . Escribimos

$$\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta),$$

donde \propto quiere decir "es proporcional a".

Consejo

A veces, será suficiente reconocer el modelo de distribución que sigue $\pi(\theta|\mathbf{x})$ sin tener que calcular el denominador.

Ejemplo

Ingredientes

Lanzamos un dado n veces y apuntamos en cada tirada si ha salido par o impar.

- $X \sim \text{Bernoulli}(p)$, $X = 1$ si el resultado es par.
- muestra: (x_1, \dots, x_n) , tenemos que

$$f(x_1, \dots, x_n|p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}.$$

- Información a priori sobre p : puede tomar cualquier valor en $(0, 1)$ con igual probabilidad:

$$\pi(p) = \begin{cases} 1 & \text{si } p \in (0, 1) \\ 0 & \text{en otro caso} \end{cases}$$

$$\pi(p|\mathbf{x}) \propto p^{n\bar{x}}(1-p)^{n(1-\bar{x})}.$$

Reconocemos una distribución beta

Sean dos parámetros positivos a y b , $X \sim \text{Be}(a, b)$, si

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \text{ para todo } x \in (0, 1),$$

siendo Γ la función gamma.

- Su esperanza: $E[X] = \frac{a}{a+b}$
- Su varianza: $\text{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}$.
- Incluye la distribución uniforme en $(0, 1)$, si $a = 1$ y $b = 1$.
- En [R](#), [dbeta](#), [pbeta](#), [qbeta](#).

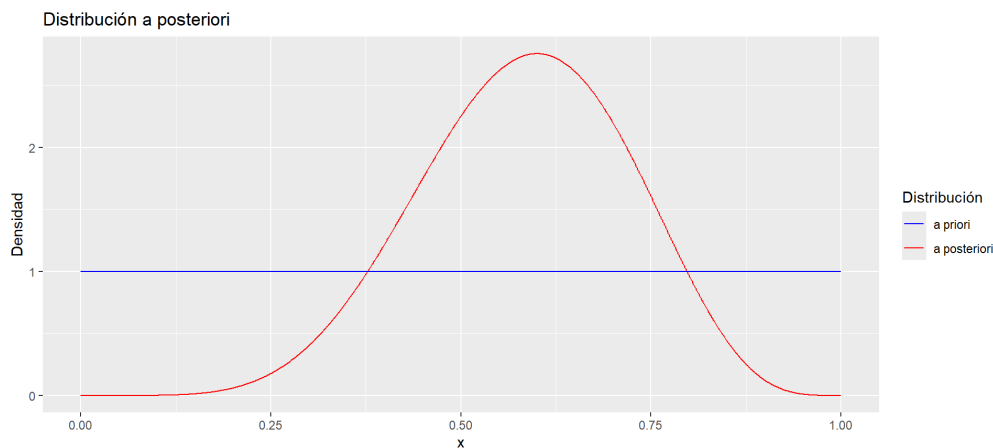
Tiramos 10 veces el dado, obtenemos 6 veces un número par

$$\pi(p|\mathbf{x}) \propto p^{n\bar{x}}(1-p)^{n(1-\bar{x})}.$$

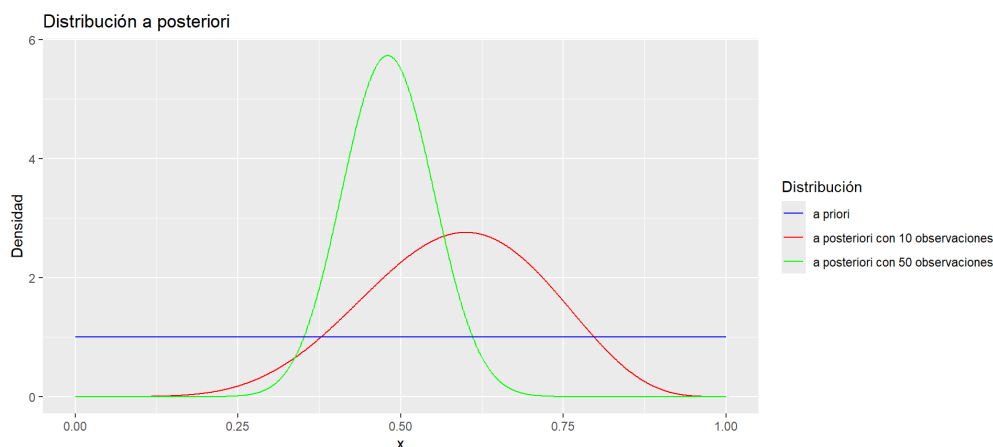
$$\pi(p|\mathbf{x}) \sim \text{Be}(1+n\bar{x}, 1+n(1-\bar{x}))$$

