Estadística I

Guillermo Julián Moreno
Eduardo Miravalls Sierra

UAM - 13/14 C1
19 de enero de 2016 01:02

Apuntes UAM
Doble Grado Mat.Inf.

Código en Github

Índice general

I Estadística descriptiva **3**
I.1 Estadística descriptiva de datos univariantes . . . 3
I.1.1 Estadísticos de tendencia central . . . 3
I.1.2 Estadísticos de dispersión . . . 3
I.1.3 Representación gráfica de datos . . . 4
I.2 Estadística descriptiva de datos bivariantes . . . 8
I.2.1 Representación gráfica . . . 8
I.2.2 Regresión . . . 9

II Muestreo aleatorio **11**
II.1 Conceptos de probabilidad . . . 11
II.1.1 Distribuciones aleatorias . . . 12
II.2 Problema de inferencia . . . 15
II.2.1 Interpretación estadística de la ley de los grandes números . . . 15
II.2.2 Función de distribución empírica . . . 15
II.3 Estadísticos . . . 18
II.3.1 Media muestral y poblacional . . . 18
II.3.2 Varianza muestral y poblacional . . . 20
II.3.3 Estadísticos de orden . . . 20

III Estimación paramétrica **22**
III.1 Estimadores . . . 22
III.1.1 Propiedades interesantes de los estimadores . . . 23
III.1.2 Estimador de máxima verosimilitud (EMV) . . . 26
III.1.3 Método de los momentos . . . 42
III.1.4 Metodología bayesiana . . . 43
III.2 Estimación por intervalos de confianza . . . 46
III.2.1 Intervalos de confianza asintóticos basados en el TCL . . . 47
III.2.2 Método de la cantidad pivotal . . . 48
III.2.3 Construcción de intervalos de confianza habituales . . . 48
III.2.4 Intervalos de confianza bayesianos . . . 49

IV Contraste de hipótesis **50**
IV.1 Conceptos básicos . . . 50
IV.1.1 Teoría de Neyman-Pearson . . . 51
IV.2 Problema de una muestra . . . 52
IV.2.1 Regiones de rechazo para contrastes habituales . . . 53
IV.3 Contrastes para dos muestras . . . 54

[0] Documento compilado el 19 de enero de 2016 a las 01:02

IV.4 Consistencia de tests. Tests insesgados y UMP . . . 57
IV.4.1 Lema de Neyman-Pearson . . . 57
IV.4.2 Familias paramétricas con cociente de verosimilitudes monótono y tests óptimos . . . 59
IV.4.3 Construcción de tests. Test de cociente de verosimilitudes . . . 60
IV.4.4 Tests Bayesianos . . . 64

A Anexos **65**
A.1 Condiciones suficientes para permutar la derivada con la integral . . . 65
A.2 Distribuciones notables . . . 66
A.3 Regiones de rechazo . . . 72

B Ejercicios **75**
B.1 Tema 1 - Estadística descriptiva . . . 76
B.2 Tema 2 - Muestreo aleatorio . . . 91
B.3 Tema 3 - Estimación puntual paramétrica . . . 99
B.4 Tema 4 - Intervalos de confianza . . . 110
B.5 Tema 5 - Contraste de hipótesis . . . 120
B.5.1 Hoja 5A . . . 120
B.5.2 Hoja 5B . . . 133

C Exámenes **141**
C.1 Enero 2013 . . . 142
C.1.1 Solución . . . 144
C.2 Junio 2013 . . . 148
C.2.1 Solución . . . 150
C.3 Enero 2014 . . . 154
C.3.1 Solución . . . 156

Índice alfabético **158**

Capítulo I

Estadística descriptiva

I.1. Estadística descriptiva de datos univariantes

La estadística descriptiva es el conjunto de técnicas para resumir la información proporcionada por una gran masa de datos. El primer objetivo natural es resumir la información que proporcionan esos datos.

I.1.1. Estadísticos de tendencia central

Media

Definición I.1 **Media**.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Es la medida de tendencia central más utilizada. Es bastante sensible a los valores atípicos (outliers), observaciones anormalmente grandes que aparecen en el conjunto de datos por errores de transcripción o medición.

Mediana

Definición I.2 **Mediana**. *Es el valor que divide a los datos en dos mitades, de tal forma que la mitad son menores y la otra mitad mayores que la mediana.*

La mediana se calcula de la siguiente forma: dado un conjunto de datos $\{x_1, \dots, x_n\}$, *la mediana es* $x_{\frac{n+1}{2}}$ *si* n *es impar y el promedio entre* $x_{\frac{n}{2}}$ *y* $x_{\frac{n}{2}+1}$ *si* n *es par.*

I.1.2. Estadísticos de dispersión

Varianza

Definición I.3 **Varianza**.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2$$

Desviación típica

Definición I.4 **Desviación típica.**

$$\sigma = \sqrt{\sigma^2}$$

La desviación típica es la raíz de la varianza.

Cuantil

Definición I.5 **Cuantil**. *Para $p \in (0,1)$ se llama cuantil p o q_p al valor que deja el $100p\,\%$ de los datos a la izquierda.*

Cuartil

Definición I.6 **Cuartil**. *Los cuartiles son los tres datos que dejan a la izquierda el 25, 50 y 75 por ciento de los datos respectivamente. Es decir:*

- $Q_1 = q_{0.25}$
- $Q_2 = q_{0.5}$. *El cuartil dos es la mediana.*
- $Q_3 = q_{0.75}$

Hay varios métodos para el cálculo de cuantiles. Para hacerlo a mano, podemos usar el siguiente método.

Si el dato en la posición $p(n+1)$ no es un número entero, entonces se interpola entre las observaciones ordenadas que están en la posición $\lfloor p(n+1) \rfloor$ y $\lfloor p(n+1) \rfloor + 1$ de la siguiente forma: sea j la parte entera de $p(n+1)$ y m la parte decimal. Entonces,

$$q_p = (1-m)x_j + mx_{j+1}$$

Coeficiente de asimetría

Definición I.7 **Coeficiente de asimetría**. *El tercer momento con respecto a la media se define como*

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3$$

que, en su versión adimensional dividimos por σ^3.

Un valor diferente de 0 indica asimetría de las muestras. Sin embargo, 0 no garantiza simetría, solo que ambas colas se compensan.

I.1.3. Representación gráfica de datos

Box-plot

Definición I.8 **Box-plot**. *El diagrama de caja o box-plot (imagen I.1) nos permite visualizar las medidas de dispersión respecto a la mediana. Hay que añadir una nueva medida, el* ***rango intercuartílico****, la diferencia entre el primer y el tercer cuartil:*

$$RI = Q_3 - Q_1$$

A partir del rango intercuartílico obtenemos los límites inferior y superior de la representación:

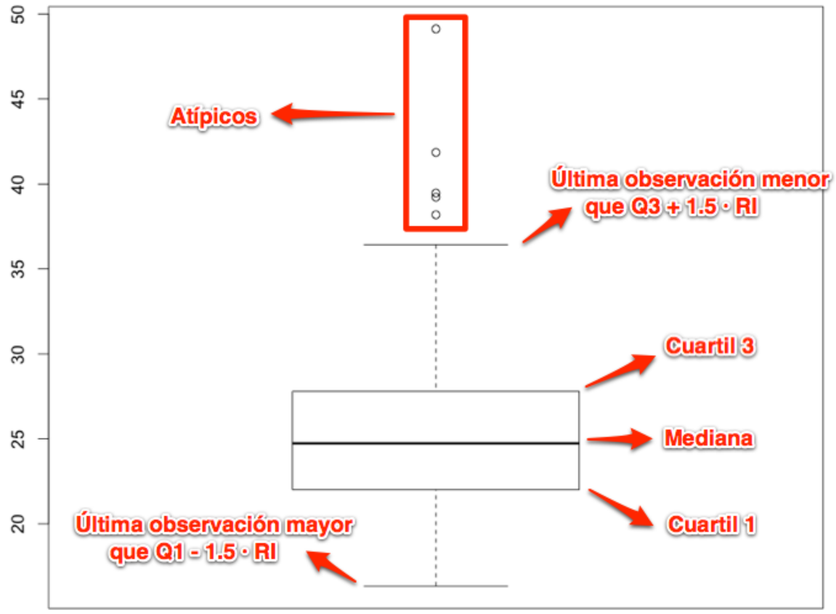


Figura I.1: *Diagrama de caja*

Límite inferior/-superior

Definición I.9 **Límite inferior/superior**. *Se define el límite superior (LS) y el inferior (LI) de la siguiente forma:*

$$LS = Q_3 + 1.5RI$$
$$LI = Q_1 - 1.5RI$$

Cualquier dato fuera del intervalo $[LI, LS]$ se considera un atípico.

Histograma

Definición I.10 **Histograma**. *El histograma se trata de una aproximación discreta a la función de densidad continua $f(t)$ de la variable que estamos midiendo. Es un diagrama de frecuencias que mantiene la forma de esa función de densidad.*

Definimos una serie, las marcas de intervalos $a_1^n, \dots, a_n^n$, donde n es el número de intervalos y la longitud de cada intervalo es $h_n = a_{j+1}^n - a_j^n$. Sea el conjunto $\{x_i\}_{i=0,\dots,m}$ los datos de nuestra muestra. Entonces, el estimador, la función $\hat{f}_n$, se define de la siguiente forma:

$$\hat{f}^n(t) = \frac{\left|\left\{i \;/\; x_i \in \left(a_j^n, a_{j+1}^n\right]\right\}\right|}{nh_n} = \frac{\sum_{i=1}^{m} \mathbb{1}_{(a_j^n, a_{j+1}^n]}(x_i)}{nh_n}$$

*Recordemos la **función indicatriz***

$$\mathbb{1}_A(n) = \begin{cases} 1 & n \in A \\ 0 & n \notin A \end{cases}$$

A grandes rasgos, lo que hace en una función es definir un número de intervalos fijos de ancho h_n. Al evaluar $\hat{f}^n(t)$ buscamos en qué intervalo cae t y contamos cuántas de nuestras mediciones están también en ese intervalo.

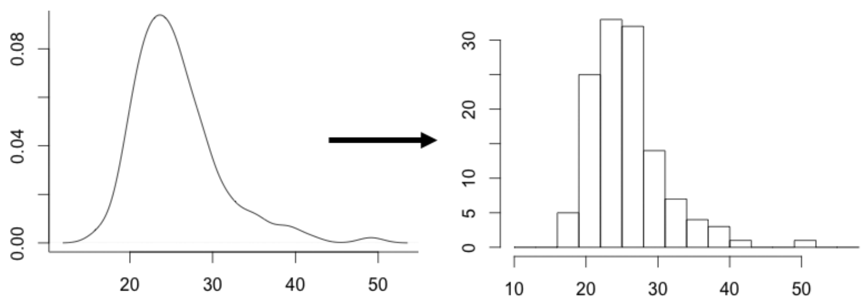


Figura I.2: *El histograma es una aproximación de la función de densidad real en base a la muestra que hemos obtenido.*

I.1.3.1. Estimadores núcleo o kernel

Método de ventana móvil

Definición I.11 **Método de ventana móvil**. *El método de ventana móvil nos da una estimación de la función de densidad en un punto t midiendo los x_i que están en el intervalo de radio h_n centrado en t. Matemáticamente:*

$$\hat{f}_n(t) = \frac{1}{n2h_n} \sum_{i=1}^{n} \mathbb{1}_{[t-h_n, t+h_n]}(x_i) = \frac{1}{n2h_n} \sum_{i=1}^{n} \mathbb{1}_{[-1,1]} \left(\frac{t - x_i}{h_n} \right)$$

Podemos reemplazar la función $\frac{1}{2}\mathbb{1}_{[-1,1]}$ por otra, llamada la función de densidad K, kernel o núcleo:

Estimador núcleo

Definición I.12 **Estimador núcleo**. *Dada una función de densidad K simétrica, no necesariamente positiva, definimos el estimador kernel como:*

$$\hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^{n} K_h(t - x_i) = \frac{1}{nh_n} \sum_{i=1}^{n} K \left(\frac{t - x_i}{h_n} \right)$$

con $K_h(x) = \frac{1}{h}K(\frac{x}{h})$.

La elección del núcleo K no afecta especialmente a lo bien aproximada que esté la función de densidad. Sin embargo, sí que influye la selección de la ventana h_n (figura

I.3), también llamada *bandwith* en inglés. Si escogemos una ventana muy pequeña, damos demasiado peso a los datos de nuestra muestra. Si elegimos una ventana muy grande, nuestra muestra pierde importancia y podemos perder información importante.

La elección del h_n más habitual es el que minimiza la distancia L^2 entre $\hat{f}$ y f, es decir, el parámetro que minimice $\int \left(\hat{f}_h - f\right)^2$. Sin embargo, hay un problema: no sabemos qué es f. Hay trucos que imagino que veremos más tarde.

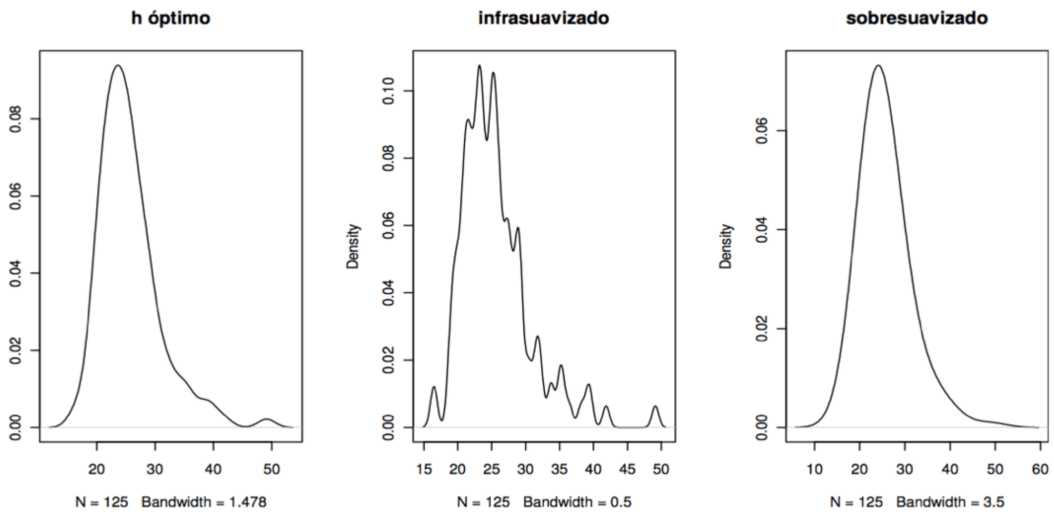


Figura I.3: *Los efectos que causa elegir una ventana más grande o más pequeña en el estimador*

Las funciones kernel más usadas son la uniforme, $\frac{1}{2}\mathbb{1}_{[-1,1]}$, la gaussiana $\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$ y la de Epanechnikov, que matemáticamente es la que mejor aproxima f.

El estimador kernel $\hat{f}_n(t)$ es la función de densidad de una medida de probabilidad que es la convolución [1] de dos medidas de probabilidad: una, $K_h(x)$ (el kernel reescalado) y otra que da probabilidad $\frac{1}{n}$ a cada punto de la muestra $\{x_i\}$ (distribución o medida empírica).

Generación de datos del estimador kernel Supongamos que K es el núcleo gaussiano. Podemos generar datos artificiales de la densidad así:

$$x_i^0 = x_i^* + h_n Z_i, \; i = 1, \ldots, k$$

donde x_i^* es una observación elegida al azar entre los datos originales y Z_i una observación aleatoria con probabilidad $N(0,1)$. Es decir, lo que hacemos es añadir un dato aleatorio de la muestra y sumamos una pequeña perturbación aleatoria.

[1] http://en.wikipedia.org/wiki/Convolution

I.2. Estadística descriptiva de datos bivariantes

En esta sección estudiaremos dos variables (X, Y) para explorar la relación entre ambas y tratar de inferir si existe una relación funcional para predecir los valores de una variable en función de los de la otra.

I.2.1. Representación gráfica

Diagrama de dispersión

Definición I.13 **Diagrama de dispersión**. *El diagrama de dispersión representa cada variable en función de la otra para que podamos ver la posible relación entre ambas. Ver figura I.4.*

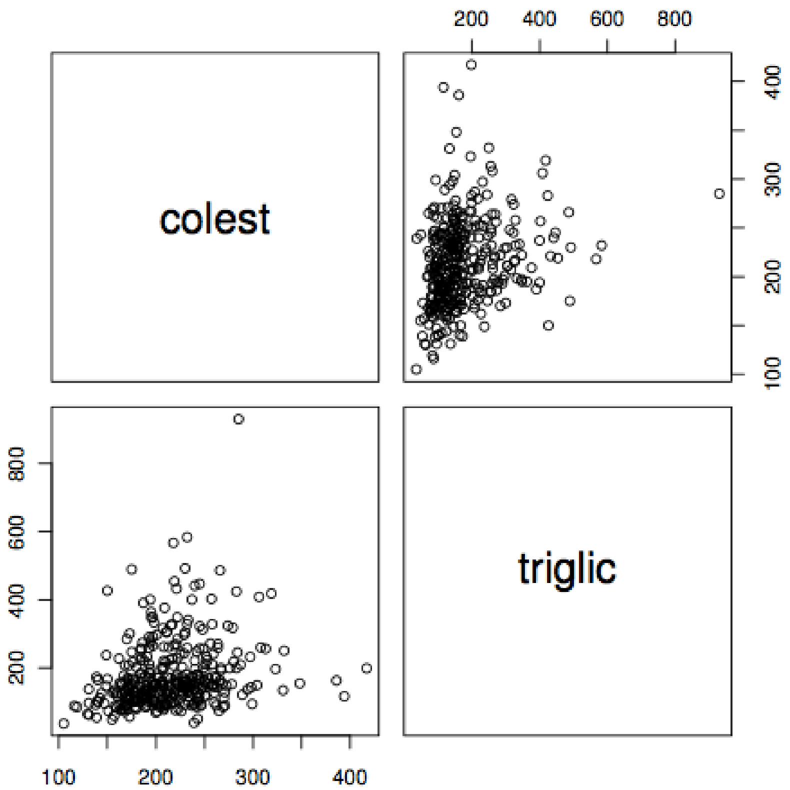


Figura I.4: *Diagrama de dispersión*

I.2.2. Regresión

Recta de regresión

Definición I.14 **Recta de regresión**.

La recta de regresión de y sobre x es la recta de forma $\hat{y} = \hat{a} + \hat{b}x$ que más se aproxima a los datos, minimizando los cuadrados de la distancia:

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

La recta de regresión se calcula obteniendo primero $\hat{b}$:

$$\hat{b} = \frac{\sigma_{x,y}}{\sigma_x^2}$$

donde $\sigma_{x,y}$ se denomina **covarianza muestral de x e y**:

$$\sigma_{x,y} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \overline{xy}$$

y después, sabiendo que la recta pasa por el punto $(\overline{x}, \overline{y})$, obtenemos $\hat{a}$

$$\hat{a} = \overline{y} - \hat{b}\overline{x}$$

El valor b se denomina **coeficiente de regresión lineal** o parámetro de la regresión. Cada valor $e_i = y_i - \hat{y}_i$ se denomina **residuo**. Hay que notar que

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}\left(y_i - \hat{a} - \hat{b}x_i\right) = \sum_{i=1}^{n}\left(y_i - (\overline{y} - \hat{b}\overline{x}) - \hat{b}x_i\right) =$$
$$= \sum_{i=1}^{n}\left(y_i - \hat{b}x_i\right) - n\overline{y} + n\hat{b}\overline{x} = n\overline{y} - n\hat{b}\overline{x} - n\overline{y} + n\hat{b}\overline{x} = 0$$

Esta ecuación ($\sum_{i=1}^{n} e_i = 0$) junto con

$$\sum_{i=1}^{n} x_i e_1 = 0$$

son las dos restricciones entre los residuos que nos dan la recta.

Varianza residual

Definición I.15 **Varianza residual**. *La varianza residual s_R^2 o $\hat{\sigma}_e^2$ mide, aproximadamente el error cuadrático cometido en la aproximación dada por la recta de regresión:*

$$s_R^2 = \hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Coeficiente de correlación lineal

Definición I.16 **Coeficiente de correlación lineal**. *El coeficiente de correlación lineal o coeficiente de Pearson*

$$r = \frac{\hat{\sigma}_{x,y}}{\hat{\sigma}_x \hat{\sigma}_y}$$

que cumple las siguientes condiciones:

$$0 \leq r^2 \leq 1$$
$$\hat{\sigma}_e^2 = \hat{\sigma}_y^2(1 - r^2)$$
$$r = \hat{b}\frac{\hat{\sigma}_x}{\hat{\sigma}_y}$$

nos indica el grado de ajuste lineal entre las dos variables. Un valor absoluto más cercano a 1 indica una correlación más fuerte. Un valor absoluto cercano a cero indica una correlación débil. El signo, positivo o negativo, indica si la correlación es creciente o decreciente.

Capítulo II

Muestreo aleatorio

La muestra aleatoria de una cierta v.a.[1] X se denomina como la **muestra aleatoria** o simplemente **muestra**.

Durante este tema, usaremos conceptos de Probabilidad, que repasaré aquí brevemente[2].

II.1. Conceptos de probabilidad

Distribución de una v.a.

Definición II.1 **Distribución de una v.a.**.

$$\mathbb{P}_X\{B\} = \mathbb{P}\{X \in B\}$$

Función de distribución

Definición II.2 **Función de distribución**.

$$F(t) = \mathbb{P}\{X \leq t\}$$

Media de una distribución

Definición II.3 **Media de una distribución**. *También llamada esperanza de X:*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} F(t)\, dt$$

[1] variable aleatoria

[2] repasa PROB I

Teorema II.1 (Teorema de cambio de espacio de integración). *Sea g una función real medible tal que $\mathbb{E}\left(g(X)\right)$ es finita, entonces*

$$\mathbb{E}\left(g(X)\right) = \int_{\mathbb{R}} g(x)\,\mathrm{d}F(x) = \int_{\mathbb{R}} g(x)\;\mathrm{d}P(x)$$

En particular

$$\mathbb{E}\left(X\right) := \mu = \int_{\mathbb{R}} x\;\mathrm{d}F(x)$$

y

$$\mathbb{V}\left(X\right) := \sigma^2 = \int_{\mathbb{R}} \left(x-\mu\right)^2 \mathrm{d}F(x)$$

Momento

Definición II.4 **Momento**. *El momento μ_k es la esperanza de X elevado a una potencia de orden k. Es el valor esperado de la distancia de orden k con respecto a la media*

$$\mu_k = \mathbb{E}\left(\left(X-\mu\right)^k\right)$$

II.1.1. Distribuciones aleatorias

Ver apéndice A.2 (página 66).

II.1.1.1. Criterios de convergencia

Queremos buscar convergencias entre variables aleatorias.

Convergencia en distribución

Definición II.5 **Convergencia en distribución**.

Se dice que X_n converge débilmente o en distribución a X si la función de distribución de X_n, $F_n(x)$, tiende a $F(x)$ para todo x punto de continuidad de F; donde F y F_n son las funciones de distribución de X y X_n respectivamente.

Esto es equivalente a decir que

$$\lim_{n\to\infty} \mathbb{P}\left\{X_n \in (-\infty, x]\right\} = \mathbb{P}\left\{X \in (-\infty, x]\right\}$$

Notación:

$$X_n \xrightarrow[n\to\infty]{d} X \text{ ó } X_n \xrightarrow[n\to\infty]{w} X$$

Convergencia en probabilidad

Definición II.6 **Convergencia en probabilidad**. *Se dice que X_n converge en probabilidad a X si $\forall\varepsilon > 0$ se tiene que*

$$\mathbb{P}\left\{|X_n - X| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

Es decir, que para cualquier error que tomemos el error cometido en la aproximación va a tender a cero siempre que tomemos un X_n suficientemente grande.

Notación:

$$X_n \xrightarrow[n\to\infty]{P} X$$

Convergencia casi segura

Definición II.7 **Convergencia casi segura**. *También denotada c.s o a.s en inglés, convergencia en casi todo punto (c.t.p) o convergencia con probabilidad 1.*
Se dice que X_n converge a X casi seguro si el conjunto de puntos que no son convergentes tiende a ser vacío. Es decir

$$\mathbb{P}\left\{X_n \xrightarrow[n\to\infty]{} X\right\} = 1$$

Otra forma de interpretarlo es: $X_n \xrightarrow[n\to\infty]{} X$ cuando el conjunto de los ω tales que $X(\omega)$ es el límite de la sucesión $X_n(\omega)$ tiene probabilidad 1.

Más estrictamente, la condición se expresa como

$$\mathbb{P}\left\{\omega \in \Omega \;/\; X_n(\omega) \xrightarrow[n\to\infty]{} X(\omega)\right\} = 1$$

Notación

$$X_n \xrightarrow[n\to\infty]{c.s} X$$

Teorema II.2. *Se puede probar que si $\{X_n\}$ es una sucesión de variables aleatorias y X es variable aleatoria,*

$$X_n \xrightarrow[n\to\infty]{c.s} X \implies X_n \xrightarrow[n\to\infty]{P} X \implies X_n \xrightarrow[n\to\infty]{d} X$$

La recíproca no es cierta.

Teorema II.3 (Teorema de Slutsky). *Sean* $\{X_n\}$, $\{Y_n\}$ *sucesiones de variables aleatorias tales que* $X_n \xrightarrow[n\to\infty]{d} X$, $Y_n \xrightarrow[n\to\infty]{P} c$ *con* $c \in \mathbb{R}$ *constante. Entonces*

1. $X_n + Y_n \xrightarrow[n\to\infty]{d} X + c$
2. $X_n \cdot Y_n \xrightarrow[n\to\infty]{d} X \cdot c$
3. $\frac{X_n}{Y_n} \xrightarrow[n\to\infty]{d} \frac{X}{c}$ *si* $c \neq 0$.

II.1.1.2. Desigualdades básicas

Teorema II.4 (Desigualdad de Markov). *Sea* X *v.a. Entonces,* $\forall \varepsilon > 0$,

$$\mathbb{P}\left\{|X| > \varepsilon\right\} \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

Teorema II.5 (Desigualdad de Chebichev). *Sea* X *v.a. Entonces,* $\forall \varepsilon > 0$, *se cumple que*

$$\mathbb{P}\left\{\left|X - \mathbb{E}(X)\right| > \varepsilon\right\} \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

II.2. Problema de inferencia

II.2.1. Interpretación estadística de la ley de los grandes números

Teorema II.6 (Ley de los grandes números). *Sea $\{x_k\}$ una sucesión de v.a.i.i.d. con media finita μ. Se verifica entonces que*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow[n\to\infty]{c.s} \mu$$

II.2.2. Función de distribución empírica

Función de distribución empírica

Definición II.8 **Función de distribución empírica**. *La función de distribución empírica asociada a la muestra $\{x_n\}$ se define mediante*

$$\mathbb{P}\{X \leq t\} = \mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(x_i)$$

Es decir, $\mathbb{F}_n(t)$ es la proporción de puntos de la muestra que caen en el intervalo $(-\infty, t]$.

Sin embargo, surge una duda: ¿converge la función de distribución empírica a la función de distribución original?

Intuitivamente, podemos pensar que cuantos más puntos cojamos, más se aproximará a la función de distribución original. De hecho, eso es lo que demuestra el siguiente teorema:

Teorema II.7 (Teorema de Glivenko-Cantelli). *Sean $\{x_n\}$ v.a.i.i.d. con función de distribución F. Se verifica que*

$$\|\mathbb{F}_n - F\|_\infty = \sup_{t\in\mathbb{R}} \left|\mathbb{F}_n(t) - F(t)\right| \xrightarrow[n\to\infty]{c.s} 0$$

donde $\|\mathbb{F}_n - F\|_\infty$ es el ***estadístico de Kolmogorov-Smirnov****.*

Demostración. Empezamos demostrando la convergencia de los términos intermedios. Es decir, queremos demostrar que

$$\mathbb{F}_n(t) \xrightarrow[n\to\infty]{c.s} F(t) \tag{II.1}$$

Tenemos que

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(x_i)$$

A cada uno de los términos de los términos de la suma $\mathbb{1}_{(-\infty,t]}(x_i)$ los podemos llamar y_i. Estos valores son una muestra de la distribución

$$Y = \mathbb{1}_{(-\infty,t]}(X)$$

Por lo tanto y por la LGN (II.6)

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y} \xrightarrow[n\to\infty]{c.s} \mathbb{E}\left(Y\right)$$

pero

$$\mathbb{E}\left(Y\right) = \mathbb{E}\left(\mathbb{1}_{(-\infty,t]}(X)\right) = \mathbb{P}\left\{X \in (-\infty,t]\right\} = F(t)$$

por lo tanto hemos demostrado (II.1).

Ahora tenemos que demostrar que el límite por la izquierda converge. Es decir, hay que demostrar que

$$\mathbb{F}_n(t^-) \xrightarrow[n\to\infty]{c.s} F(t^-) \tag{II.2}$$

. Esa convergencia se da si y sólo si en un conjunto de probabilidad 1 se tiene que $\mathbb{F}_n(t^-) \xrightarrow[n\to\infty]{} F(t^-)$. Según la definición de límite, esto se da si y sólo si

$$\forall\varepsilon > 0\ \exists N \;/\; n \geq N \implies \left|\mathbb{F}_n(t^-) - F(t^-)\right| < \varepsilon \tag{II.3}$$

Sabemos que

$$\exists\varepsilon > 0 \;/\; \mathbb{F}_n(t^-) = \mathbb{F}_n(x)\ \forall x \in (t-\delta, t+\delta) \tag{II.4}$$

Seguimos:

$$F(t^-) = \lim_{x\to t^-} F(x) \iff \forall\varepsilon > 0\ \exists\delta > 0 \;/\; x \in (t-\delta, t) \implies \left|F(x) - F(t^-)\right| < \frac{\varepsilon}{2} \tag{II.5}$$

Tomamos $x \in (t-\delta, t)$ con un delta que cumpla tanto la condición en (II.4) como en (II.5). Entonces

$$\left|\mathbb{F}_n(t^-) - F(t^-)\right| = \left|\mathbb{F}_n(x) - F(x) + F(x) - F(t^-)\right| \leq \underbrace{\left|\mathbb{F}_n(x) - F(x)\right|}_{(a)} + \underbrace{\left|F(x) - F(t^-)\right|}_{(b)}$$

Sabemos que (a) es menor que $\frac{\varepsilon}{2}$ por (II.3) y que (b) también es menor que $\frac{\varepsilon}{2}$ por (II.5), por lo tanto

$$\left|\mathbb{F}_n(t^-) - F(t^-)\right| < \varepsilon$$

Buscamos ahora una partición finita de $\mathbb{R}$ dada por $t_0 = -\infty \leq t_1 \leq \cdots \leq t_k = \infty$ tal que para todo $\varepsilon > 0$ se cumpla que $\left|F(t_i^-) - F(t_{i-1})\right| \leq \varepsilon$. Lo construimos de forma recursiva: dado t_{i-1} tomamos

$$t_i = \sup_{z\in\mathbb{R}}\{F(z) \leq F(t_{i-1} + \varepsilon\}$$

El siguiente paso: para todo $t_{i-1} \leq t \leq t_i$ se tiene que

$$\mathbb{F}_n(t) - F(t) \leq \mathbb{F}_n(t_i^-) - F(t_i^-) + \varepsilon$$

Como $\mathbb{F}_n$ es no decreciente (es una función de distribución), tenemos también que

$$\mathbb{F}_n(t) - F(t) \geq \mathbb{F}_n(t_{i-1}) - F(t_{i-1}) - \varepsilon$$

Con estas dos últimas ecuaciones, llegamos a que

$$\sup_{t\in\mathbb{R}}\left|\mathbb{F}_n(t) - F(t)\right| \leq \max\left\{\max_{i=1,\dots,k}\left|\mathbb{F}_n(t_i) - F(t_i)\right|, \max_{i=1,\dots,k}\left|\mathbb{F}_n(t_i^-) - F(t_i^n)\right|\right\} + \varepsilon$$

Por (II.1), sabemos que $\left|\mathbb{F}_n(t_i) - F(t_i)\right| \xrightarrow[n\to\infty]{c.s} 0$, y por lo tanto

$$\max_{i=1,\dots,k}\left|\mathbb{F}_n(t_i) - F(t_i)\right| \xrightarrow[n\to\infty]{c.s} 0$$

De la misma forma, usando (II.2) tenemos que

$$\max_{i=1,\dots,k}\left|\mathbb{F}_n(t_i^-) - F(t_i^n)\right| \xrightarrow[n\to\infty]{c.s} 0$$

Por lo tanto, todo ese máximo enorme vale 0, de tal forma que

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|\mathbb{F}_n(t) - F(t)\right| = \lim_{n\to\infty}\|\mathbb{F}_n - F\|_\infty \leq \varepsilon$$

para cualquier $\varepsilon > 0$ arbitrario que cojamos. Es decir, que

$$\|\mathbb{F}_n - F\|_\infty = \sup_{t\in\mathbb{R}}\left|\mathbb{F}_n(t) - F(t)\right| \xrightarrow[n\to\infty]{c.s} 0$$

□

II.3. Estadísticos

Cuando extraemos una muestra $\{x_n\}$ de X se pueden calcular algunas *medidas resumen*. Cualquiera de ellas se puede expresar matemáticamente como una función $T(x_1, \dots, x_n)$ de la muestra.

Estadístico

Definición II.9 **Estadístico**. *Sea $T(x_1, \dots, x_n)$ una función cuyo dominio incluye el espacio muestral del vector aleatorio $(X_1, \dots, X_n)$. Entonces la variable aleatoria T se denomina **estadístico**. La única restricción es que un estadístico no puede ser función de un parámetro.*

Como la distribución de T se calcula a partir de la distribución de las variables X_i que constituyen la muestra, la denominaremos distribución de T en el muestreo (*sampling distribution*).

Error típico

Definición II.10 **Error típico**. *El error estándar o error típico σ de un estadístico T, es la desviación típica de su distribución en el muestreo. Como en ocasiones depende de alguna cantidad desconocida, también se denomina error típico a una estimación de ese valor.*

En ocasiones, se cumple que $\frac{T}{\sigma}$ sigue una distribución t de Student, lo que nos permitirá definir intervalos de confianza.

II.3.1. Media muestral y poblacional

Media muestral

Definición II.11 **Media muestral**. *La media muestral*

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

se puede expresar de la siguiente forma

$$\overline{X} = \int_{\mathbb{R}} x \, d\mathbb{F}_n(x)$$

La definición es análoga con la de la **media poblacional**

$$\mu = \int_{\mathbb{R}} x \, dF(x)$$

Esto nos da una clave de la estadística: sustituir todo lo que desconozco de la población con su análogo muestral[3] (en este caso, pasamos de la función de distribución teórica a la función de distribución empírica). Sólo quedaría ver si los estimadores que resultan son adecuados.

[3] método plugin

La media muestral tiene otras relaciones muy importantes con μ:

1. $\overline{X}$ es **estimador insesgado o centrado** de μ: $\mathbb{E}\left(\overline{X}\right) = \mu$
2. $\mathbb{V}\left(\overline{X}\right) = \frac{\sigma^2}{n}$. Como es inversamente proporcional, está claro que cuantos más datos haya, mejor nos aproximaremos a lo que queremos estimar.

Teorema II.8 (Teorema central del límite). *Suponemos que $\{X_n\}$ son v.a.i.i.d. con media μ y desviación típica σ finitas. Entonces*

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \xrightarrow[n\to\infty]{d} Z \sim N(0,1)$$

Si denotamos la función de distribución de la normal como

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

entonces

$$\forall t \in \mathbb{R} \quad \mathbb{P}\left\{\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \leq t\right\} \xrightarrow[n\to\infty]{} \Phi(t) = \mathbb{P}\left\{Z \leq t\right\}$$

Por tanto, para n grande se cumple

$$\mathbb{P}\left\{\sqrt{n}\left(\overline{X}-\mu\right) \leq x\right\} \approx \Phi(\frac{x}{\sigma})$$

***aunque las X_i no tengan distribución normal**.*
Es decir:

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma} \iff \sqrt{n}(\overline{X}-\mu) \approx N(0,\sigma) \iff$$
$$X-\mu \approx N(0, \frac{\sigma}{\sqrt{n}}) \iff X \approx N(\mu, \frac{\sigma}{\sqrt{n}})$$

II.3.2. Varianza muestral y poblacional

Una medida importante de dispersión de una variable aleatoria es la varianza

$$\mathbb{V}(X) = \sigma^2 = \int_{\mathbb{R}} (x-\mu)^2 \, dF(x) \tag{II.6}$$

Varianza muestral

Definición II.12 **Varianza muestral**.*El análogo muestral de σ^2 es la* ***varianza muestral****. Utilizando el criterio plugin en (II.6)*

$$\hat{\sigma}_n^2 = \int_{\mathbb{R}} (x-\overline{X})^2 \, d\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Teorema II.9. *La varianza muestral cumple lo siguiente*

$$\mathbb{E}\left(\hat{\sigma}_n^2\right) = \frac{n-1}{n}\sigma^2$$
$$\hat{\sigma}_n^2 \xrightarrow[n\to\infty]{c.s} \sigma^2$$

Por lo tanto, la varianza muestral es un estimador sesgado. No es un problema grande ya que cuando $n \to \infty$ acaba convergiendo a σ^2 y el sesgo

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

también tiende a cero. Es decir, es **asintóticamente insesgado**.

Cuasivarianza muestral

Definición II.13 **Cuasivarianza muestral**. *En lugar de usar $\hat{\sigma}_n^2$ usamos la cuasivarianza muestral, definida como*

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

de tal forma que se tiene

$$\mathbb{E}\left(S^2\right) = \sigma^2$$
$$S^2 \xrightarrow[n\to\infty]{c.s} \sigma^2$$

II.3.3. Estadísticos de orden

Estadístico de orden

Definición II.14 **Estadístico de orden**. *Dada una muestra $\{X_n\}$, se denotan como*

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

las observaciones de la muestra ordenadas de menor a mayor, llamados ***estadísticos de orden****. Cuando la distribución de las v.a. es continua, la probabilidad de coincidencia en valores es 0 y con probabilidad 1 se tiene que*

$$X_{(1)} < \cdots < X_{(n)}$$

Los estadísticos de orden pueden utilizarse para definir la mediana o los cuartiles. Sin embargo, podemos usar la función cuantílica para definir mejor estos conceptos.

Función cuantílica

Definición II.15 **Función cuantílica**. *La función cuantílica en p es el punto que deja una probabilidad p a la izquierda, de tal forma que una proporción p de los individuos de la población X sería menor que el cuantil poblacional de orden p.*

La función cuantílica correspondiente a la función de distribución F se define

$$F^{-1} : \mathbb{R} \longmapsto (0, 1)$$
$$F^{-1}(p) = \text{ínf}\left\{x \,/\, F(x) \geq p\right\}$$

La función cuantílica nos permite obtener los **cuantiles poblacionales de orden** p al valor $F^{-1}(p)$. El análogo es el **cuantil muestral de orden** p, se define a partir de la función de distribución empírica como ${\mathbb{F}_n}^{-1}(p)$.

Capítulo III

Estimación paramétrica

En este tema supondremos que la muestra es absolutamente continua o discreta, con función de densidad o probabilidad $f(.;\theta)$ que es totalmente conocida salvo el valor de un parámetro θ del cuál sólo se conoce su rango de posibles valores Θ, al que se llama el **espacio paramétrico.**

III.1. Estimadores

Estimador

Definición III.1 **Estimador**. *Sean $\{X_n\}$ v.a.i.i.d. con distribución común caracterizada por la función de densidad/masa $f(\cdot;\theta)$, con θ un parámetro desconocido del que sólo se sabe que pertenece al espacio paramétrico $\Theta \subset \mathbb{R}$.*

*El **estimador** es una función medible $\hat{\theta}_n = T_n(X_1, \dots, X_n)$ que se utiliza para estimar o aproximar el valor de θ.*

Cuando tenemos una muestra aleatoria $\{X_n\}$, cada $T_n(X_1, \dots, X_n)$ es un estimador de θ, una variable aleatoria. Si por el contrario tenemos una serie de observaciones de una muestra $\{x_n\}$ entonces $T_n(x_1, \dots, x_n)$ es una **estimación** de θ.

Podemos evaluar la calidad de un estimador con el **error cuadrático medio** (ECM):

$$\mathrm{ECM}(T_n) = \mathbb{E}\left((T_n - \theta)^2\right)$$

Si sumamos y restamos $\mathbb{E}(T_n)$, nos queda que

$$\mathrm{ECM}(T_n) = \mathbb{V}(T_n) + \mathrm{sesgo}^2(T_n)$$

que nos describe el error cuadrático medio en función de la varianza y del sesgo de T_n.

III.1.1. Propiedades interesantes de los estimadores

Buscaremos varias propiedades interesantes de los estimadores:

III.1.1.1. Ausencia de sesgo

Se dice que un estimador T_n es **insesgado** si, siempre que $X_i \sim f(\cdot;\theta)$ se tiene que

$$\mathbb{E}(T_n) = \theta \; \forall\theta \in \Theta$$

III.1.1.2. Consistencia

Se dice que $\{T_n\} = \{T_n(X_1, \ldots, X_n)\}$ es consistente en probabilidad si, siempre que $X_i \sim f(.;\theta)$ se tiene que $T_n \xrightarrow[n\to\infty]{P} \theta \; \forall\theta \in \Theta$.

Si reemplazamos la consistencia en probabilidad por la convergencia casi segura, se obtiene la **consistencia fuerte** o casi segura.