¿Qué vale $n\bar{x}$?

$$\pi(p|\mathbf{x}) \sim \text{Be}(7, 5)$$



Seguimos tirando el dado, hasta 50 veces, obtenemos otras 18 veces un número par, en total 24 veces: $\pi(p) = 1, \pi(p|\mathbf{x}) \sim \text{Be}(25, 27)$



5.4) Distribución a posteriori

Con el enfoque bayesiano, el resultado de la inferencia, una vez observada una muestra, es una distribución entera, la distribución a posteriori.

- Podemos representarla, contiene toda la información que hemos inferido sobre θ , a partir de la muestra y nuestra especificación de la distribución a priori.
- Podemos resumirla, por ejemplo con la media, la mediana.
- Podemos incluso dar un intervalo que contiene el 95% de los valores de θ . ("intervalo de credibilidad al 95%").

Estimador Bayes

Si consideramos una función de pérdida $L(\theta, a)$ asociada a estimar θ por a , la pérdida esperada a partir de la distribución a posteriori sería:

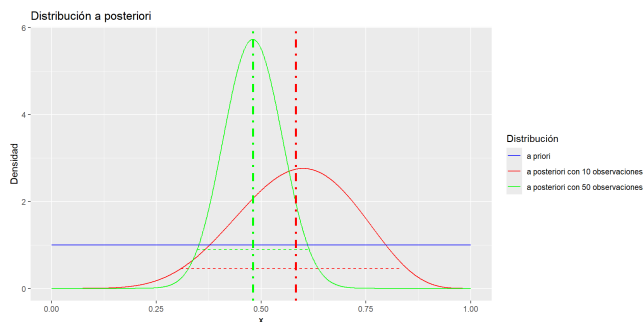
$$E[L(\theta, a)]|\mathbf{x} = \int_{\Omega} L(\theta, a) \pi(\theta|\mathbf{x}) d\theta$$

para cada valor posible \mathbf{x} del vector aleatorio \mathbf{X} , sea $\delta^*(\mathbf{x})$ un valor de a cuya pérdida esperada, dada por la expresión anterior sea mínima. La función $\delta^*(\mathbf{X})$ define un estimado llamado estimador Bayes:

$$\delta^*(\mathbf{X}) = \min_{a \in \Theta} E[L(\theta, a)|\mathbf{x}]$$

- Si $L(\theta, a) = (\theta - a)^2 \rightarrow \delta^*(\mathbf{X}) = E(\theta|\mathbf{X})$.
- Si $L(\theta, a) = |\theta - a| \rightarrow \delta^*(\mathbf{X}) = \text{mediana}$.

En el ejemplo de las tiradas de dados



Cuando usamos las 10 primeras observaciones

- $E[p|\mathbf{x}] = \frac{7}{12} = 0.58$
- Intervalo credibilidad al 95%:
 $(\text{qbeta}(0.025, 7, 5), \text{qbeta}(0.975, 7, 5)) = (0.31, 0.83)$

Cuando usamos las 50 observaciones

- $E[p|\mathbf{x}] = \frac{25}{52} = 0.48$
- Intervalo credibilidad al 95%:
 $(\text{qbeta}(0.025, 25, 27), \text{qbeta}(0.975, 25, 27)) = (0.35, 0.62)$