Para probar la consistencia fuerte, usaremos el siguiente teorema:

Teorema III.1 (Teorema de la aplicación continua). *Sea $g : \mathbb{R} \longmapsto \mathbb{R}$ continua en todo punto de un conjunto C tal que $\mathbb{P}\{X \in C\} = 1$, entonces*

- *Si $X_n \xrightarrow[n\to\infty]{d} X$ entonces $g(X_n) \xrightarrow[n\to\infty]{d} g(X)$.*
- *Si $X_n \xrightarrow[n\to\infty]{P} X$ entonces $g(X_n) \xrightarrow[n\to\infty]{P} g(X)$.*
- *Si $X_n \xrightarrow[n\to\infty]{c.s} X$ entonces $g(X_n) \xrightarrow[n\to\infty]{c.s} g(X)$.*

Otra forma de probarlo sería usar la desigualdad de Markov (II.4). Buscamos probar que

$$\mathbb{P}\left\{|T_n - \theta| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

entonces

$$\mathbb{P}\left\{|T_n - \theta| > \varepsilon\right\} = \mathbb{P}\left\{(T_n - \theta)^2 > \varepsilon^2\right\}$$

que por Markov tenemos que

$$\mathbb{P}\left\{(T_n - \theta)^2 > \varepsilon^2\right\} \leq \frac{\mathbb{E}\left(T_n - \theta\right)^2}{\varepsilon^2}$$

y entonces sólo nos quedaría probar que $\mathbb{E}\left(T_n - \theta\right)^2 \xrightarrow[n\to\infty]{} 0$.

También podemos usar condiciones suficientes

Teorema III.2 (Condición de Borel-Cantelli). *Si se cumple que*

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{|T_n - \theta| > \varepsilon\right\} < \infty \; \forall \varepsilon > 0$$

entonces $T_n \xrightarrow[n\to\infty]{c.s} \theta$.

Con esta condición, bastaría ver que la probabilidad o la esperanza convergen y automáticamente se cumpliría la condición.

Ejemplo: *Sean $\{X_n\}$ v.a.i.i.d. con distribución uniforme en el intervalo $[0,\theta]$ con $\theta > 0$. Estudiar la consistencia de los siguientes estimadores de θ*

a)

$$T_n = 2\overline{X}$$

Este estimador se basa en que $\mathbb{E}(X) = \frac{\theta}{2}$. Esto se estima mediante la media muestral $\overline{X}$, y por lo tanto un estimador razonable sería duplicar esa media muestral: $T_n = 2\overline{X}$.

Como T_n se expresa como una función continua de la media muestral, por la LFGN y el teorema de la aplicación continua

$$T_n = g(\overline{X}) \xrightarrow[n\to\infty]{c.s} g(\mu) = 2\mu = 2\mathbb{E}(X) = \theta$$

y por lo tanto tiene consistencia fuerte.

b)

$$T_n = X_{(n)} = \max\{X_1, \dots, X_n\}$$

Aquí usaremos la segunda herramienta: estudiar la probabilidad que el estimador no se aleja del valor esperado en más de ε:

$$\mathbb{P}\left\{|T_n - \theta| > \varepsilon\right\} = \mathbb{P}\left\{\left|X_{(n)} - \theta\right| > \varepsilon\right\} = \mathbb{P}\left\{\theta - X_{(n)} > \varepsilon\right\} = \mathbb{P}\left\{X_{(n)} < \theta - \varepsilon\right\}$$

Si pedimos que el máximo sea menor que $\theta - \varepsilon$, es lo mismo que pedir que lo sean todas las observaciones:

$$\mathbb{P}\left\{X_{(n)} < \theta - \varepsilon\right\} = \mathbb{P}\left\{X_1 < \theta - \varepsilon, \dots, X_n < \theta - \varepsilon\right\}$$

Y con esto logramos quitarnos los estadísticos de orden, que nos causan problemas al tratar de seguir con la demostración. Como las variables de la muestra son independientes, podemos expresarlo todo como producto

$$\prod_{i=1}^{n} \mathbb{P}\left\{X_i < \theta - \varepsilon\right\} = \left(\frac{\theta - \varepsilon}{\theta}\right)^n$$

Esta probabilidad está contenida en el intervalo $(0,1)$ y por lo tanto converge a cero cuando $n \to \infty$. Entonces, T_n es un estimador de θ consistente en probabilidad.

Para examinar si se cumple la condición de Borel-Cantelli, examinamos la serie

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{|T_n - \theta| > \varepsilon\right\} = \sum_{n=1}^{\infty} \left(\frac{\theta - \varepsilon}{\theta}\right)^n < \infty$$

se cumple la condición y es un estimador consistente casi seguro.

Si quisiésemos explorar cuál de los dos estimadores es mejor, usaríamos el error cuadrático medio.

III.1.1.3. Normalidad asintótica

Se dice que una sucesión de estimadores $\{T_n\}$ del parámetro θ es **asintóticamente normal** con tasa $\sqrt{n}$ si

$$\sqrt{n}(T_n - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma)$$

¿Cómo se puede probar la normalidad asintótica? La herramienta se llama el **método delta** y es consecuencia casi inmediata del teorema del valor medio y de las propiedades de la convergencia en distribución: intentaremos expresar el estimador que se propone como una función C^1 de la media muestral y aplicar entonces el Teorema Central del Límite (II.8).

Si llamamos $T_n = g(\overline{X})$ con $g \in C^1$ entonces podemos expresar, con un μ^* entre $\overline{X}$ y μ

$$\sqrt{n}(g(\overline{X}) - g(\mu)) \underset{TVM}{=} g'(\mu^*)\sqrt{n}(\overline{X} - \mu)$$

Como $\overline{X} \xrightarrow[n\to\infty]{c.s} \mu$ entonces $\mu^* \xrightarrow[n\to\infty]{c.s} \mu$ y por lo tanto y usando el Thm. de la aplicación continua (III.1) $g'(\mu^*) \xrightarrow[n\to\infty]{c.s} g'(\mu)$. Al final

$$g'(\mu^*)\sqrt{n}(\overline{X} - \mu) \xrightarrow[n\to\infty]{d} N(0, \left|g'(\mu)\right| \sigma)$$

En general, se habla de normalidad asintótica con tasa a_n si se cumple que

$$a_n(T_n - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma)$$

con a_n sucesión creciente y mayor que cero.

III.1.2. Estimador de máxima verosimilitud (EMV)

En lo que sigue vamos a suponer que $\{X_n\}$ es una muestra formada por v.a.i.i.d. cuya distribución tiene una función de densidad o de masa $f(.;\theta_0)$ perteneciente a una familia de funciones $\{f(.;\theta) \;/\; \theta \in \Theta\}$. θ_0 nos indica el valor real, y θ es un parámetro genérico.

Intuitivamente, lo que pensamos con este método es que la función de masa mide lo verosímil que es que salga un cierto parámetro.

Función de verosimilitud

Definición III.2 **Función de verosimilitud**. *También llamada likelihood function. Dada una muestra fija $\{x_n\}$, se define como*

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

Estimador de máxima verosimilitud

Definición III.3 **Estimador de máxima verosimilitud**. *También llamado EMV o MLE (maximum likelihood estimator) es el argumento que maximiza la función de verosimilitud:*

$$\hat{\theta}_n = \hat{\theta}_n(x, \dots, x_n) = \arg\max_{\theta\in\Theta} L_n(\theta; x_1, \dots, x_n)$$

cuando ese máximo está bien definido.

Para evitar usar derivadas en un producto potencialmente muy largo, podemos maximizar el logaritmo de la verosimilitud, que es creciente y está bien definido porque la densidad es siempre mayor que cero, y los casos en los que sea cero no los estudiamos porque no ocurren (ocurren con probabilidad 0).

III.1.2.1. Cálculo efectivo

El valor del estimador se obtiene como solución de la **ecuación de verosimilitud**.

$$\frac{\partial}{\partial\,\theta} \log L_n = \frac{\partial}{\partial\,\theta} \sum_{i=1}^{n} \log f(\theta; x_i) = 0$$

Ejemplos

Distribución de Poisson de parámetro λ . Suponemos que $X \sim$ Poisson (λ) con $\lambda > 0$, de tal forma que

$$\mathbb{P}\left\{X = x\right\} = e^{-\lambda}\frac{\lambda^x}{x!};\ x \in \mathbb{Z}^+$$

Dada una muestra $\{x_n\}$ de X. Entonces

$$L_n(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \mathbb{P}\left\{X = x\right\} = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-n\lambda}\frac{\lambda^{\sum\limits_{i=1}^{n} x_i}}{x_1!\cdots x_n!}$$

Tomamos logaritmos:

$$\log L_n(\lambda) = -n\lambda + \log\lambda\sum_{i=1}^{n} x_i - \log\left(x_1!\cdots x_n!\right)$$

y derivando

$$\frac{\partial}{\partial\,\lambda}\log L_n(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

de tal forma que nos queda

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

En la imagen (III.1) vemos cómo las diferentes funciones se aproximan a $\lambda = 1$.

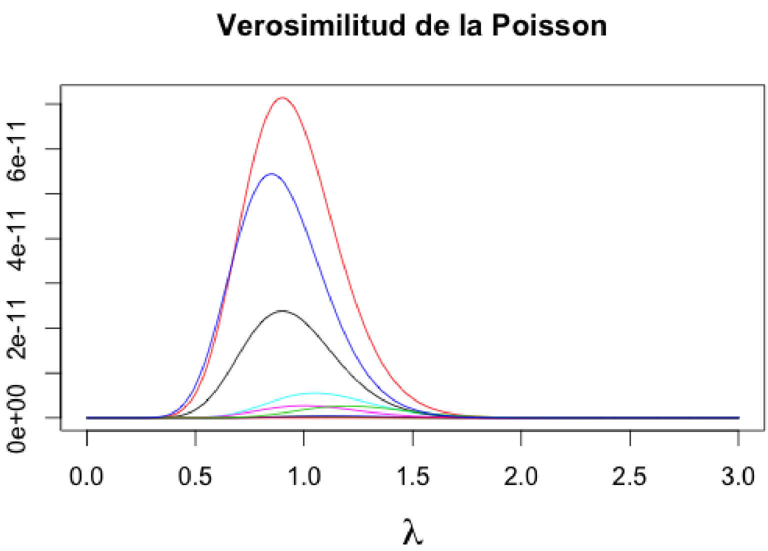


Figura III.1: *Diferentes funciones de verosimilitud para diferentes muestras de la distribución de Poisson*

Distribución normal de parámetros μ, σ Tenemos

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

La función de verosimilitud es

$$L_n = \prod_{i=1}^{n} f(x_i; \mu, \sigma) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Tomamos logaritmos:

$$\log L_n = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Derivamos con respecto de μ

$$\frac{\partial \log L_n}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)(-1) = -\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right) = 0 \leftrightarrow \mu = \overline{x}$$

de tal forma que $\hat{\mu} = \overline{x}$.

Hacemos lo mismo con σ

$$\frac{\partial \log L_n}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 =$$

$$= \frac{1}{\sigma}\left(-n + \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) = 0 \leftrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

luego $\hat{\sigma}^2 = \sigma^2$.

Distribución Weibull La función de densidad de la distribución de Weibull, que toma dos parámetros k y θ, es

$$f(x;\theta,k) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} e^{-\left(\frac{x}{\theta}\right)^k} \mathbb{1}_{[0,\infty)}(x)$$

La función de verosimilitud para los dos parámetros es:

$$L_n(k,\theta) = \prod_{i=1}^{n} f(x_i,\theta,k) = \prod_{i=1}^{n} \frac{k}{\theta}\left(\frac{x_i}{\theta}\right)^{k-1} e^{-\left(\frac{x_i}{\theta}\right)^k} =$$

$$= k^n\theta^{-n}\left(\prod_{i=1}^{n} x_i\right)^{k-1} \theta^{-n(k-1)} e^{-\frac{1}{\theta^k}\sum_{i=1}^{n} x_i^k} = k^n\theta^{-nk}\left(\prod_{i=1}^{n} x_i\right)^{k-1} e^{-\frac{1}{\theta^k}\sum_{i=1}^{n} x_i^k}$$

Tomamos logaritmos:

$$\log L = n \log k - nk \log \theta + (k-1) \sum_{i=1}^{n} \log x_i - \frac{1}{\theta^k} \sum_{i=1}^{n} x_i^k$$

y derivamos con respecto de ambas variables

$$\frac{\partial \log L}{\partial \theta} = -nk\frac{1}{\theta} - (-k)\theta^{-k-1} \sum_{i=1}^{n} x_i^k = \frac{k}{\theta} \left(-n + \frac{1}{\theta^k} \sum_{i=1}^{n} x_k^k \right) = 0$$

$$\frac{\partial \log L}{\partial k} = \frac{n}{k} - n \log \theta + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\theta}^k \right) \log \frac{x_1}{\theta} = 0$$

Con la primera ecuación, tenemos que

$$\theta^k = \frac{1}{n} \sum_{i=1}^{n} x_i^k \iff \hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^k \right)^{\frac{1}{k}}$$

De la segunda ecuación resulta prácticamente imposible despejar k. Sin embargo, podemos usar métodos numéricos para obtener el valor de k.

Teorema III.3 (Invarianza del EMV). *Si τ es una función biyectiva y $\hat{\theta}$ es el e.m.v. de θ, entonces el e.m.v. de $\tau(\theta)$ es $\tau(\hat{\theta})$*

Por ejemplo, tomamos $X \sim N(\mu, \sigma)$. Ya habíamos calculado el e.m.v. de la varianza, que era la varianza muestral. ¿Cómo calcular entonces el e.m.v. de la desviación típica? Sabiendo que $\sigma = \sqrt{\sigma^2}$, tomamos $\tau(\theta) = \sqrt{\theta}$ que es una función biyectiva en $\mathbb{R}^+$ y por lo tanto podemos decir que $\operatorname{emv}(\sigma) = \sqrt{\hat{\sigma}^2}$

III.1.2.2. Motivación del método

Estudiamos la siguiente función

$$\frac{1}{n} \log L_n(\theta) = \frac{1}{n} \log \left(\prod_{i=1}^{n} f(X_i, \theta) \right) = \frac{1}{n} \sum_{i=1}^{n} \log f(X_i; \theta)$$

que por la L.G.N. (II.6) converge a una función $\Psi(\theta)$ que es el valor esperado de esos logaritmos de las muestras:

$$\frac{1}{n} log L_n(\theta) \xrightarrow[n \to \infty]{} \Psi(\theta)$$

donde

$$\Psi(\theta) = \mathbb{E}_{\theta_0} \left[\log f(X; \theta) \right] = \int \log f(x; \theta) f(x; \theta_0) \, dx$$

Teorema III.4 (Teorema MV1). *Sea $X \sim f(.;\theta_0)$. Supongamos que se satisfacen las siguientes condiciones:*

MV0) Parametrización adecuada *Las distribuciones son distintas si el parámetro θ es distinto.*

MV1) Soporte común *Las distribuciones $f(.;\theta)$ tienen un soporte común. Es decir, que las funciones de densidad o de masa tienen valor distinto de cero en los mismos puntos.*

MV2) *$\Psi(\theta)$ es finita para todo $\theta \in \Theta$.*

Entonces θ_0 es el único máximo de la función $\Psi(\theta)$ y además

$$\mathbb{P}_{\theta_0}\left\{L_n(\theta_0; X_1, \dots, X_n) > L_n(\theta; X_1, \dots, X_n)\right\} \xrightarrow[n\to\infty]{} 1 \;\forall \theta \neq \theta_0$$

En el teorema se habla del soporte, definámoslo formalmente:

Soporte

Definición III.4 **Soporte**. *El soporte de una función de distribución o masa f es el conjunto de puntos en el que el valor de f es distinto de 0. Es decir,*

$$\text{soporte } f = \{x \in \mathbb{R} \;/\; f(x) \neq 0\}$$

Para la demostración, primero veremos la siguiente desigualdad:

Teorema III.5 (Desigualdad de Jensen). *Supongamos que X es una v.a. tal que $\mathbb{E}(X) < \infty$ (su esperanza existe y es finita) y que φ es una función convexa*[1] *tal que $\mathbb{E}\left(\varphi(X)\right) < \infty$.*

Entonces

$$\mathbb{E}\left(\varphi(X)\right) \geq \varphi(\mathbb{E}(X))$$

Con esto, podemos pasar a la demostración del teorema (III.4):

Demostración. Decir que

$$L_n(\theta_0; X_1, \dots, X_n) > L_n(\theta; X_1, \dots, x_n)$$

[1] como una parábola $y = x^2$, más o menos

es equivalente a que

$$\log L_n(\theta_0; X_1, \dots, X_n) > \log L_n(\theta; X_1, \dots, x_n)$$
$$\sum_{i=1}^n \log f(X_i;\theta_0) > \sum_{i=1}^n \log f(X_i;\theta)$$
$$0 > \frac{1}{n}\sum_{i=1}^n \big[\log f(X_i;\theta) - \log f(X_i;\theta_0)\big] =$$
$$= \frac{1}{n}\sum_{i=1}^n \log \frac{f(X_i;\theta)}{f(X_i;\theta_0)} \xrightarrow[n\to\infty]{P} \mathbb{E}_{\theta_0}\left(\log \frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right) = -\mathbb{E}_{\theta_0}\left(-\log \frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right) < 0$$

usando la L.G.N (II.6). Aplicando ahora la desigualdad de Jensen (III.5)

$$-\mathbb{E}_{\theta_0}\left(-\log \frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right) > -\log \mathbb{E}_{\theta_0}\frac{f(X_i;\theta)}{f(X_i;\theta_0)}$$

Entonces

$$\mathbb{E}_{\theta_0}\frac{f(X_i;\theta)}{f(X_i;\theta_0)} = \int \frac{f(x;\theta)}{f(x;\theta_0)} f(x;\theta_0)\,dx = \int f(x;\theta)\,dx = 1$$

y por lo tanto

$$-\mathbb{E}_{\theta_0}\left(-\log \frac{f(X_i;\theta)}{f(X_i;\theta_0)}\right) = -\log 1 = 0$$

Entonces, $\forall \varepsilon > 0$

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n \log \frac{f(X_i;\theta)}{f(X_i;\theta_0)} - \mathbb{E}_{\theta_0}\left(\log \frac{f(X;\theta)}{f(X;\theta_0)}\right)\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0 \iff$$
$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n \log \frac{f(X_i;\theta)}{f(X_i;\theta_0)} - \mathbb{E}_{\theta_0}\left(\log \frac{f(X;\theta)}{f(X;\theta_0)}\right)\right| \leq \varepsilon\right\} \xrightarrow[n\to\infty]{} 1$$

Tomo $\varepsilon = \frac{1}{2}\left|\mathbb{E}_{\theta_0}\left(\log \frac{f(X;\theta)}{f(X;\theta_0)}\right)\right|$ y entonces

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n \log \frac{f(X_i;\theta)}{f(X_i;\theta_0)} < \frac{1}{2}\left|\mathbb{E}_{\theta_0}\left(\log \frac{f(X;\theta)}{f(X;\theta_0)}\right)\right| < 0\right\} \xrightarrow[n\to\infty]{} 1$$

□

III.1.2.3. Consistencia del método

Teorema III.6 (Teorema MV2). *Supongamos que se cumplen las condiciones del teorema MV1 (III.4) y adicionalmente*

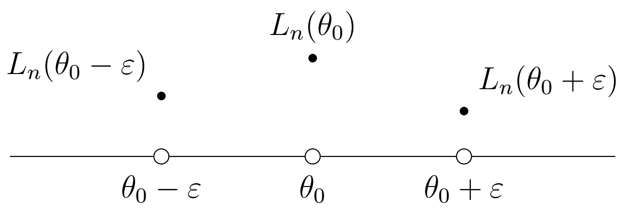
MV3) *El espacio paremétrico Θ es un intervalo abierto no necesariamente finito y, para casi todo x, $f(x;\theta)$ es diferenciable respecto a θ con derivada continua.*

Entonces, con probabilidad tendiente a 1, la ecuación

$$\frac{\partial}{\partial \theta} \log L_n(\theta; X_1, \dots, X_n) = 0 \tag{III.1}$$

tiene una raíz $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$ que ***converge en probabilidad a*** *θ_0 (el verdadero valor del parámetro). Si además suponemos que la raíz es única, entonces $\hat{\theta}_n$ maximiza la verosimilitud L_n y por lo tanto es el estimador de máxima verosimilitud.*

Demostración. Sea $\varepsilon > 0$. Entonces para casi todo[2] x en el intervalo $\Omega = (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ se tiene que $f(x;\theta)$ es diferenciable con derivada continua.



Cogemos entonces un conjunto S_n definido de la siguiente forma:

$$\begin{aligned} S_n = \{(x_1, \dots, x_n) \,/\, & L_n(\theta_0; x_1, \dots, x_n) > L_n(\theta_0 - \varepsilon; x_1, \dots, x_n) \\ & \wedge L_n(\theta_0; x_1, \dots, x_n) > L_n(\theta_0 + \varepsilon; x_1, \dots, x_n)\} \end{aligned}$$

Aplicando el teorema MV1 (III.4), tenemos que $\mathbb{P}_{\theta_0}(S_n) \xrightarrow[n\to\infty]{} 1$.

En algún punto del interior del intervalo Ω hay un máximo local. Como puede haber varios máximos locales, tomo $\hat{\theta}_n$ como el punto de máximo local más cercano a θ_0.