5.5) Cambiamos la distribución a priori

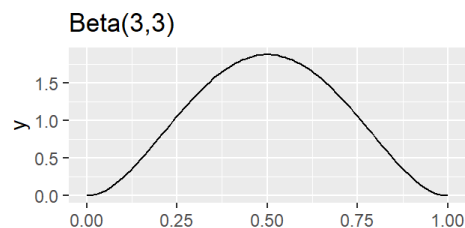
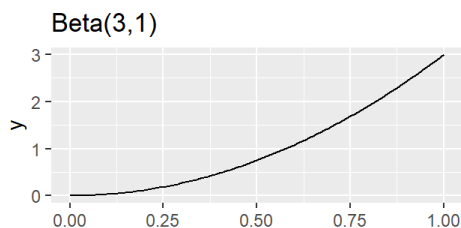
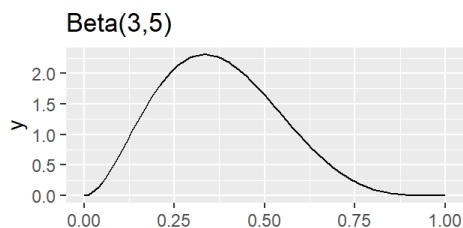
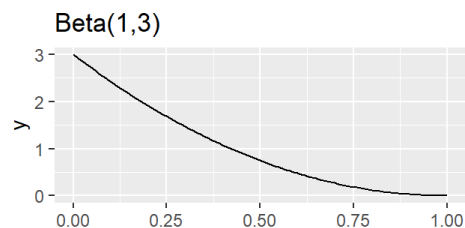
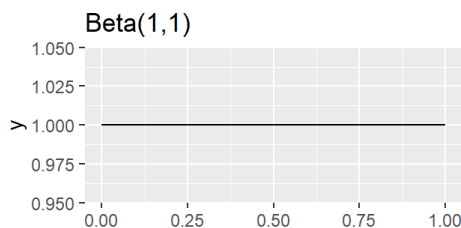
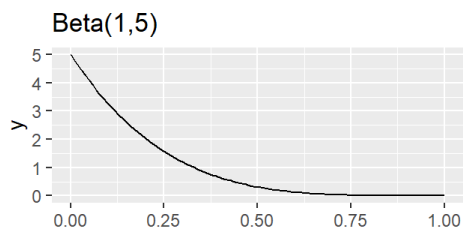
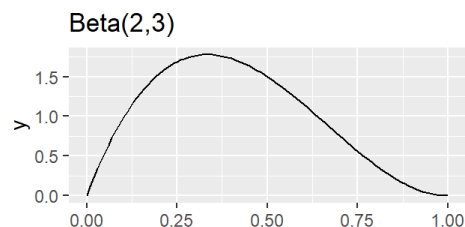
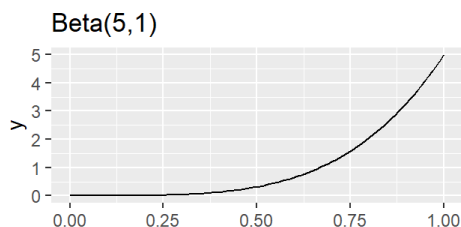
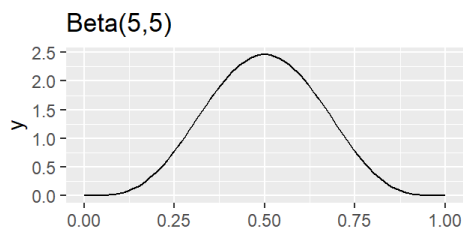
Si $\pi(p) \sim \text{Be}(a_0, b_0)$, y $X|p \sim \text{Bernoulli}(p)$ entonces

$$\pi(p|\mathbf{x}) \sim \text{Be}(a' + n\bar{x}, b_0 + n(1 - \bar{x})).$$

Clases conjugadas

Decimos que las clase Bernoulli (para X) y Beta (para la priori) son **conjugadas**: la posteriori se queda en la misma clase de distribuciones que la a priori.