Se cumple que cada uno de esos puntos de máximo satisfacen la ecuación de verosimilitud (III.1). En consecuencia $\hat{\theta}_n$ satisface también esa misma ecuación. Por lo tanto

$$\mathbb{P}\left\{\left|\hat{\theta}_n - \theta_0\right| < \varepsilon\right\} \xrightarrow[n\to\infty]{} 1 \iff \mathbb{P}\left\{\left|\hat{\theta}_n - \theta_0\right| \geq \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

y entonces

$$\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta_0$$

□

III.1.2.4. Información de Fisher

Supongamos el conjunto de todos los estimadores de un parámetro θ. Su error cuadrático medio es

$$\text{ECM}(\hat{\theta}) = \mathbb{V}\left(\hat{\theta}\right) + \text{sesgo}^2(\hat{\theta})$$

Si queremos buscar el *mejor* estimador, buscamos los que minimicen el ECM. Por lo tanto, nos interesaremos en el subconjunto de estimadores insesgados ($\text{sesgo}\,\hat{\theta} = 0$). Sin embargo, no tenemos una forma clara de distinguir cuál es mejor entre esos estimadores insesgados. En esta sección vamos a buscar una *escala*, a la que llamaremos la **información de Fisher**, que nos dará una cota para la varianza de un estimador.

Suponemos que en la integral $\int f(x;\theta)\,dx$ se puede derivar dos veces bajo el signo integral (esto es, que $\int \frac{\partial^2}{\partial\theta^2} f(x;\theta)\,dx$ existe) y que además se puede permutar la integral y la derivada parcial (vemos condiciones suficientes en el apéndice A.1, página 65). Entonces

$$\int f(x;\theta)\,dx = 1 \implies \frac{\partial}{\partial\,\theta}\int f(x;\theta)\,dx = 0$$

Por tanto

$$\int \frac{\partial}{\partial\,\theta}(\log f(x;\theta))f(x;\theta)\,dx = \mathbb{E}_\theta\left(\frac{\partial}{\partial\,\theta}\log f(X;\theta)\right) = 0$$

Si derivamos de nuevo en la integral

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\int f(x;\theta)\,dx = 0 = \int \frac{\partial^2}{\partial\theta^2} f(x;\theta)\,dx = \\
= \int \frac{\partial^2}{\partial\theta^2}\log f(x;\theta) f(x;\theta)\,dx + \int \frac{\partial}{\partial\,\theta}\log f(x;\theta)\cdot\frac{\partial}{\partial\,\theta} f(x;\theta)\,dx = (?) \\
= \int \frac{\partial^2}{\partial\theta^2}\log f(x;\theta) f(x;\theta)\,dx + \int \left(\frac{\partial}{\partial\,\theta}\log f(x;\theta)\right)^2 f(x;\theta)\,dx = \\
= \mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right] + \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\,\theta}\log f(X;\theta)\right)^2\right] = 0
\end{aligned}$$

El segundo valor se llama información de Fisher:

[2] Casi todo: puntos con probabilidad no nula

Información de Fisher

Definición III.5 **Información de Fisher**. *Se denota por $I(\theta)$ la información de Fisher del parámetro θ*

$$I(\theta) = \mathbb{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right) = \mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\,\theta}\log f(X;\theta)\right)^2\right)$$

Representa intuitivamente la cantidad de información acerca del valor del parámetro θ contenida en una observación de X.

¿En qué consiste esa cantidad de información? Tomemos, por ejemplo, una normal $N(0,\theta)$ con θ pequeña. Una observación X que hagamos nos dará mucha información sobre el modelo, ya que todos los valores de la normal están muy agrupados, y por lo tanto $I(\theta)$ será grande. Si tomamos θ grande, una observación X no nos dará mucha información sobre el modelo porque los valores están más dispersos, y por lo tanto tendremos un valor de $I(\theta)$ pequeño.

La información de Fisher nos da una cota inferior para la varianza.

Teorema III.7 (Cota de Fréchet-Cramér-Rao). *Dado $\hat{\theta}$ un estimador insesgado de θ, entonces*

$$\mathbb{V}\left(\hat{\theta}\right) \geq \frac{1}{nI(\theta)}$$

donde $\frac{1}{nI(\theta)}$ se llama la ***cota de Fréchet-Cramér-Rao****.*

Demostración. Tomamos la v.a. Z como la derivada del logaritmo de la verosimilitud

$$Z = \frac{\partial}{\partial\,\theta}\log L_n(X,\theta) = \sum_{i=1}^n \frac{\partial}{\partial\,\theta}\log f(X_i;\theta)$$

La desigualdad de Cauchy-Schwartz establece que [3]

$$\mathbb{V}_\theta\left(T_n\right) \geq \frac{\mathrm{Cov}_\theta^2(Z,T_n)}{\mathbb{V}_\theta\left(Z\right)}$$

Veremos que el numerador vale 1 si T_n es un estimador insesgado, y que $\mathbb{V}_\theta\left(Z\right) = nI(\theta)$.

Primero observamos que

$$\mathbb{E}_\theta\left(Z\right) = \sum_{i=1}^n \mathbb{E}\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right) = 0$$

Y la varianza

$$\mathbb{V}_\theta\left(Z\right) = \sum_{i=1}^n \mathbb{V}_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i;\theta)\right) = \sum_{i=1}^n \mathbb{E}\left(\left(\frac{\partial}{\partial\,\theta}\log f(X;\theta)\right)^2\right)[\theta] = nI(\theta)$$

La primera parte está demostrada.

Ahora vemos que, si $\mathbb{E}_\theta(Z) = 0$, entonces

$$\text{Cov}(Z, T_n) = \mathbb{E}_\theta(ZT_n) - \underbrace{\mathbb{E}_\theta(Z)}_{0}\mathbb{E}_\theta(T_n) = \mathbb{E}_\theta(ZT_n)$$

Como Z y T_n dependen de la muestra

$$\mathbb{E}_\theta(ZT_n) = \mathbb{E}_\theta\big(Z(X_1,\dots,X_n)\cdot T_n(X_1,\dots,X_n)\big) = \int_{\mathbb{R}^n} Z(x_1,\dots,x_n)\cdot T_n(x_1,\dots,x_n)\cdot f_\theta(x_1,\dots,x_n)$$

Como las $X_1,\dots,X_n$ son independientes,

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

y la integral nos queda entonces como una serie de integrales iteradas

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} Z(x_1,\dots,x_n)\cdot T_n(x_1,\dots,x_n)\prod_{i=1}^n f(x_i;\theta)\,dx_i$$

Vemos cuánto vale Z:

$$Z = \frac{\partial}{\partial\theta}\log f(x_i;\theta) = \sum_{i=1}^n \frac{\frac{\partial}{\partial\theta}f(x_i;\theta)}{f(x_i;\theta)}$$

Pero

$$\sum_{i=1}^n \frac{\frac{\partial}{\partial\theta}f(x_i;\theta)}{f(x_i;\theta)}\prod_{i=1}^n f(x_i;\theta)\,dx_i = \sum_{i=1}^n\left[\frac{\partial}{\partial\theta}f(x_i;\theta\cdot\prod_{\substack{j=1\\j\neq i}}^n f(x_j;\theta)\right]$$

que por la regla de la cadena es igual a

$$\frac{\partial}{\partial\theta}\left[\prod_{i=1}^n f(x_i;\theta)\right]$$

y entonces nos queda que

$$\operatorname{cov}\left(()\, Z, T_n\right) = \mathbb{E}_\theta\left(ZT_n\right) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} T_n(x_1, \dots, x_n) \frac{\partial}{\partial\,\theta} \prod_{i=1}^{n} f(x_i;\theta)\, dx_i =$$

$$= \frac{\partial}{\partial\,\theta} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} T_n(x_1, \dots, x_n) \prod_{i=1}^{n} f(x_i;\theta)\, dx_i =$$

$$= \frac{\partial}{\partial\,\theta} \mathbb{E}_\theta\left(T_n\right)$$

Como T_n es un estimador insesgado $\mathbb{E}_\theta\left(T_n\right) = \theta$ y entonces $\text{Cov}\left(Z, T_n\right) = 1$. Por lo tanto, nos queda que

$$\mathbb{V}\left(\hat{\theta}\right) \geq \frac{1}{nI(\theta)}$$

Además, si T_n no fuese un estimador insesgado

$$\mathbb{V}\left(\hat{\theta}\right) \geq \frac{\left(dpa\theta\mathbb{E}_\theta\left(T_n\right)\right)^2}{nI(\theta)}$$

y por lo tanto

$$\text{ECM}(T_n) \geq \frac{\left(\frac{\partial}{\partial\,\theta}\mathbb{E}_\theta\left(T_n\right)\right)^2}{nI(\theta)} + \text{Sesgo}\ ^2(T_n)$$

□

Estimador eficiente

Definición III.6 **Estimador eficiente**. *Se dice que un estimador es eficiente si su varianza es igual a la cota de Fréchet-Cramér-Rao (III.7), es decir*

$$\mathbb{V}\left(\hat{\theta}\right) = \frac{1}{nI(\theta)}$$

III.1.2.5. Eficiencia asintótica

Teorema III.8 (Teorema MV3). *Supongamos que se verifican las condiciones MV0 - MV3 (ver teoremas III.4, III.6) y además:*

MV4) *La integral $\int f(x;\theta)\,dx$ se puede derivar dos veces bajo el signo integral.*

MV5) *Para cada x la densidad $f(x;\theta)$ es tres veces diferenciable con respecto a θ, con la tercera derivada continua en θ.*

[3] Por ejemplo, porque no tengo ni idea de dónde sale esto.

MV6) *La información de Fisher es estrictamente positiva y finita:* $0 < I(\theta_0) < \infty$

MV7) *Para cada* $\theta_0 \in \Theta$ *existen un número* $c > 0$ *y una función* $M(x)$*, que pueden depender de* θ_0*, tales que*

$$\mathbb{E}_{\theta_0}\left(M(X)\right) < \infty$$

y

$$\left|\frac{\partial^3 \log f}{\partial \theta^3}(x;\theta)\right| \leq M(x)\ \forall x;\ \forall \theta \in (\theta_0 - c, \theta_0 + c)$$

Entonces, si $\hat{\theta}_n(X_1, \ldots, X_n)$ *es cualquier sucesión consistente de soluciones de las ecuaciones de verosimilitud, se verifica*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{\sqrt{I(\theta_0)}}\right)$$

Demostración.

$$\tilde{\Psi}_n(\theta) = L_n \tag{III.2}$$

$$\tilde{\Psi}'_n = \frac{\partial\, \Psi_n(\theta)}{\partial\, \theta} \tag{III.3}$$

$$f' = \frac{\partial\, f}{\partial\, \theta} f \tag{III.4}$$

donde la función III.3 se llama el **score** (quizás).

Recordemos que $\hat{\Psi}_n(\theta)$ depende de la muestra. Para cada muestra fija se tiene

$$\tilde{\Psi}_n(\hat{\theta}_n) = \hat{\Psi}'_n(\theta_0) + (\hat{\theta}_n - \theta_0)\Psi''_n(\theta_0) + \frac{\left(\hat{\theta}_n - \theta_0\right)^2}{2}\tilde{\Psi}'''_n(\theta^*_n)$$

Para algún θ^*_n entre $\hat{\theta}_n$ y θ_0. Como el primer miembro es 0, resulta

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = \frac{\frac{1}{\sqrt{n}}\tilde{\Psi}'_n(\theta)^2}{-\frac{1}{n}\Psi''_n(\theta_0) - \frac{1}{2n}\left(\hat{\theta}_n - \theta_0\right)\tilde{\Psi}'''_n(\theta^*_n)}$$

Vamos a demostrar que esto converge en tres pasos:

- Numerador converge a $N(0, \sqrt{I(\theta_0)})$.
- Primera parte converge a $I(\theta_0)$.

- Segunda parte denom. converge a 0 en prob.

Tendremos por lo tanto que

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow[n\to\infty]{d} \frac{N(0, \sqrt{I(\theta_0)}}{I(\theta_0)+0}$$

Y usando la tercera condición del teorema de Slutsky (II.3), tendremos que

$$\frac{N(0, \sqrt{I(\theta_0)}}{I(\theta_0)+0} \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{\sqrt{I(\theta_0)}}\right)$$

Parte 1: Numerador

$$\frac{1}{\sqrt{n}}\tilde{\Psi}'_n(\theta_0) = \frac{\sqrt{n}}{n}\sum_{i=1}^n \left[\frac{f'(X_i;\theta_0)}{f(X_i;\theta_0)} - \mathbb{E}_{\theta_0}\left(\frac{f'(X_i;\theta_0)}{f(X_i;\theta_0)}\right)\right]$$

Como $\mathbb{E}_{\theta_0}\left(\frac{f'(X_i;\theta_0)}{f(X_i;\theta_0)}\right) = 0$ (vete tú a saber por qué), la aplicación del TCL (II.8) a las variables $Y_i = \frac{f'(X_i;\theta_0)}{f(X_i;\theta_0)}$ y la definición de $I(\theta_0)$ proporcionan directamente

$$\frac{1}{\sqrt{n}}\hat{\Psi}'_n(\theta_0) \xrightarrow[n\to\infty]{d} N(0, \sqrt{\mathbb{V}(Y)})$$

Calculamos ahora esa desviación típica:

$$\mathbb{V}_\theta(Y) = \mathbb{E}_\theta\left(Y^2\right) - \mathbb{E}_\theta(Y)^2 = \mathbb{E}_\theta\left(Y^2\right) =$$
$$= \mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right) = I(\theta)$$

Y por lo tanto nos queda que

$$\frac{1}{\sqrt{n}}\hat{\Psi}'_n(\theta_0) \xrightarrow[n\to\infty]{d} N\left(0, \sqrt{I(\theta_0)}\right)$$

Parte 2: Denominador A Operamos con

$$-\frac{1}{n}\Psi''_n(\theta_0)$$

Si derivamos de nuevo $\tilde{\Psi}''_n$ con respecto a θ tenemos que

$$\tilde{\Psi}''_n(\theta) = \sum_{i=1}^n \frac{f''(x_i;\theta)f(x_i;\theta) - \left(f'(x_i;\theta)\right)^2}{f^2(x_i;\theta)}$$

Entonces $\frac{1}{n}\Psi_n''(\theta_0)$ es un promedio, y por la LGN (II.6)

$$-\frac{1}{n}\Psi_n''(\theta_0) \xrightarrow[n\to\infty]{P} -\mathbb{E}_{\theta_0}\left(\frac{f''(x_i;\theta)f(x_i;\theta) - \left(f'(x_i;\theta)\right)^2}{f^2(x_i;\theta)}\right) =$$

$$= \underbrace{\mathbb{E}_{\theta_0}\left(\left(\frac{f'(X_i;\theta_0)}{f(X_i;\theta_0)}\right)^2\right)}_{I(\theta_0)} - \mathbb{E}_{\theta_0}\left(\frac{f''(X_i;\theta_0)}{f(X_i;\theta_0)}\right)$$

Operamos ahora con la segunda parte

$$\mathbb{E}_{\theta_0}\left(\frac{f''(X_i;\theta_0)}{f(X_i;\theta_0)}\right) = \int_{\mathbb{R}} \frac{f''(X_i;\theta_0)}{f(X_i;\theta_0)} f(X_i;\theta_0)\,dx =$$

$$= \int_{\mathbb{R}} \left.\frac{\partial^2}{\partial\theta^2} f(x;\theta)\right|_{\theta=\theta_0} dx$$

y como según el enunciado del teorema podemos permutar la derivada con la integral dos veces, tenemos que

$$\int_{\mathbb{R}} \left.\frac{\partial^2}{\partial\theta^2} f(x;\theta)\right|_{\theta=\theta_0} dx = \left.\frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}} f(x;\theta)\,dx\right|_{\theta=\theta_0} = \left.\frac{\partial^2}{\partial\theta^2} 0\right|_{\theta=\theta_0} = 0$$

Por lo tanto

$$-\frac{1}{n}\Psi_n''(\theta_0) \xrightarrow[n\to\infty]{P} I(\theta_0)$$

Paso 3: Segunda parte del denominador

$$\frac{1}{2n}\left(\hat{\theta}_n - \theta_0\right)\tilde{\Psi}_n'''(\theta_n^*) \xrightarrow[n\to\infty]{P} 0$$

Por hipótesis del teorema, $\hat{\theta}_n$ se considera consistente y entonces

$$\left(\hat{\theta}_n - \theta_0\right) \xrightarrow[n\to\infty]{} 0$$

Analizaremos ahora la segunda parte de esa ecuación, $\tilde{\Psi}_n'''(\theta_n^*)$, y demostraremos que tiende a una constante.

$$\tilde{\Psi}_n'''(\theta_n^*) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^3}{\partial\theta^3}\log f(X_i;\theta)$$

Como $\hat{\theta}_n$ es consistente, θ_n^*, que es un punto intermedio entre $\hat{\theta}_n$ y θ_0, también tiende a θ_0 en probabilidad. Entonces podemos aplicar la hipótesis MV7 del teorema y acotar la derivada parcial:

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(X_i; \theta) \right| \leq M(X_i)$$

y por lo tanto podemos acotar en probabilidad

$$\left| \tilde{\Psi}_n'''(\theta_n^*) \right| < \frac{1}{n} \sum_{i=1}^n M(X_i)$$

Este término converge a una constante por lo tanto, y entonces se cumple que

$$\frac{1}{2n} \left(\hat{\theta}_n - \theta_0 \right) \tilde{\Psi}_n'''(\theta_n^*) \xrightarrow[n \to \infty]{P} 0$$

□

III.1.3. Método de los momentos

Sea $X \sim f(x;\theta)$, donde $\theta = (\theta_1, \dots, \theta_p)$ es un parámetro p-dimensional, con $p \geq 1$.

Si los momentos $\alpha_k(\theta) = \mathbb{E}_{\theta_0}\left(X^k\right),\ k = 1, \dots, p$ son funciones sencillas de los θ_i, un procedimiento natural para obtener un estimador de θ es resolver en $\theta_1, \dots, \theta_p$ el sistema de ecuaciones

$$
\begin{gathered}
m_1 = \alpha_1(\theta) \\
\dots \\
m_p = \alpha_p(\theta)
\end{gathered}
$$

donde cada m_k es el momento muestral:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

La idea es estimar el parámetro de tal forma que los momentos muestrales coincidan con los momentos poblacionales. Por la LGN, cuando $n \to \infty$ entonces $m_k \to \alpha_k(\theta_0)$.

El método de los momentos se utiliza poco ya que da peores estimadores que el EMV. Sin embargo, puede resultar muy útil en casos en los que el EMV se calcula difícilmente o directamente no se puede calcular. Ahí hay que usar métodos numéricos de aproximación, y usando el método de los momentos podemos encontrar una primera aproximación que mejore la convergencia de los algoritmos numéricos de búsqueda de raíces.

III.1.3.1. Ejemplos

Si se tiene el modelo

$$f(x;\theta) = \frac{1+\theta x}{2}\mathbb{1}_{[-1,1](x)}\,\theta \in [-1,1]$$

no es sencillo calcular el EMV pero sí obtener el estimador por el método de los momentos:

$$\mathbb{E}_\theta\left(X\right) = \int_{-1}^{1} x f(x;\theta)\,dx = \frac{\theta}{3}$$

Por tanto, la solución de $\overline{X} = \mathbb{E}_\theta\left(X\right)$ es $\tilde{\theta}_n = 3\overline{X}$, cuya varianza es

$$\mathbb{V}_\theta\left(\tilde{\theta}_n\right) = \mathbb{V}_\theta\left(3\overline{X}\right) = 9\frac{\sigma^2}{n} = \frac{3-\theta^2}{n}$$

ya que

$$\sigma^2 = \mathbb{V}_\theta\left(X\right) = \mathbb{E}_\theta\left(X^2\right) - \mathbb{E}_\theta\left(X\right)^2 = \frac{1}{3} - \frac{\theta^2}{9}$$

Este estimador es consistente ya que, por la LGN, $3\overline{X} \xrightarrow[n\to\infty]{} 3\mathbb{E}_\theta(X)$.

Supongamos un ejemplo más complicado: $X \sim Beta(a, b)$.

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$$

y

$$\begin{aligned}\mathbb{E}_\theta(X) &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^a (1-x)^{b-1}\,dx = \\ &= \frac{\Gamma(a+b)}{\Gamma(a)}\frac{\Gamma(a+1)}{\Gamma(a+b+1)} \underbrace{\int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} x^a(1-x)^{b-1}\,dx}_{=1 \text{ (f. densidad)}} =\end{aligned}$$

Sabiendo que $\Gamma(p+1) = p\Gamma(p)$

$$\frac{\Gamma(a+b)}{\Gamma(a)}\frac{\Gamma(a+1)}{\Gamma(a+b+1)} = \frac{\Gamma(a+b)}{\Gamma(a)}\frac{a\Gamma(a)}{(a+b)\Gamma(a+b)} = \frac{a}{a+b}$$

y los estimadores quedan como

$$\hat{a} = \overline{X}\left(\frac{\overline{X}(1-\overline{X})}{s^2} - 1\right)$$
$$\hat{b} = (1-\overline{X})\left(\frac{\overline{X}(1-\overline{X})}{s^2} - 1\right)$$

III.1.4. Metodología bayesiana

En muchos casos se tiene cierta información a priori, antes de extraer la muestra, sobre la probabilidad de los diferentes valores del parámetro θ. En estos casos se sabe, o se supone, que ciertos intervalos de valores de θ son *más probables que otros* y se concreta esta información en una **distribución a priori sobre** θ cuya función de densidad se denota $\pi(\theta)$.

De manera formal, la estadística bayesiana considera que el parámetro es una variable aleatoria y que la información previa se puede expresar a través de la distribución a priori del parámetro.

Entonces, si antes teníamos una v.a. $X \sim f(x;\theta)$, ahora lo que diremos es que X sigue una distribución condicionada por un parámetro: $X \sim f(x|\theta)$.

En este caso, la muestra $X_1, \dots, X_n$ contiene información de la muestra y también de nuestro parámetro. Es decir, que podemos considerar la función de distribución de la muestra como

$$\prod_{i=1}^n f(x_i|\theta)$$

Para juntar toda esta información usaremos el Teorema de Bayes:

Teorema III.9 (Teorema de Bayes). *Sea $A_1, A_2, \dots$ una partición del espacio muestral y sea B un suceso cualquiera. Entonces*

$$\mathbb{P}\left\{A_i|B\right\} = \frac{\mathbb{P}\left\{A_i \cap B\right\}}{\mathbb{P}\left\{B\right\}} = \frac{\mathbb{P}\left\{B|A_i\right\} \cdot \mathbb{P}\left\{A_i\right\}}{\sum_j \mathbb{P}\left\{B|A_j\right\} \cdot \mathbb{P}\left\{A_j\right\}}$$

Esta formulación se refiere a sucesos probabilísticos. Podemos reformularla con la información a priori del parámetro:

$$\pi(\theta|x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{\displaystyle\int_\Theta f(x_1, \dots, x_n|\tau)\pi(\tau)\,d\tau} \tag{III.5}$$

donde Θ es todo el espacio paramétrico. A $\pi(\theta|x_1, \dots, x_n)$ se le denomina ***distribución a posteriori***

Como π es una función de distribución, tenemos que

$$\int_\Theta \pi(\theta|x_1, \dots, x_n)\,d\theta = 1$$

para toda posible muestra $(x_1, \dots, x_n)$. Estudiaremos entonces la siguiente integral

$$\int_\Theta \frac{f(x_1, \dots, x_n|\theta)\pi(\theta)}{\displaystyle\int_\Theta f(x_1, \dots, x_n|\tau)\pi(\tau)\,d\tau}$$

En esta, integral, el término

$$\int_\Theta f(x_1, \dots, x_n|\tau)\pi(\tau)\,d\tau$$

es constante. Por lo tanto, lo que nos interesará será el numerador, la integral

$$\int_\Theta f(x_1, \dots, x_n|\theta)\pi(\theta)\,d\theta$$

que nos dará la información que necesitamos.

Estimador Bayes

Definición III.7 **Estimador Bayes**. *Se define, para cada muestra dada $(x_1, \dots, x_n)$ como la esperanza de la distribución a posteriori:*

$$T_n(x_1, \dots, x_n) = \int_\Theta \theta\pi(\theta|x_1, \dots, x_n)\,d\theta$$

III.1.4.1. Ejemplos

La estadística bayesiana se suele usar para estimar los votantes de un partido político. Por ejemplo, sea θ la proporción de votantes de un partido P, y sea X la v.a. Bernoulli que toma valor 1 cuando un votante elige P y 0 en otro caso. Es decir

$$\begin{cases} f(x|\theta) = \theta & \text{si } x = 1 \\ f(x|\theta) = 1 - \theta & \text{si } x = 0 \end{cases}$$

Entonces tenemos que

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

Suponemos que la distribución a priori es una Beta(4,10):

$$\pi(\theta) = \frac{\Gamma(14)}{\Gamma(4)\Gamma(10)}\theta^3(1-\theta)^9 \mathbb{1}_{[0,1]}(\theta)$$

Así pues, aplicando la fórmula de Bayes (III.5) nos queda

$$\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\theta^3(1-\theta)^9 = \theta^{3+\sum x_i}(1-\theta)^{9+n-\sum x_i} \tag{III.6}$$

y entonces

$$\pi(\theta|x_1, \dots, x_n) \sim Beta(4 + \sum x_i, 10 + n - \sum x_i)$$

El estimador Bayes es, por lo tanto

$$T_n = \frac{4 + \sum x_i}{14 + n} = \underbrace{\frac{n}{4 + 10 + n}\overline{x}}_{(A)} + \underbrace{\frac{4+10}{4+10+n}\frac{4}{4+10}}_{(B)}$$

Es decir, pondera las dos información que teníamos: la media de la distribución a priori (B) y la media muestral (A). Si nos fijamos en la expresión, si tenemos un tamaño muestral muy grande ($n \to \infty$) damos mucho más peso a la información de la muestra que a la distribución a priori. Sin embargo, si tenemos menos muestras nuestra distribución a priori influirá más en el resultado.

Con los datos $\sum x_i = 125$ y $n = 1000$, el estimador Bayes toma valor 0.127, mientras que el e.m.v. valdría 0.125. Es decir, nuestro estimador bayesiano pondera la información que teníamos previamente y considera que en nuestra distribución a priori era más probable valores más altos.

Curiosamente, en (III.6) hemos pasado de una distribución a priori a una distribución a posteriori fácilmente identificable con una distribución Beta. Esto tiene que ver con el concepto de familias conjugadas.

III.1.4.2. Familias conjugadas

Familia conjugada

Definición III.8 **Familia conjugada**. *Sea $\mathcal{F}$ una familia de distribuciones paramétricas $f(\cdot|\theta)$, $\theta \in \Theta$; y sea Π una familia de distribuciones a priori $\pi(\theta)$ sobre el parámetro θ.*

Diremos que Π es la familia de dsitribuciones a priori conjugada de $\mathcal{F}$ si la distribución a posteriori $\pi(\theta|x_1, \dots, x_n)$ también pertence a Π para toda muestra $(x_1, \dots, x_n)$ y para toda a priori de Π.

Tenemos varias familias conjugadas identificadas:

$\mathcal{F}$	Π
Binomial	Beta
Normal	Normal

Cuadro III.1: *Familias conjugadas*

III.2. Estimación por intervalos de confianza

Al igual que en el tema anterior, vamos a obtener información sobre un parámetro desconocido $\theta \in \Theta$ a partir de una muestra $X_1, \dots, X_n$. Habíamos logrado una estimación puntual, pero, ¿por qué va a ser válido sólo ese valor? ¿Podría ser válido un valor cercano al estimador?

Este tema responde a esa pregunta: ofrece un intervalo que típicamente contiene a un estimador puntual, de posibles valores para un parámetro. Veremos cómo construir ese intervalo y la información que ofrecen.

Intervalo de confianza

Definición III.9 **Intervalo de confianza**. *Sea una muestra $X_1, \dots, X_n$ de una v.a. con una función de distribución $F(.;\theta)$, con $\theta \in \Theta \subset \mathbb{R}$ un parámetro desconocido. Sean dos estadísticos $T_n^{(1)}(X_1, \dots, X_n)$ y $T_n^{(2)}(X_1, \dots, X_n)$ con $T_n^{(1)} < T_n^{(2)}$ y un valor $\alpha \in (0, 1)$. Supongamos que se verifica*

$$\mathbb{P}_\theta \left\{ T_n^{(1)}(X_1, \dots, X_n) < \theta < T_n^{(2)}(X_1, \dots, X_n) \right\} = 1 - \alpha \;\forall \theta$$

Entonces para una realización concreta de la muestra $x_1, \dots, x_n$ se dice que el intervalo $(T_n^{(1)}(x_1, \dots, x_n), T_n^{(2)}(x_1, \dots, x_n))$ es un intervalo de confianza para θ con nivel de confianza $1 - \alpha$ y lo denotaremos como

$$IC_{1-\alpha}(\theta)$$

Probemos esta definición con una muestra $X_1, \dots, X_n$ de v.a.i.i.d. $N(\mu, \sigma)$ donde μ es un parámetro desconocido y σ es conocida. Se sabe que

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

y, tipificando,

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

Por tanto, si para cualquier $\alpha \in (0,1)$, z_α denota el cuantil $1-\alpha$ en la normal estándar ($\Phi(z_\alpha) = 1 - \alpha$, siendo Φ la función de distribución de la $N(0,1)$) tenemos

$$\mathbb{P}_\mu\left\{-z_{\alpha/2} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha/2}\right\} = 1 - \alpha$$

y, despejando

$$\mathbb{P}_\mu\left\{\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

Y por lo tanto, el intervalo

$$\left(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

es un **intervalo de confianza de nivel** $1-\alpha$ **para** μ.

Intuitivamente y en términos frecuentistas, si por ejemplo $1-\alpha = 0.95$ y extraemos muchas muestras de una $N(0,1)$ aproximadamente en el 95 % de los casos el intervalo contendrá el verdadero valor de μ.

III.2.1. Intervalos de confianza asintóticos basados en el TCL

Si X no es normal, sabemos que si μ y σ son finitas, encontes $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ por el TCL (II.8). Entonces

$$1 - \alpha =\simeq \mathbb{P}\left\{-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}\right\}$$

Es decir, obtenemos un intervalo de confianza aproximado si el tamaño de la muestra es grande.

Aplicación: Intervalo de confianza aproximado para una proporción p Sean $X_1, \dots, X_n$ i.i.d. Bernoulli(p). Por el TCL

$$\frac{\overline{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

y reemplazando p por su estimador natural $\hat{p} = \overline{X}$ obtenemos que el intervalo de confianza aproximado para p es

$$\left(\overline{x} - z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}, \overline{x} + z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x}}{n}}\right)$$

III.2.2. Método de la cantidad pivotal

Una metodología general para obtener un intervalo de confianza para θ consiste en encontrar una función $Q(\theta; X_1, \dots, X_n)$, llamada **cantidad pivotal** cuya distribución no dependa de θ y sea conocida, al menos de modo aproximado. A partir de esta distribución, fijado un valor $\alpha \in (0,1)$ se obtienen dos valores $q_1(\alpha), q_2(\alpha)$ tales que

$$\mathbb{P}_\theta \left\{q_1(\alpha) < Q(\theta; X_1, \dots, X_n) < q_2(\alpha)\right\} = 1 - \alpha$$

Despejando θ se obtiene una expresión del tipo

$$\mathbb{P}_\theta \left\{T_n^{(1)}(X_1, \dots, X_n) < T_n^{(2)}(X_1, \dots, X_n)\right\} = 1 - \alpha$$

III.2.3. Construcción de intervalos de confianza habituales

III.2.3.1. Distribución χ^2

Estamos interesados en obtener intervalos de confianza exactos, válidos para cualquier n, para σ^2 en una normal. Para ello presentaremos una distribución auxiliar que tiene una especial importancia en estadística, la **distribución** χ_k^2, que en realidad es la distribución $\Gamma(\frac{1}{2}, \frac{k}{2})$. Esta distribución surge del estudio de la distribución de las formas cuadráticas $X'AX$. En particular, si $\{Z_n\}$ son variables aleatorias normales estandarizadas, entonces

$$\sum Z_k^2 \sim \chi^2$$

De hecho, aplicando esto a una suma de varias v.a. $X_1, \dots, X_n$ S^2, nos queda que

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

Este resultado proporciona directamente una cantidad pivotal y, en consecuencia, un intervalo de confianza de nivel $1-\alpha$ para σ^2:

$$\left(\frac{(n-1)s^2}{\chi_{n-1;\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{n-1;1-\alpha/2}^2}\right)$$

III.2.3.2. Distribución t de Student

Sea $Z \sim N(0,1)$ y $W \sim \chi^2_k$. Supongamos que Z y W son independientes. Entonces la distribución de la v.a.

$$T = \frac{Z}{\sqrt{W/k}}$$

se denomina distribución t de Student con k grados de libertad. Su forma se aproxima a una normal $N(0,1)$.

Teorema III.10 (Lema de Fischer-Cochran). *Si $X_1, \dots, X_n$son v.a.i.i.d. con distribución $N(\mu, \sigma)$ entonces $\overline{X}$ y S^2 (desviación) son estadísticos independientes.*

Este teorema tiene una consecuencia importante, y es que podemos obtener un intervalo de confianza exacto para μ en $N(\mu, \sigma)$ aún cuando σ es desconocida.

III.2.4. Intervalos de confianza bayesianos

En un problema de inferencia con un enfoque bayesiano, el elemento fundamental para realizar la inferencia es la distribución a posteriori $\pi(\theta|x_1, \dots, x_n)$. A partir de esa distribución se define una **región creíble** de nivel $1-\alpha$ como un subconjunto $A \subseteq \Theta$ tal que

$$\int_A \pi(\theta|x_1, \dots, x_n)\, d\theta = 1 - \alpha$$

Capítulo IV

Contraste de hipótesis

IV.1. Conceptos básicos

El objetivo de la teoría de contraste de hipótesis es **elegir entre dos posibilidades excluyentes**, las hipótesis nula e alternativa, relativas al valor de un parámetro poblacional a partir de la información proporcionada por los datos muestrales.

Sea $X_1, \dots, X_n$ una muestra aleatoria de una v.a. X con función de distribución F_θ donde $\theta \in \Theta$. Dada una partición del espacio paramétrico $\Theta = \Theta_0 \cup \Theta_1$, deseamos decidir, en base a la muestra obtenida, si θ está en Θ_0 o en Θ_1. En el primer caso se cumple la hipótesis nula, en el segundo la alternativa. Ambas hipótesis son excluyentes.

Para resolver el problema definiremos una región de rechazo. Esta región $R \subseteq \mathbb{R}^n$ nos permitirá valorar si el parámetro está en Θ_0 o en Θ_1 en base a la muestra obtenida. De esta forma, si $(x_1, \dots, x_n) \in R$, se rechaza la hipótesis nula.

El paso más importante del contraste de hipótesis es construir la región de rechazo R, y a partir de entonces los pasos son muy mecánicos. En el apéndice A.3, página 72, tenemos varias muestras de regiones de rechazo.

En el test de hipótesis podemos cometer dos tipos de fallos:

- **Error de tipo I** Rechazar H_0 cuando H_0 es cierta.
- **Error de tipo II** Aceptar H_0 cuando H_0 es falsa.

Para medir la probabilidad de cometer uno de esos fallos definimos la función de potencia

Función de potencia

Definición IV.1 **Función de potencia**. *La función de potencia de un test con región de rechazo R para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ es la función*

$$\begin{aligned} \beta_n : \Theta &\longmapsto [0, 1] \\ \theta &\longmapsto \beta_n(\theta) = \mathbb{P}_\theta \left\{ (X_1, \dots, X_n) \in R \right\} \end{aligned}$$

y nos da la probabilidad de rechazar la hipótesis Θ_0.

IV.1.1. Teoría de Neyman-Pearson

Nos gustaría que $\beta_n(\Theta_0) = 0$ y que $\beta_n(\Theta_1) = 1$, pero normalmente no pasará esto, sino que β_n será una función continua y suave del parámetro.

La teoría de Neyman-Pearson trata de responder a este problema con los dos siguientes pasos:

Acotar la máxima probabilidad de error de tipo I

- Se fija un **nivel de significación** $\alpha \in (0, 1)$. Típicamente se toma $\alpha = 0.05$.
- Se define el **tamaño de un test** como la máxima probabilidad de error de tipo I, o como

$$\max_{\theta \in \Theta_0} \mathbb{P}_\theta \{R\} = \max_{\theta \in \Theta} \beta_n(\theta)$$

- Se busca una región de rechazo R tal que

$$\max_{\theta \in \Theta_0} \mathbb{P}_\theta \{R\} \leq \alpha$$

Tal y como hemos definido α, se puede considerar que el nivel de significación nos indica la probabilidad de cometer un error de tipo I, es decir, de rechazar H_0 cuando es cierta. Por lo tanto, cuanto menor es el nivel de significación más *seguros* estamos de que no estamos rechazando H_0 por error.

Minimizar la probabilidad de error de tipo II Se intenta buscar una región de rechazo R que maximice la función de potencia cuando $\theta \in \Theta_1$.

Aquí podemos ver por qué las dos hipótesis no son simétricas. Los tests de hipótesis están diseñados para controlar la probabilidad máxima de rechazar H_0 cuando es cierta. En consecuencia, suelen ser conservadores con la hipótesis nula: hace falta mucha evidencia muestral para rechazar H_0. Observemos que es posible que, con los mismos datos, H_0 se rechace para un nivel de significación $\alpha = 0.05$ y se acepte para $\alpha = 0.01$.

Además de la asimetría, tenemos que pensar que al aceptar H_0 no significa que la hayamos demostrado, sino simplemente que no se ha encontrado suficiente evidencia empírica a nivel prefijado α en contra de H_0. **No es una demostración matemática**.

IV.2. Problema de una muestra

En una primera aproximación, los problemas de contraste de hipótesis pueden clasificarse en problemas de una muestra o de dos, según haya sólo una población de interés o queramos comparar dos poblaciones y dispongamos de una muestra de cada una de ellas. Presentaremos las ideas básicas en el caso de los problemas de una muestra pero pueden extenderse de modo análogo a los de dos muestras.

Dualidad con los intervalos de confianza En algunos casos de hipótesis nula simple, aparece una dualidad entre el contraste de hipótesis y los intervalos de confianza (III.2). Si tenemos $H_0 : \; \mu = \mu_0$, entonces aceptar H_0 significa que $\mu \in IC_{1-\alpha}(\theta)$, es decir, que está en el intervalo de confianza. La región de rechazo sería entonces

$$R = \{(x_1, \dots, x_n) \;/\; \theta(x_1, \dots, x_n) \notin IC_{1-\alpha}(\theta)\}$$

P-valor del contraste

Definición IV.2 **p-valor del contraste**. *Se define el p-valor del contraste como el ínfimo de los niveles de significación α para los que se rechaza H_0.*

De esta forma, si α es menor que el p-valor, aceptaremos H_0 y si es mayor, la rechazaremos.

¿Qué información nos va a dar el p-valor? Supongamos que tenemos, por ejemplo, un **p-valor pequeño** (< 0.01). Con este valor rechazaríamos la hipótesis nula para los valores más habituales de niveles de significación ($0.01, 0.05, 0.1$). Por lo tanto, en este caso **lo razonable sería rechazar** H_0.

Por otra parte, supongamos que tenemos un **p-valor grande** (> 0.1). En este caso, aceptaríamos la hipótesis nula para los valores más habituales de α, y entonces **lo razonable sería aceptar** H_0.

Un p-valor que se encuentra entre 0.01 y 0.1 se considera **dudoso**. Lo razonable es revisar la muestra, y si es posible, aumentar su tamaño. **No se puede decidir** de manera razonable entre H_0 y H_1.

De forma general, el p-valor de contraste nos dice la probabilidad de observar la muestra que hemos obtenido suponiendo que H_0 es cierta. Si es muy bajo, nos indicará que es muy poco probable que la muestra obtenida haya salido así por pura casualidad.

IV.2.1. Regiones de rechazo para contrastes habituales

IV.2.1.1. Contraste de la media de una distribución

En todo caso se rechaza H_0 cuando $(X_1, \dots, X_n) \in R$. Para hallar las regiones de rechazo buscaremos los **estadísticos de contraste**, medidas de lo razonable que es la hipótesis nula y que depende de la muestra obtenida. Cuando la hipótesis nula sea cierta, el estadístico del contraste estará en zonas de alta probabilidad.

Distribución normal con varianza conocida Primero construiremos el estadístico del contraste Z, que depende de la media muestral obtenida.

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Si $H_0 : \mu = \mu_0$ es cierta entonces $Z \sim N(0,1)$. Entonces las regiones de rechazo son

H_0	R
$\mu = \mu_0$	$\{(x_1, \dots, x_n) \,/\, \lvert Z \rvert \geq z_{\frac{\alpha}{2}}\}$
$\mu \leq \mu_0$	$\{(x_1, \dots, x_n) \,/\, Z \geq z_{\frac{\alpha}{2}}\}$
$\mu \geq \mu_0$	$\{(x_1, \dots, x_n) \,/\, Z \leq z_{\frac{\alpha}{2}}\}$

Cuadro IV.1: *Regiones de rechazo para una normal $N(\mu, \sigma)$.*

Distribución normal con varianza desconocida Sea $X_1, \dots, X_n$ una muestra aleatoria de $X \sim N(\mu, \sigma)$ con σ desconocido. Entonces el estadístico del contraste sigue una distribución T de Student de $n-1$ grados de libertad:

$$T = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

H_0	R
$\mu = \mu_0$	$\{(x_1, \dots, x_n) \,/\, \lvert T \rvert \geq t_{\frac{\alpha}{2}}\}$
$\mu \leq \mu_0$	$\{(x_1, \dots, x_n) \,/\, T \geq t_{\frac{\alpha}{2}}\}$
$\mu \geq \mu_0$	$\{(x_1, \dots, x_n) \,/\, T \leq t_{\frac{\alpha}{2}}\}$

Cuadro IV.2: *Regiones de rechazo para una normal $N(\mu, \sigma)$ con σ desconocida.*

Tests de nivel aproximado α (muestras grandes) para la media de cualquier distribución Sea $X_1, \dots, X_n$ una muestra aleatoria de X con $\mathbb{E}(X) = \mu < \infty$. Entonces el estadístico del contraste es

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \overset{TCL}{\sim} N(0,1)$$

si $H_0 : \mu = \mu_0$ es cierta. Por lo tanto, nos quedamos con las siguientes regiones:

H_0	R
$\mu = \mu_0$	$\{(x_1, \dots, x_n) \;/\; \lvert Z\rvert \geq z_{\frac{\alpha}{2}}\}$
$\mu \leq \mu_0$	$\{(x_1, \dots, x_n) \;/\; Z \geq z_{\frac{\alpha}{2}}\}$
$\mu \geq \mu_0$	$\{(x_1, \dots, x_n) \;/\; Z \leq z_{\frac{\alpha}{2}}\}$

Cuadro IV.3: *Regiones de rechazo para la media de cualquier distribución*

IV.3. Contrastes para dos muestras

Supongamos que tenemos 2 muestras $X_1, ..., X_N$ y $Y_1, ..., Y_N$. Siendo μ_1 la esperanza de X y μ_2 la esperanza de Y.