5.5.1) La distribución $\text{Be}(a, b)$ es bastante flexible como a priori



5.6) Aplicaciones de la inferencia Bayesiana

La inferencia bayesiana se utiliza en varias áreas del aprendizaje máquina

- **Redes Bayesianas** (Belief Networks): Modelos gráficos que permiten representar las dependencias condicionales entre variables aleatorias de forma compacta.
- **Optimización Bayesiana:** Se utiliza para optimizar funciones objetivas que son costosas de evaluar, útil en el ajuste de hiperparámetros para modelos de aprendizaje máquina.
- **Procesamiento de Lenguaje Natural** (PLN): por ejemplo para la clasificación de texto y el análisis de sentimientos.
- **Sistemas de Recomendación:** Los modelos bayesianos se pueden utilizar para predecir las preferencias de los usuarios y hacer recomendaciones.

5.6.1) Ilustración para un ejemplo simple de sistema de recomendación:

Contexto:

- Queremos predecir la probabilidad de que un usuario aprecie una nueva película, para incluirla en las recomendaciones. Sabemos que, entre las 10 películas de este mismo tipo que ha visto, siete le han gustado.
- $X \sim \text{Bernoulli}(p) : X = 1 \longleftrightarrow$ "el usuario aprecia una película de este tipo".
- Observamos x_1, x_2, \dots, x_{10} .
- Usamos una a priori no informativa para $p : \pi(p) = \begin{cases} 1 & \text{si } p \in (0, 1) \\ 0 & \text{en caso contrario} \end{cases}$

$$\pi(p|\mathbf{x}) \sim \text{Be}(1 + n\bar{x}, 1 + n(1 - \bar{x})) \sim \text{Be}(8, 4).$$

Inferencia bayesiana predictiva:

- Tenemos la posteriori $\pi(p|\mathbf{x})$.
- Nos interesa $P(X = 1|\mathbf{x})$, que es la llamada probabilidad predictiva.

Para calcularla la obtenemos como marginal usando la posteriori:

$$P(X = 1|\mathbf{x}) = \int P(X = 1|p)\pi(p|\mathbf{x})dp = \int p\pi(p|\mathbf{x})dp.$$

En este caso: $P(X = 1|\mathbf{x}) = E[p|\mathbf{x}] = \frac{8}{12} = \frac{2}{3}$.

Hoja de ejercicios Tema 5: Inferencia Bayesiana

- 1) Consideramos el caso en que $X \sim P(\lambda)$ y $\pi(\lambda)$ sigue una distribución gamma con densidad

$$\pi(\lambda) \propto \lambda^{\alpha-1} \exp\left(-\frac{\lambda}{\beta}\right)$$

Observamos una m.a.s de X , estudiar la distribución a posteriori de λ .

- 2) Consideramos una m.a.s. de $X \sim \mathcal{N}(\mu, \sigma^2)$, con σ^2 conocida y escogemos como distribución a priori $\pi(\mu)$, una distribución normal con densidad

$$\pi(\mu) \propto \exp\left(-\frac{1}{2\tau^2}(\mu - \mu_0)^2\right)$$

Demostrar que la distribución a posteriori de μ es Normal con media y varianza:

$$E[\mu|x] = \mu_0 \frac{\frac{\sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}} + \bar{x} \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}}$$
$$\text{Var}[\mu|x] = \kappa^2 = \frac{\frac{\tau^2 \sigma^2}{n}}{\tau^2 + \frac{\sigma^2}{n}}$$

- 3) Consideramos una m.a.s. de $X \sim \text{Exp}(1/\kappa)$. Nota: κ es la esperanza de X . Consideramos a priori $\pi(\kappa) \propto \frac{1}{\kappa}$, que se considera una a priori no informativa y es la usual en este caso.

- a) Demostrar que la densidad a posteriori de $\kappa|\mathbf{x}$ es proporcional $\kappa^{-n-1} \exp\left(-\sum_i \frac{x_i}{\kappa}\right)$

- b) Demostrar que, si transformamos κ , considerando $\lambda = \frac{1}{\kappa}$, $\lambda|\mathbf{x}$ admite una densidad gamma con parámetros n y $\frac{1}{\sum_i x_i}$.

- c) En un experimento sobre duración de dispositivos, obtenemos los siguientes valores en horas: 6562, 2280, 50, 989, 1797. Obtener un intervalo de credibilidad al 90% para la inversa de k . Obtener el mismo intervalo para κ .