Podemos plantear hipótesis del tipo

- $H_0 : \;\; \mu_1 = \mu_2$
- $H_0 : \;\; \mu_1 \leq \mu_2$
- $H_0 : \;\; \sigma_1 = \sigma_2$

Este último caso (si las varianzas son iguales) suele ser un requisito previo antes de plantearte contrastes como el segundo ejemplo.

Uno de los test más usuales es el de igualdad de medias para dos poblaciones **homocedásticas** , es decir, con $\sigma_1 = \sigma_2$.

Si

$$\left.\begin{array}{l} X \sim N(\mu_1, \sigma) \\ Y \sim N(\mu_2, \sigma) \end{array}\right\} \text{Independientes} \quad \begin{array}{l} \overline{X} - \mu_1 \sim N\left(0, \frac{\sigma}{\sqrt{n_1}}\right) \\ \overline{Y} - \mu_2 \sim N\left(0, \frac{\sigma}{\sqrt{n_2}}\right) \end{array}$$

Entonces:

$$\frac{(\overline{X} - \mu_1) - (\overline{Y} - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

Todo esto suponiendo que $\sigma_1 = \sigma_2$, desconociendo su valor real. Nos gustaría por tanto, tener en el estadístico un estimador de σ.

Con este razonamiento podemos deducir que la región de rechazo es:

Contraste de igualdad de medias Si

$$\left.\begin{array}{l} X \sim N(\mu_1, \sigma) \\ Y \sim N(\mu_2, \sigma) \end{array}\right\} \text{Independientes} \quad \begin{array}{l} X_1, ..., X_{n_1} \to \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{(n_1-1)} \\ Y_1, ..., Y_{n_1} \to \frac{(n_1-1)S_2^2}{\sigma_2^2} \sim \chi^2_{(n_2-1)} \end{array}$$

Para seguir con el contraste de igualdad de medias necesitamos definir la distribución **Fisher-Snedecor** con $n_1 \, y \, n_2$ grados de libertad. . La distribución se parece mucho a la χ^2, y su función de distribución se obtiene así:

$$Q_1 \sim \chi^2_{n_1} \, ; \, Q_2 \sim \chi^2_{n_2}$$
$$F \sim \frac{Q_1/n_1}{Q_2/n_2}$$

Volviendo al caso donde estábamos podemos definir un estadístico de esta manera:

$$\frac{\frac{(n_1-1)S_1^2}{\sigma_1^2(n_1-1)}}{\frac{(n_2-1)S_2^2}{\sigma_2^2(n_2-1)}} \sim F_{n_1-1,n_2-1}$$

Sigue una F de Fisher.

Simplificando y suponiendo cierta la hipótesis de homocedasticidad ($\sigma_1 = \sigma_2$) tenemos que $F = \dfrac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$.

Este es el estadístico del contraste para comparar varianzas de dos poblaciones normales. Si el valor nos queda en las colas de la distribución, rechazaremos la hipótesis de igualdad de varianzas.

Con este razonamiento podemos construir la región de rechazo, que es

$$R = \left\{ |\overline{x} - \overline{y}| > t_{n_1+n_2-2;\sigma/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

siendo

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

la **varianza combinada**.

Ejemplo: *Sean X, Y poblaciones de datos emparejados tal que $\mathbb{E}(X) = \mu_1$ y $\mathbb{E}(Y) = \mu_2$.*

¿Qué significa datos emparejados? Muestras tomadas ambas a los mismo individuos de la mezcla después de una medicina por ejemplo, siendo X la medida antes e Y después. Esto quiere decir que X, Y no son independientes.

*El **procedimiento estándar** para este tipo de casos es suponer que*

$$D = X - Y \sim N(\mu_d, \sigma)$$

Y ahora expresamos nuestra hipótesis en función de D, de la que sabemos que

$$\mathbb{E}(D) = \mu_d = \mu_1 - \mu_2$$

- *Si $H_0 : \mu_1 = \mu_2 \iff H_0 : \mu_d = 0$. La región de rechazo de esta hipótesis será*

$$R = \left\{ \frac{\left|\overline{d}\right|}{S_d/\sqrt{n}} > t_{n-1;\frac{\alpha}{2}} \right\}$$

- *Si $H_0 : \mu_1 \leq \mu_2 \iff H_0 : \mu_d \leq 0$*
- *Si $H_0 : \mu_1 \geq \mu_2 \iff H_0 : \mu_d \geq 0$*

En el apéndice encontramos un ejercicio realizado en $\mathcal{R}$ ¿¿¿¿?????

IV.4. Consistencia de tests. Tests insesgados y UMP

Sucesión consistente

Definición IV.3 **Sucesión consistente**. *Se dice que una sucesión de tests con un nivel prefijado α es consistente cuando*

$$\lim_{n\to\infty} \beta_n(\theta) = 1 \;\forall \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

Es decir, que la probabilidad de rechazar la hipótesis nula cuando es falsa, dada por la función de potencia (IV.1), tienda a uno con muestras suficientemente grandes.

Test insesgado

Definición IV.4 **Test insesgado**. *Se dice que un test es insesgado cuando*

$$\beta_n(\theta) \leq \alpha \;\forall \theta \in \Theta_0$$

es decir, cuando cumple la teoría de Neyman-Pearson (ver sección IV.1.1); y además

$$\beta_n(\theta) \geq \alpha \;\forall \theta \in \Theta_1$$

Test UMP

Definición IV.5 **Test UMP**. *Se dice que un test es uniformemente más potente (UMP) dentro de una clase $\mathcal{B}_{n,\alpha}$ de tests de nivel α basados en muestras de tamaño n cuando*

$$\beta_n(\theta) \geq \tilde{\beta}_n(\theta), \;\forall \theta \in \Theta_1$$

siendo $\tilde{\beta}_n$ la función de potencia de cualquier otro test de la clase $\mathcal{B}_{n,\alpha}$.

IV.4.1. Lema de Neyman-Pearson

Recordemos la función de verosimilitud, que medía lo verosímil que es el valor del parámetro θ a la vista de la muestra. Para comparar dos hipótesis simples $H_i : \theta = \theta_i$, calcularíamos la función de verosimilitud para esos dos valores y veríamos cuál es más probable. Extendiendo esta idea, llegamos al lema de Neyman-Pearson.

Teorema IV.1 (Lema de Neyman-Pearson). *Se considera el problema de hipótesis simple y alternativa simple, es decir, que*

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_1 &: \theta = \theta_1 \end{aligned}$$

Denotemos

$$f_n(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Dado $\alpha \in (0,1)$, *supongamos que la región de rechazo*

$$R^* = \left\{ (x_1, \dots, x_n \diagup \frac{f_n(x_1, \dots, x_n; \theta_1)}{f_n(x_1, \dots, x_n; \theta_0)} > k \right\}$$

verifica $P_{\theta_0}(R^*) = \alpha$. *Entonces*

$$\mathbb{P}_{\theta_1}\{R^*\} \geq \mathbb{P}_{\theta_1}\{R\}$$

siendo R *la región crítica de cualquier otro test tal que* $\mathbb{P}_{\theta_0}\{R\} \leq \alpha$.
En otras palabras, R^* *es el* ***test óptimo de nivel*** α *para el problema considerado.*

Demostración. Denotamos $\mathbf{x} = (x_1, \dots, x_n)$ para cortar.

Tenemos que probar que $\mathbb{P}_{\theta_1}\{R^*\} - \mathbb{P}_{\theta_1}\{R\}$ es mayor o igual que cero.

$$\mathbb{P}_{\theta_1}\{R^*\} - \mathbb{P}_{\theta_1}\{R\} = \int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_1)\,\mathrm{d}\mathbf{x} - \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_1)\,\mathrm{d}\mathbf{x}$$

Por definición de R^*

$$\int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_1)\,\mathrm{d}\mathbf{x} \geq k \int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}\mathbf{x}$$

y también

$$\int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_1)\,\mathrm{d}\mathbf{x} \geq k \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}\mathbf{x}$$

Por lo tanto,

$$\begin{aligned} \mathbb{P}_{\theta_1}\{R^*\} - \mathbb{P}_{\theta_1}\{R\} \geq k \left[\int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}\mathbf{x} - \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}\mathbf{x} \right] &= \\ = k \left[\int_{R^*} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}\mathbf{x} - \int_{R} f_n(\mathbf{x}; \theta_0)\,\mathrm{d}x \right] &= \\ = k \left[\mathbb{P}_{\theta_0}\{R^*\} - \mathbb{P}_{\theta_0}\{R\} \right] \geq 0 & \end{aligned}$$

$\square$

IV.4.2. Familias paramétricas con cociente de verosimilitudes monótono y tests óptimos

En la subsección anterior hemos construido tests óptimos en problemas de hipótesis simple y alternativa simple. Pasaremos ahora a definirlos en modelos más complejos.

Familia paramétrica CVM

Definición IV.6 **Familia paramétrica CVM**. *Se dice que $f(\cdot|\theta)$ es una familia paramétrica con* ***cociente de verosimilitudes monótono*** *(CVM) si existe un estadístico $T_n(x_1,\dots,x_n)$ tal que, para todo θ_1,θ_2 con $\theta_1<\theta_2$ la razón de verosimilitudes*

$$\frac{f_n(x_1,\dots,x_n;\theta_2)}{f_n(x_1,\dots,x_n;\theta_1)}$$

es una función monótona no decreciente de $T_n(x_1,\dots,x_n)$.

Podemos ver algunos ejemplos de este tipo de familias.

Distribución exponencial Tomemos $X\sim\exp(\theta)$ con $\theta>0$ y $f(x;\theta)=\theta e^{-\theta x}$ para $x>0$. El cociente de las dos funciones es

$$\frac{\theta_2^n e^{-\theta_2\sum x_i}}{\theta_1^n e^{-\theta_1\sum x_i}}=\left(\frac{\theta_2}{\theta_1}\right)^n e^{(\theta_1-\theta_2)\sum x_i}$$

con $\theta_1-\theta_2<0$. Entonces, si consideramos

$$T_n(x_1,\dots,x_n)=\frac{1}{\sum x_i}\ \text{ó}\ T_n(x_1,\dots,x_n)=-\sum x_i$$

Tenemos tenemos un estimador monótonamente creciente y

$$\left(\frac{\theta_2}{\theta_1}\right)^n e^{(\theta_1-\theta_2)\sum x_i}=\left(\frac{\theta_2}{\theta_1}\right)^n e^{(\theta_1-\theta_2)\frac{1}{T}}$$

Teorema IV.2. *Supongamos que $F(\cdot;\theta)$ cumpla la propiedad CVM (cociente de verosimilitudes monótono) y que k_α es tal que:*

$$P_{\theta_0}\{t_n>k_\alpha\}=\alpha$$

Además suponemos que $P_{\theta_0}\{T_n=c\}=0,\ \forall\theta,c.$

Entonces:

$$R=\{(x_1,\dots,x_n):T_n(x_1,\dots,x_n)>k_\alpha\}$$

es la región crítica de un ***test óptimo***[1] ***de nivel*** *α para contrastar*

$$H_0:\theta\le\theta_0$$
$$H_1:\theta>\theta_0.$$

Vamos a ver otro ejemplo:

Ejemplo: *Ya hemos visto que la exponencial tiene CVM (cociente verosimilitudes monótono).*

Por el teorema tenemos que el ***test óptimo*** *de nivel α para $H_0 : \theta \leq \theta_0 ; H_1 : \theta > \theta_0$.*

Podemos construir la región de rechazo

$$R = \{(x_1, \dots, x_n) : \frac{1}{\sum_{i=1}^n x_i} > k_\alpha\}$$

donde

$$P_{\theta_0} \left\{ \sum_{i=1}^n X_i < \frac{1}{k_\alpha} \right\} = \alpha$$

Ejemplo: *Sea $f(\cdot;\theta)$ una uniforme en $(0,\theta)$. Se deja como ejercicio para el lector la comprobación de que la propiedad de CVM y la obtención del estadístico (que es el máximo de la muestra)*

IV.4.3. Construcción de tests. Test de cociente de verosimilitudes

Estadístico del contraste de razón de verosimilitudes

Definición IV.7 **Estadístico del contraste de razón de verosimilitudes.**

Sea $f(*;\theta)$ donde $\theta = (\theta_1, \dots, \theta_n) \in \Theta \subseteq \mathbb{R}^k$, siendo Θ un intervalo $\mathbb{R}^k$. Dada una muestra $x = (x_1, \dots, x_n)$, sea

$$f_n(x;\theta) = \prod_{i=1}^n f(x_1;\theta)$$

Consideremos el problema de contrastar a nivel α:

$$H_0 : \theta_i = c_i \text{ para } i = 1, \dots, r \leq k$$
$$H_1 : \theta_1 \neq c_i \text{ para algún } i = 1, \dots, r.$$

El estadístico del ***contraste de razón de verosimilitudes*** *es*

$$\Lambda_n = \frac{\sup_{\theta \in \Theta_0} f_n(x;\theta)}{\sup_{\theta \in \Theta} f_n(x;\theta)} = \frac{\sup_{\theta \in \Theta_0} f_n(x;\theta)}{f_n(x;\hat{\theta})}$$

donde $\hat{\theta}$ es el e.m.v. (III.3) de θ, y

[1]Uniformemente más potente (UMP)

- Si H_0 es cierta y el verdadero valor de θ están en Θ_0 entonces $\Lambda_n \to 1$, porque $\hat{\theta}_n$ tiende al verdadero valor del parámetro.
- Si H_0 es falsa, el e.m.v.(θ) tiende a un valor fuera de Θ_0. Entonces Λ_n tomará un valor significativamente menor que 1.

De esta forma, podemos construir una región de rechazo

$$R = \{(x_1, \dots, x_n \;/\; \Lambda_n(x_1, \dots, x_n) < k_\alpha \;\}$$

Hallar k_α según la probabilidad de error que queramos es algo complejo. Por eso nos apoyamos en el siguiente teorema:

Teorema IV.3. *Supongamos que*

1. *El e.m.v. $\hat{\theta}_n$ es estimador consistente en probabilidad del parámetro θ.*
2. *Para todo x, la función $\log f(x;\theta)$ tiene derivadas parciales terceras respecto a los componentes de θ contínuas.*
3. *En las integrales que involucran a la función $f(x;\theta)$ se pueden permutar las derivadas con el signo integral.*
4. *La matriz de información de Fisher*

$$\mathcal{I}(\theta) = \left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X;\theta)\right)_{1\le i,j\le k}$$

es invertible para cada θ.

Entonces, bajo H_0,

$$-2\log\Lambda_n \xrightarrow[n\to\infty]{d} \chi^2_r$$

IV.4.3.1. Aplicación a tests de bondad de ajuste

Sea X una v.a. discreta que toma los valores $a_1, \dots a_k$. Denotemos $p_i = \mathbb{P}\{X = a_i\}$. Supongamos que se desea contrastar

$$H_0 : \; p_i = p_{i0} \; i = 1, \dots, k$$

basado en una muestra $x_1, \dots, x_n$. Obsérvese que, en este caso, con la notación del teorema, $r = k - 1$ porque cuando se fijan $k-1$ probabilidades p_i, queda fijada la probabilidad restante. Por tanto, se rechaza H_0 al nivel α cuando

$$-2\log\Lambda_n > \chi^2_{k-1;\alpha}$$

Consideramos $f(x_1, \dots, x_n; p_1, \dots, p_k)$ como la probabilidad de haber observado la muestra $x_1, \dots, x_n$ con los valores de los parámetros $p_1, \dots, p_k$.
Entonces el numerador de Λ_n es

$$\frac{n!}{O_1! \cdots O_k!} p_{10}^{O_1} \cdots p_{k0}^{O_k}$$

siendo $O_j = \left| i \;/\; x_i = a_j \right|$ las *frecuencias observadas* de los distintos valores de la variable. Nótese que, bajo H_0, $(O_1, \dots, O_k)$ tiene distribución multinomial $\mathcal{M}(n : p_{10}, \dots, p_{k0})$.

En el denominador tenemos que poner los e.m.v. de cada p, de la siguiente forma

$$\hat{p}_k = \frac{o_k}{n}$$

y por lo tanto el denominador queda

$$\frac{n!}{O_1! \cdots O_k!} \left(\frac{O_1}{n}\right)^{O_1} \cdots \left(\frac{O_k}{n}\right)^{O_k}$$

Sustituyendo en Λ_n es inmediato ver que que el estadístico de contraste se puede expresar en la forma

$$-2 \log \Lambda_n \; = 2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{e_i}\right)$$

donde $e_i = np_{i0}$ $i = 1, \dots, k$ son las "*frecuencias esperadas (bajo H_0)*" de los distintos valores de la variable en una muestra de tamaño n.

Ejemplo: Experimento de Mendel *Un ejemplo clásico de este tipo de ajuste se puede ver en el experimento de Mendel, en el que se cruzaron plantas de guisantes con fenotipo rugoso-amarillo con otras de fenotipo liso-verde. En la segunda generación se podían observar cuatro fenotipos cuyas respectivas probabilidades, según la teoría de la herencia mendeliana, debían ser*

$$p_{10} = \frac{9}{16}, \, p_{20} = \frac{3}{16}, \, p_{30} = \frac{3}{16}, \, p_{40} = \frac{1}{16}$$

Observados $n = 556$ guisantes en la segunda generación del experimento, se obtuvieron los siguientes números de guisantes con estos fenotipos:

$$0_1 = 315, \, O_2 = 101, \, O_3 = 108, \, O_4 = 32.$$

¿Proporcionan estos resultados alguna evidencia en contra de la teoría mendeliana?

Aplicamos el test para contrastar $H_0 : p_1 = \frac{9}{16}, \dots, p_4 = \frac{1}{16}$.

$$e_1 = 556 \, \frac{9}{16} = 312.75, \;\; e_2 = e_3 = 556 \, \frac{3}{16} = 104.25, \;\; e_4 = 556 \, \frac{1}{16} = 34.75,$$

Obtenemos el estadístico

$$-2\log\Lambda_n = 2\sum_{i=1}^{k} O_i \log\left(\frac{O_i}{e_i}\right) = 0.4754$$

El p-valor, calculado a partir de la distribución χ^2_3, es 0.9281 lo que no indica ninguna evidencia estadística en contra de H_0.

Hay una controversia clásica en la historia de la ciencia en el sentido de que los resultados de Mendel eran "demasiado buenos", es decir, había demasiada concordancia entre las O_i y las e_i (por ejemplo, R.A. Fisher era de esta opinión; ver su artículo de 1936, "Has Mendel's work been rediscovered?", en The Annals of Science).

Se ha sugerido que este supuesto "exceso de concordancia" podría deberse a un "sesgo de repetición" (confirmation bias) producido por la repetición de los resultados hasta que las O_i concordasen fuertemente con las e_i. También se ha conjeturado que algún ayudante de Mendel pudo actuar con "exceso de celo" manipulando los resultados. En todo caso, las ideas básicas de Mendel eran acertadas y han tenido una influencia decisiva.

IV.4.4. Tests Bayesianos

Se desea contrastar

$$H_0 : \theta \in \Theta_0 \text{ frente a } H_1 : \theta \in \Theta \setminus \Theta_0$$

Obteniendo la información de una muestra $x_1, \ldots, x_n$.

La metodología bayesiana supone que la densidad que ha generado los datos es $f(\cdot|\theta)$ y que el parámetro θ puede considerarse como una v.a. con distribución a priori $\pi(\theta)$. A partir de aquí, se calcula la distribución a posteriori $\pi(\theta|x_1, \ldots, x_n)$ dada por

$$\pi(\theta|x_1, \ldots, x_n) = \frac{f_n(x_1, \ldots, x_n|\theta)\pi(\theta)}{\int_\Theta f_n(x_1, \ldots, x_n|\theta)\pi(\theta)d\theta}, \text{ donde}$$

$$f_n(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

El elemento fundamental en la inferencia bayesiana es siempre la distribución a posteriori. A partir de ella se pueden calcular las probabilidades a posteriori de ambas hipótesis:

$$\mathbb{P}\left\{\theta \in \Theta_0|x_1, \ldots, x_n\right\} = \pi(H_0|x_1, \ldots, x_n) = \int_{\Theta_0} \pi(\theta|x_1, \ldots, x_n)d\theta,$$
$$\mathbb{P}\left\{\theta \in \Theta_1|x_1, \ldots, x_n\right\} = \pi(H_1|x_1, \ldots, x_n) = 1 - \pi(H_0|x_1, \ldots, x_n)$$

y se toma la decisión en función de sus valores. Típicamente, se optará por H_1 cuando

$$\pi(H_1|x_1, \ldots, x_n) \geq \beta, \ \beta \in (0, 1)$$

β es un valor que se fija dependiendo de la gravedad que se atribuya al error de tipo I (IV.1).

Observación: la metodología bayesiana de contraste de hipótesis depende fuertemente de la elección de la distribución a priori π.

Apéndice A

Anexos

A.1. Condiciones suficientes para permutar la derivada con la integral

Sea una función $p(x, \theta)$ con $x \in \mathbb{R}$ y $\theta \in \mathbb{T}$ donde $\mathbb{T}$ es un intervalo abierto de los reales. Supongamos que

1. $p(x, \theta)$ es integrable con respecto a x para cada θ (se cumple automáticamente si p es función de densidad.
2. Para casi todo punto[1] existe $\dfrac{\partial}{\partial\theta}p(x, \theta)\ \forall\theta$.
3. Existe una función integrable $g : \mathbb{R} \longmapsto \mathbb{R}$ tal que

$$\left|\frac{\partial}{\partial\theta}p(x, \theta)\right| \leq g(x)\ \forall\theta$$

Entonces para todo θ

$$\frac{\partial}{\partial\,\theta}\int_{\mathbb{R}} p(x, \theta)\,dx = \int_{\mathbb{R}} \frac{\partial}{\partial\,\theta}p(x, \theta)\,dx$$

[1]Para todo x salvo los que tienen probabilidad 0

A.2. Distribuciones notables

La distribución normal

Función de densidad:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ x \in \mathbb{R}, \ \mu \in \mathbb{R}, \sigma > 0$$

Momentos:

$$\mathbb{E}(X) = \mu, \ \mathbb{V}(X) = \sigma^2$$

Aplicaciones: Es un modelo muy habitual para la distribución de magnitudes (en física, genética, etc.) que se pueden considerar como la suma de muchos pequeños efectos independientes (TCL). En Estadística aparece como distribución límite de muchos estadísticos que se usan para la inferencia.

La distribución exponencial

Función de densidad:

$$f(x) = \theta e^{-\theta x} \mathbb{I}_{[0,\infty)}(x) \ (\theta > 0)$$

Momentos:

$$\mathbb{E}(X) = \frac{1}{\theta}, \ \mathbb{V}(X) = \frac{1}{\theta^2}$$

Aplicaciones: en modelos de fiabilidad (tiempo de espera hasta que se produce una avería en un sistema).
Una propiedad interesante ("falta de memoria"): Si X sigue una distribución exponencial de parámetro θ, se tiene para $a > 0$ y $x > 0$,

$$\mathbb{P}\{X > x + a | X > x\} = e^{-\theta a}$$

(no depende de x).

La distribución gamma

Función de densidad:

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1} \mathbb{I}_{[0,\infty)}(x), \ (a > 0, \ p > 0),$$

donde $\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx$. Esta función verifica $\Gamma(p) = (p-1)!$ cuando $p \in \mathbb{N}$ y $\Gamma(p+1) = p\Gamma(p)$
Momentos:

$$\mathbb{E}(X) = \frac{p}{a}, \ \mathbb{V}(X) = \frac{p}{a^2}$$

Aplicaciones: Cuando $p \in \mathbb{N}$ se llama distribución de Erlang y se usa en problemas de fiabilidad (tiempo de espera hasta p fallos), cantidad de lluvia caída, cuantía de las reclamaciones a las compañías de seguro, modelos de supervivencia,.... Para $a = 1/2$ $p = n/2$, con $n \in \mathbb{N}$, se llama distribución χ^2 con n grados de libertad y desempeña un importante papel en Estadística.

La distribución uniforme

Función de densidad:

$$f(x; a, \theta) = \frac{1}{\theta - a}\mathbb{I}_{[a,\theta]}(x), \ (a, \theta \in \mathbb{R}, a < \theta)$$

Momentos:

$$\mathbb{E}(X) = \frac{\theta + a}{2}, \ \mathbb{V}(X) = \frac{(\theta - a)^2}{12}$$

Aplicaciones:
La uniforme se relaciona con otras distribuciones a través de la siguiente propiedad: si X es v.a. con f. de dist. F continua, entonces $Y = F(X)$ tiene distribución uniforme estándar (i.e. con $a = 0$, $\theta = 1$). Esta propiedad se utiliza en los métodos de generación de números (pseudo-)aleatorios: se generan números de una v.a. Y uniforme estándar y se transforman con F^{-1} para obtener observaciones aleatorias con la distribución F.

La distribución beta

Función de densidad:

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbb{I}_{[0,1]}(x),$$

siendo $a, b > 0$ y Γ la función gamma que aparece en la definición de la distribución del mismo nombre.
Momentos: $\mathbb{E}(X) = \frac{a}{a+b}$, $\mathbb{V}(X) = \frac{ab}{(a+b+1)(a+b)^2}$.
Aplicaciones: Dependiendo de los valores de los parámetros la densidad beta adopta formas muy variadas. Esta distribución (o sus versiones "reescaladas" en otros intervalos diferentes a [0,1]) proporciona un modelo muy flexible para describir variables aleatorias reales de soporte compacto.

La distribución de Weibull

Función de densidad:

$$f(x; \theta, k) = \frac{k}{\theta}\left(\frac{x}{\theta}\right)^{k-1} e^{-(x/\theta)^k}\mathbb{I}_{[0,\infty)}(x), \ (k > 0, \ \theta > 0)$$

Momentos:

$$\mathbb{E}(X) = \theta\Gamma\left(1 + \frac{1}{k}\right), \ \mathbb{V}(X) = \theta^2\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right]$$

Aplicaciones:
Tiempos de supervivencia, problemas de fiabilidad en ingeniería, distribuciones de velocidad del viento en ingeniería, de periodos de incubación de algunas enfermedades, etc.

La distribución de Pareto

Función de densidad:

$$f(x;a,\theta) = \theta \frac{a^\theta}{x^{\theta+1}} \mathbb{I}_{[a,\infty)}(x), \ (a > 0, \theta > 1)$$

Momentos:

$$\mathbb{E}_\theta(X) = \frac{\theta a}{\theta - 1}, \ \mathbb{V}_\theta(X) = \left(\frac{a}{\theta-1}\right)^2 \frac{\theta}{\theta-2}, \text{ si } \theta > 2$$

Aplicaciones:
Distribución de ingresos, de reservas de petróleo, de área quemadas en bosques, de tamaños de ficheros enviados por e-mail, de tamaños de partículas,...

La distribución de Cauchy

Función de densidad:

$$f(x;\theta,a) = \frac{1}{\pi a \left[1 + \left(\frac{x-\theta}{a}\right)^2\right]}$$

Momentos: No tiene momentos finitos
Aplicaciones: En el estudio de emisiones de partículas. Si Z es un ángulo aleatorio distribuido uniformemente entre $-\pi/2$ y $\pi/2$, $tang(Z)$ tiene distribución de Cauchy. El cociente de dos v.a. normales estándar independientes tiene también distribución de Cauchy.

La distribución lognormal

Función de densidad:

$$f(x;m,a) = \frac{1}{xa\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x - m}{a}\right)^2} \mathbb{I}_{[0,\infty)}(x), \ (m \in \mathbb{R}, a > 0)$$

Momentos:

$$\mathbb{E}(X) = e^{m+\frac{1}{2}a^2}, \ \mathbb{V}(X) = (e^{a^2} - 1)e^{2m+a^2}$$

Aplicaciones: Si X tiene distribución lognormal, $\log X$ tiene distribución normal. Se usa en geología (tamaño de rocas sedimentarias) y en general en aquellos casos en los que una variable puede considerarse producto de muchos factores de pequeño efecto individual.

La distribución de Bernoulli

Función de probabilidad (o "de masa"): Se dice que una v.a. X tiene distribución de Bernoulli de parámetro $p \in [0,1]$ (y se denota $X \sim B(1,p)$ o bien $X \sim Be(p)$) si

$$\mathbb{P}(X=1)=p,\ \mathbb{P}(X=0)=1-p.$$

Momentos:

$$\mathbb{E}(X)=p,\ \mathbb{V}(X)=p(1-p)$$

Aplicaciones: Experimentos aleatorios "binarios", i.e. con sólo dos posibles resultados.

La distribución binomial

Función de probabilidad: Se dice que una v.a. X tiene distribución binomial de parámetro $p \in [0,1]$ (y se denota $X \sim B(n,p)$) si

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\ k=0,1,\ldots,n$$

Momentos:

$$\mathbb{E}(X)=np,\ \mathbb{V}(X)=np(1-p)$$

Aplicaciones: Número de éxitos en n pruebas de Bernoulli independientes en cada una de las cuales la probabilidad de éxito es p. La suma de n v.a. independientes con distribución $B(1,p)$ es $B(n,p)$.

La distribución de Poisson

Función de probabilidad: Se dice que una v.a. X tiene distribución de Poisson de parámetro $\lambda > 0$ (y se denota $X \sim \mathcal{P}(\lambda)$) si

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

Momentos:

$$\mathbb{E}(X)=\lambda,\ \mathbb{V}(X)=\lambda$$

Aplicaciones: Frecuentemente se utiliza como modelo probabilístico para el estudio de fenómenos como el número de "sucesos" (tales como llegadas de clientes a un servicio, llamadas telefónicas a una centralita, accidentes,...) que se producen en un periodo de tiempo prefijado. Aparece como límite de la binomial en el siguiente sentido: Si $X_n \sim B(n,p_n)$ y $np_n \to \lambda > 0$, entonces

$$\lim_{n\to\infty}\mathbb{P}(X_n=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ k=0,1,2,\ldots$$

La distribución binomial negativa

Función de probabilidad: Se dice que una v.a. X tiene distribución binomial negativa de parámetros $p \in [0,1]$ y $r \in \mathbb{N}$ (y se denota $X \sim BN(r,p)$) si

$$\mathbb{P}(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r},\ k = r, r+1, r+2, \dots$$

Momentos:

$$\mathbb{E}(X) = \frac{r}{p},\ \mathbb{V}(X) = r\frac{1-p}{p^2}$$

Aplicaciones: Es un modelo discreto de "tiempo de espera": En una sucesión de experimentos de Bernoulli con probabilidad éxito p, la distribución del número de pruebas necesarias para obtener r éxitos es $BN(r,p)$. La distribución $BN(1,p)$ se denomina geométrica.

A.3. Regiones de rechazo

CONTRASTES DE HIPÓTESIS

NOTACION:
α= nivel de significación del contraste.
n= tamaño de la muestra.
H_0= hipótesis nula.
R= región crítica o de rechazo de H_0.

1) $X \sim N(\mu, \sigma)$.

$H_0 : \mu = \mu_0$ (σ desconocida); $\quad R = \left\{ |\bar{x} - \mu_0| > t_{n-1;\alpha/2} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ desconocida); $\quad R = \left\{ \bar{x} - \mu_0 > t_{n-1;\alpha} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ desconocida); $\quad R = \left\{ \bar{x} - \mu_0 < t_{n-1;1-\alpha} \frac{s}{\sqrt{n}} \right\}$ $(t_{n-1;1-\alpha} = -t_{n-1;\alpha})$

$H_0 : \sigma = \sigma_0; \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 \notin \left[\chi^2_{n-1;1-\alpha/2} \ , \ \chi^2_{n-1;\alpha/2} \right] \right\}$

$H_0 : \sigma \leq \sigma_0; \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 > \chi^2_{n-1;\alpha} \right\}$

$H_0 : \sigma \geq \sigma_0; \quad R = \left\{ \frac{n-1}{\sigma_0^2} s^2 < \chi^2_{n-1;1-\alpha} \right\}$

2) $X \sim B(1, p)$ (muestras grandes)

$H_0 : p = p_0; \quad R = \left\{ |\bar{x} - p_0| > z_{\alpha/2} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \leq p_0; \quad R = \left\{ \bar{x} - p_0 > z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$

$H_0 : p \geq p_0; \quad R = \left\{ \bar{x} - p_0 < z_{1-\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$ $(z_{1-\alpha} = -z_{\alpha})$

3) Contrastes para la media de una población no necesariamente normal (muestras grandes)

$H_0 : \mu = \mu_0$ (σ desconocida); $\quad R = \left\{ |\bar{x} - \mu_0| > z_{\alpha/2} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \leq \mu_0$ (σ desconocida); $\quad R = \left\{ \bar{x} - \mu_0 > z_{\alpha} \frac{s}{\sqrt{n}} \right\}$

$H_0 : \mu \geq \mu_0$ (σ desconocida); $\quad R = \left\{ \bar{x} - \mu_0 < z_{1-\alpha} \frac{s}{\sqrt{n}} \right\}$ $(z_{1-\alpha} = -z_{\alpha})$

4) Dos poblaciones normales independientes.

$X \sim N(\mu_1, \sigma_1)$; $(X_1, \ldots, X_{n_1})$ m. a. de X; se calcula $\bar{x}$ y s_1^2.
$Y \sim N(\mu_2, \sigma_2)$; $(Y_1, \ldots, Y_{n_2})$ m. a. de Y; se calcula $\bar{y}$ y s_2^2.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$H_0 : \mu_1 = \mu_2$ $(\sigma_1 = \sigma_2)$; $\quad R = \left\{ |\bar{x} - \bar{y}| > t_{n_1+n_2-2;\alpha/2} \ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

$H_0 : \mu_1 = \mu_2$ $(\sigma_1 \neq \sigma_2)$; $\quad R = \left\{ |\bar{x} - \bar{y}| > t_{f;\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$

$H_0 : \mu_1 \leq \mu_2$ $(\sigma_1 = \sigma_2)$; $\quad R = \left\{ \bar{x} - \bar{y} > t_{n_1+n_2-2;\alpha} \ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$

$H_0 : \mu_1 \leq \mu_2$ $(\sigma_1 \neq \sigma_2)$; $\quad R = \left\{ \bar{x} - \bar{y} > t_{f;\alpha} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\}$

$$H_0: \mu_1 \geq \mu_2 \ (\sigma_1 = \sigma_2); \quad R = \left\{ \bar{x} - \bar{y} < t_{n_1+n_2-2;1-\alpha} \ s_p \sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}} \right\}$$

$$H_0: \mu_1 \geq \mu_2 \ (\sigma_1 \neq \sigma_2); \quad R = \left\{ \bar{x} - \bar{y} < t_{f;1-\alpha} \sqrt{\tfrac{s_1^2}{n_1} + \tfrac{s_2^2}{n_2}} \right\}$$

$$H_0: \sigma_1 = \sigma_2; \quad R = \left\{ s_1^2/s_2^2 \notin \left[F_{n_1-1;n_2-1;1-\alpha/2} \ , \ F_{n_1-1;n_2-1;\alpha/2} \right] \right\}$$

$$H_0: \sigma_1 \leq \sigma_2; \quad R = \left\{ s_1^2/s_2^2 > F_{n_1-1;n_2-1;\alpha} \right\}$$

$$H_0: \sigma_1 \geq \sigma_2; \quad R = \left\{ s_1^2/s_2^2 < F_{n_1-1;n_2-1;1-\alpha} \right\}$$

$$\text{donde } f = \text{entero más próximo a } \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

5) Comparación de proporciones (muestras grandes e independientes).

$X \sim B(1, p_1)$; $(X_1, \ldots X_{n_1})$ m. a. de X.
$Y \sim B(1, p_2)$; $(Y_1, \ldots Y_{n_2})$ m. a. de Y.

$$H_0: p_1 = p_2; \quad R = \left\{ |\bar{x} - \bar{y}| > z_{\alpha/2} \sqrt{\bar{p}(1-\bar{p}) \left(\tfrac{1}{n_1} + \tfrac{1}{n_2} \right)} \right\}$$

$$H_0: p_1 \leq p_2; \quad R = \left\{ \bar{x} - \bar{y} > z_{\alpha} \sqrt{\bar{p}(1-\bar{p}) \left(\tfrac{1}{n_1} + \tfrac{1}{n_2} \right)} \right\}$$

$$H_0: p_1 \geq p_2; \quad R = \left\{ \bar{x} - \bar{y} < z_{1-\alpha} \sqrt{\bar{p}(1-\bar{p}) \left(\tfrac{1}{n_1} + \tfrac{1}{n_2} \right)} \right\}$$

$$\text{donde } \ \bar{p} = \frac{\sum x_i + \sum y_i}{n_1 + n_2} = \frac{n_1 \bar{x} + n_2 \bar{y}}{n_1 + n_2}$$

Apéndice B

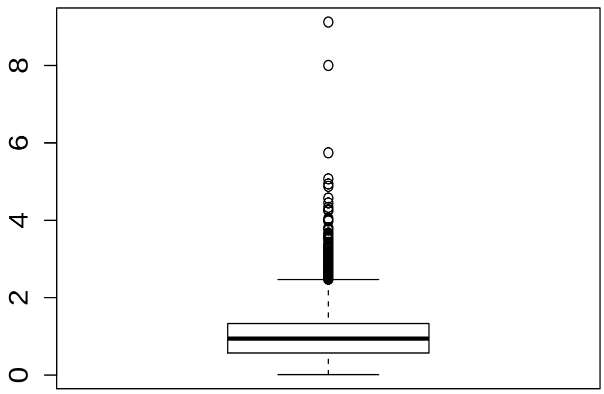
Ejercicios

B.1. Tema 1 - Estadística descriptiva

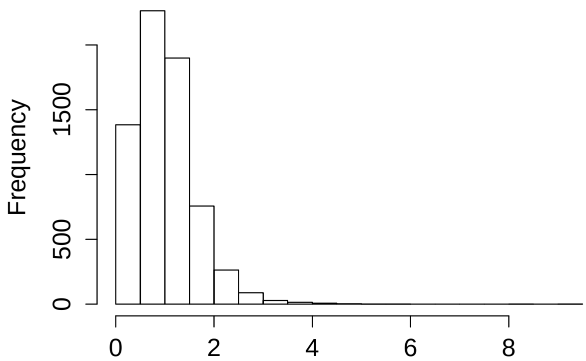
Ejercicio 1.1:

Realizar un análisis descriptivo de los datos británicos de ingresos familiares contenidos en el fichero `Datos-ingresos.txt`. En concreto, calcular los estadísticos de tendencia central, las medidas de dispersión y representar un diagrama de cajas y un estimador kernel de la función de densidad. Comentar los resultados.

```
> x = scan('Datos-ingresos.txt')
Read 6711 items
> mean(x)
[1] 1.022779
> median(x)
[1] 0.9417
> var(x)
[1] 0.3657983
> sd(x)
[1] 0.6048126
> boxplot(x)
```

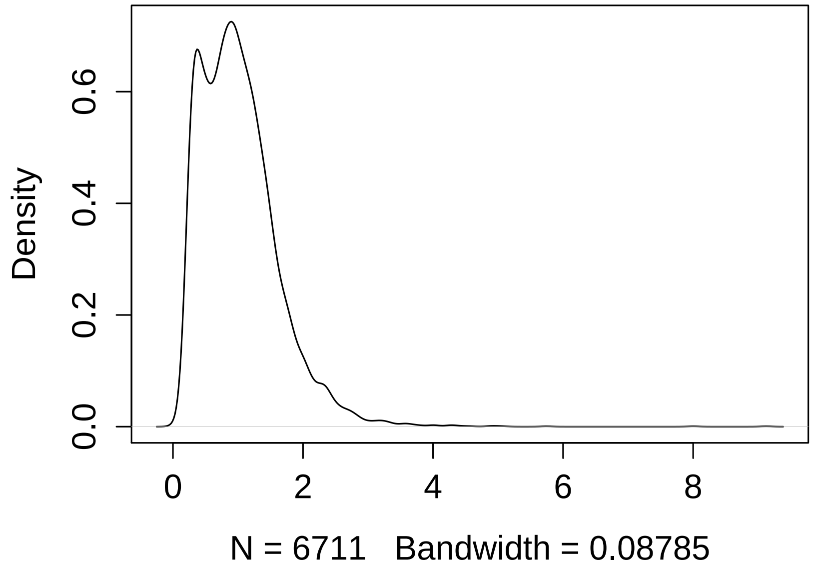



```
> hist(x)
```



1

```
> plot(density(x,kernel='gaussian'))
```




```
> sum(x>2)/length(x)
[1] 0.0606467
> skewness(x)
[1] 1.797857
```

Ejercicio 1.2: *Demostrar que*

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \min_{a \in \mathbb{R}} \sum_{i=1}^{n}(x_i - a)^2$$

Definimos una función

$$g(a) = \sum_{i=1}^{n}(x_i - a)^2$$

buscamos su derivada

$$g'(a) = -2\sum_{i=1}^{n}(x_i - a)$$

e igualamos a cero:

$$\begin{aligned} -2\sum_{i=1}^{n}(x_i - a) &= 0 \\ \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} a &= 0 \\ n\overline{x} &= na \\ \overline{x} &= a \end{aligned}$$

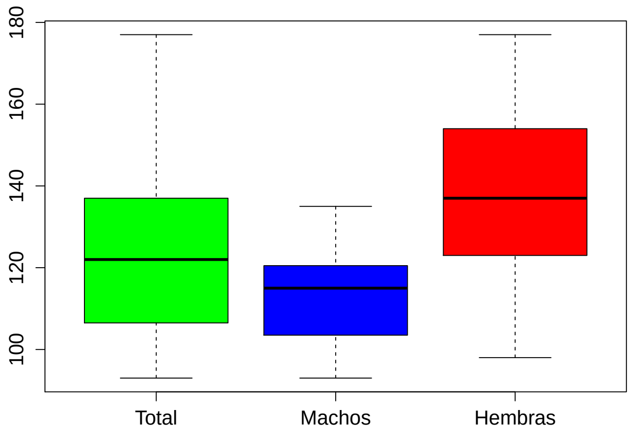
Esto quiere decir que la media muestral es el valor que minimiza la distancia con cada uno de los datos de la muestra.

Ejercicio 1.3:

Representar, en un mismo gráfico, los diagramas de cajas correspondientes a la variable "Largo" del fichero **tortugas.txt** para el conjunto de datos global, para los ejemplares hembra y para los ejemplares macho. Es decir, el gráfico debe incluir tres diagramas de cajas, de izquierda a derecha: el primero debe corresponder a la variable global (sin distinción de sexos), el segundo al subconjunto de los datos correspondiente a las hembras y el tercero al correspondiente a los machos. Emplear colores distintos para los tres diagramas.

Solución:

```
X = read.table("tortugas.txt",header=T)
boplot(X$Largo,X$Largo[X$Sexo==1],X$Largo[X$Sexo==0],
       names=cbind("Total","Machos","Hembras"),col=cbind("green","blue","red"))
```



Ejercicio 1.4:

Los datos del fichero `Datos-kevlar.txt` corresponden al tiempo hasta el fallo (en horas) de 101 barras de un material utilizado en los transbordadores espaciales, llamado Kevlar49/epoxy, sometidas a un cierto nivel de esfuerzo. Los datos han sido tomados de Barlow *et al.* (1984).

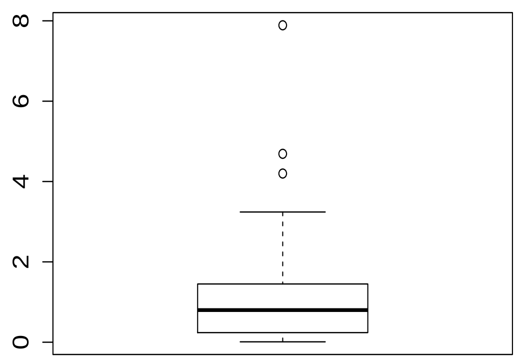
(a) Calcula las principales medidas numéricas descriptivas de estos datos.

(b) Representa un diagrama de cajas.

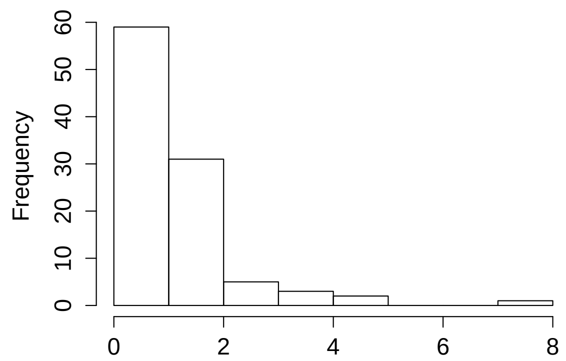
(c) Representa un histograma con un número de clases apropiado.

(d) Estudia la presencia de datos atípicos en la muestra. Si hay datos atípicos, suprímelos y repite todos los apartados anteriores. Compara los resultados obtenidos.

```
> x = scan('Datos-kevlar.txt')
Read 101 items
> mean(x)
[1] 1.024018
> median(x)
[1] 0.799838
> var(x)
[1] 1.248112
> sd(x)
[1] 1.117189
> skewness(x)
[1] 3.009575
> boxplot(x)
```

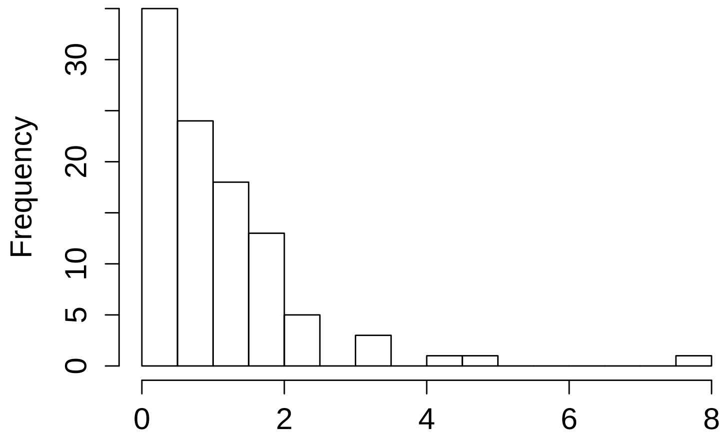



```
> hist(x)
```

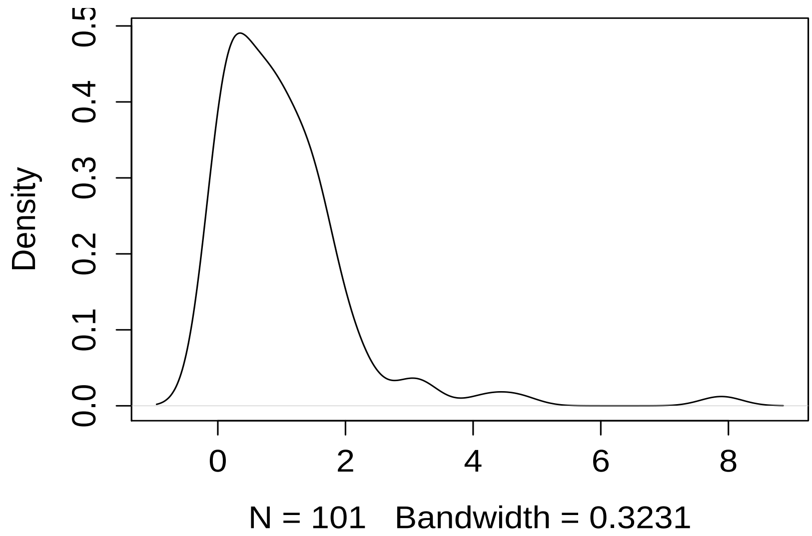


1

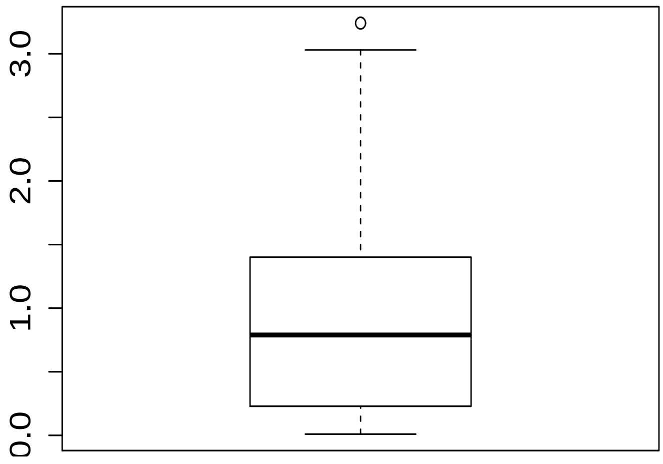
```
> hist(x)$breaks
[1] 0 1 2 3 4 5 6 7 8
> n=length(x)
> sqrt(n)
[1] 10.04988
> n=length(x)
> sqrt(n)
[1] 10.04988
> (max(x)-min(x))/sqrt(n)
[1] 0.7840221
> max(x)
[1] 7.889078
> min(x)
[1] 0.00975351
> hist(x,breaks=seq(0,8,0.5))
```



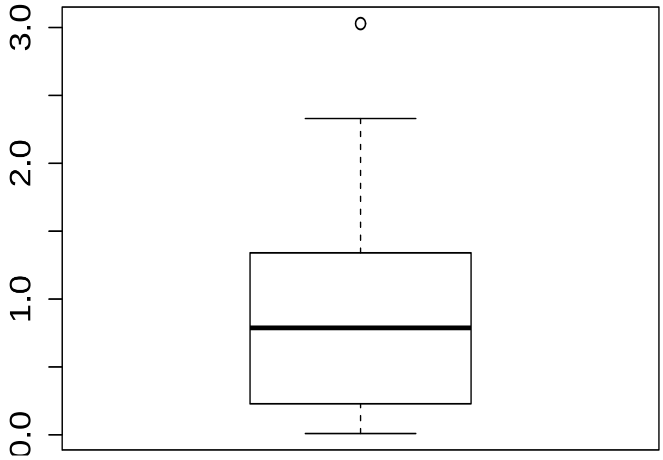

```
> plot(density(x,kernel='gaussian'))
```



```
> xOrd=sort(x)
> xOrdSin=xOrd[1:(n-3)]
> mean(xOrdSin)
[1] 0.8841606
> median(xOrdSin)
[1] 0.7889238
> var(xOrdSin)
[1] 0.5386131
> boxplot(xOrdSin)
```




```
> skewness(xOrdSin)
[1] 0.9158652
> xOrdSin=xOrd[1:(n-4)]
> boxplot(xOrdSin)
```



Ejercicio 1.5: *Determina si es verdadero o falso:*

a) *Si añadimos 7 a todos los datos de un conjunto, el primer cuartil aumenta en 7 unidades y el rango intercuartílico no cambia.*

b) *Si todos los datos de un conjunto se multiplican por -2, la desviación típica se dobla.*

c) *Si todos los datos de un conjunto se multiplican por 2, la varianza se dobla.*

d) *Al multiplicar por tres todos los datos de un conjunto, el coeficiente de asimetría no varía*

e) *Si el coeficiente de correlación entre dos variables vale -0.8, los valores por debajo del promedio de una variable están asociados con valores por debajo del promedio de la otra.*

f) *Si $\forall i \; y_i < x_i$ entonces el coeficiente de correlación es negativo.*

g) *Si cambiamos el signo de todos los datos de un conjunto, el coeficiente de asimetría también cambia de signo.*

h) *Al restar una unidad a cada dato de un conjunto, la desviación típica siempre disminuye.*

i) *Si a un conjunto de datos con media $\overline{x}$ se le añade un nuevo dato que coincide con $\overline{x}$, la media no cambia y la desviación típica disminuye.*

Apartado a)

Falso. Añadir siete a todos los datos es una traslación, así que la distribución de los datos no cambia. El rango intercuartílico se mantiene y el cuantil también.

Apartado b)

Teniendo en cuenta que si multiplicamos todos los datos del conjunto por -2 la media también se multiplica por -2, y sustituyendo en la fórmula de la varianza:

$$\sigma' = \sqrt{\frac{1}{n}\sum_{i=1} n(-2x_i)^2 - (-2\overline{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1} 4\left(nx_i^2 - \overline{x}^2\right)} = \sqrt{4\sigma^2} = 2\sigma$$

Por lo tanto, la desviación típica sí se dobla.

Apartado c)

Usando los cálculos del apartado anterior vemos que la varianza se multiplica por cuatro.

Apartado d)

Efectivamente: cambiar el signo haría una reflexión de los datos sobre el eje Y y la asimetría estaría orientada hacia el lado contrario.

Apartado e)

Teniendo en cuenta que si multiplicamos todos los datos del conjunto por 3 la media también se multiplica por 3

El coeficiente de asimetría se calcula:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3$$

Sustituyendo en la fórmula del coeficiente de asimetría

$$\frac{1}{n}\sum_{i=1}^{n}(3x_i - 3\overline{x})^3 = \frac{1}{n}\sum_{i=1}^{n}3^3(x_i - \overline{x})^3 = 27 \cdot \frac{1}{n}\sum_{i=1}^{n}(x - \overline{x})^3$$

Por lo tanto el coeficiente de asimetría sí varía.

Apartado f)

Falso.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \overline{y})^2 = \begin{cases} y_j = x_j - 1 \\ \overline{y} = \frac{1}{n}\sum_{j=1}^{n}(x_j - 1) = \frac{1}{n}(\sum_{j=1}^{n} x_j) - 1 = \overline{x} - 1 \end{cases}$$

$$= \frac{1}{n}\sum_{j=1}^{n}(x_j - 1 - (\overline{x} - 1))^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2 = \sigma^2$$

Apartado g)

Falso. 2 variables pueden tener una correlación creciente aunque $y_i < x_i$.

Apartado h)

Falso. La desviación típica se mantiene (los datos siguen estando "igual de separados").

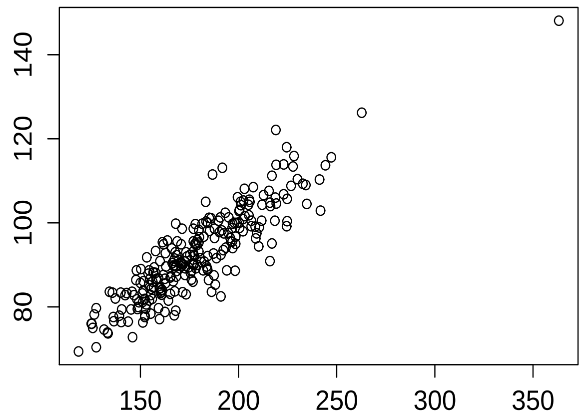
Apartado i)

Verdadero. Al hacer el cálculo de la media no varía (en la fórmula del ejercicio 2 se puede comprobar que si añadimos un $x_i = \overline{x}$ el sumatorio de la derecha queda igual) y la desviación típica disminuye.

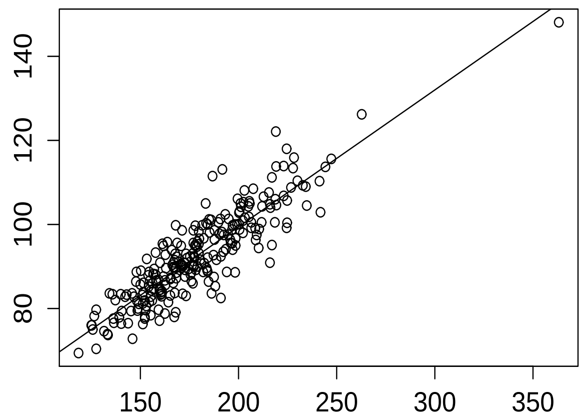
Ejercicio 1.6:

Calcula el diagrama de dispersión de las dos variables correspondientes al peso y a la circunferencia de abdomen que aparecen en el fichero `Datos-bodyfat.txt`. Calcula la recta de regresión y el coeficiente de correlación. Comenta los resultados.

```
> Datos=read.table('Datos-bodyfat.txt')
> Peso=Datos[,4]
> CircAbd=Datos[,8]
> plot(Peso,CircAbd)
```

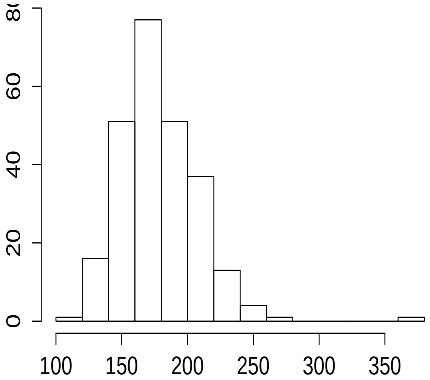


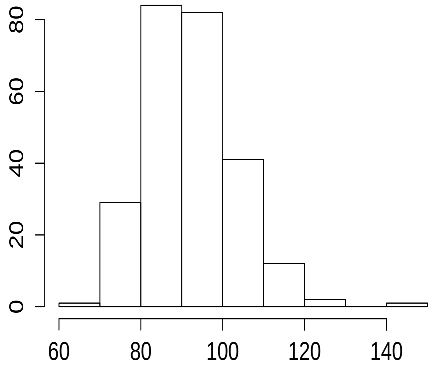

```
> lm(CircAbd~Peso)

Call:
lm(formula = CircAbd ~ Peso)

Coefficients:
(Intercept)        Peso
    34.2604      0.3258
> cor(Peso,CircAbd)
[1] 0.8879949
> zz=abline(lm(CircAbd~Peso))
```



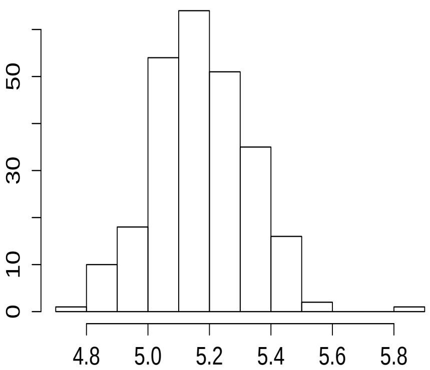
1

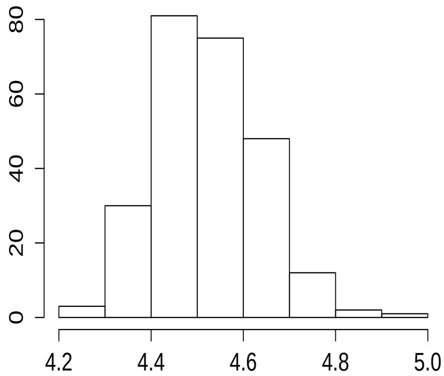
```
> hist(Peso)
> hist(CircAbd)
```



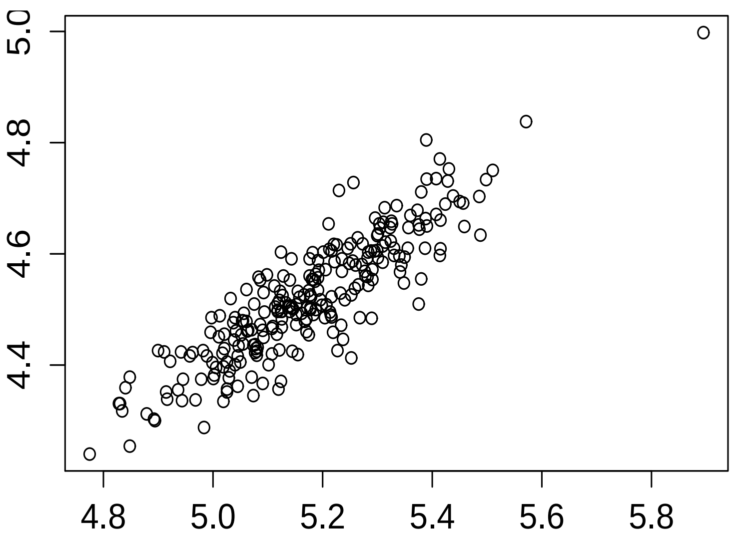



```
> hist(log(Peso))
> hist(log(CircAbd))
```



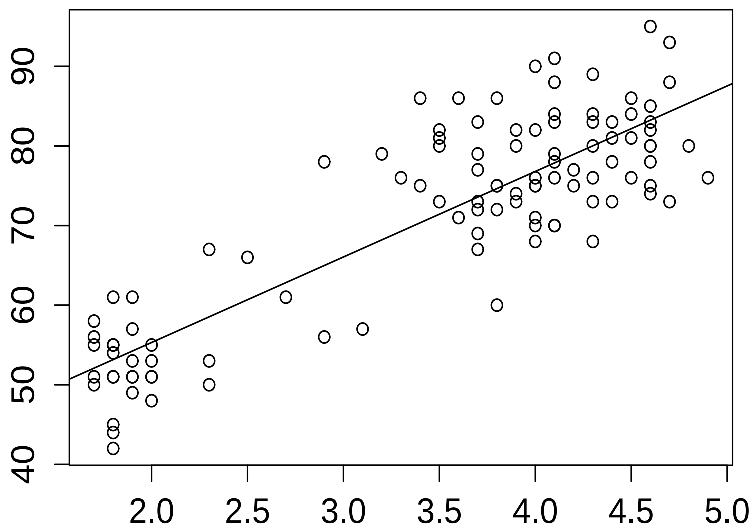



```
> skewness(Peso)
[1] 1.198077
> skewness(log(Peso))
[1] 0.317743
> skewness(log(CircAbd))
[1] 0.3548225
plot(log(Peso),log(CircAbd))
```



Análogas preguntas para las dos variables del fichero `Datos-geyser.txt`.

```
> Datos=read.table('Datos-geyser.txt',header=T)
> y=Datos[,2]
> x=Datos[,3]
> plot(x,y)
> zz=lm(y~x)
> abline(zz)
> zz

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
      33.83        10.74
> cor(x,y)
[1] 0.8584273
```



Ejercicio 1.7: *Relaciona los histogramas con los boxplot*

Fijándose en los intervalos entre los que se mueven los datos es la forma más fácil.

$$1 \to 2$$

$$2 \to 1$$

$$3 \to 3$$

Ejercicio 1.8: *Del diagrama de dispersión presentado se pregunta:*

a) *¿Existe alguna relación?*

b) *¿Hay algún dato atípico?*

c) *De los 3 valores siguientes:* $0.01, 0.83, -0,73$ *¿cuál crees que podría corresponder al coeficiente de correlación?*

Apartado a)

Parece que sí.

Apartado b)

Bastante obvio que sí

Apartado c)

0.83. Como la nube de puntos parece que se aproxima a una recta con pendiente positiva, la correlación debe ser positiva. Además, como "se parece bastante" a una recta, la correlación debe ser cercana a 1.

Ejercicio 1.9:

Un estudio sobre el efecto de la temperatura en el rendimiento de un proceso químico proporciona los siguientes resultados:

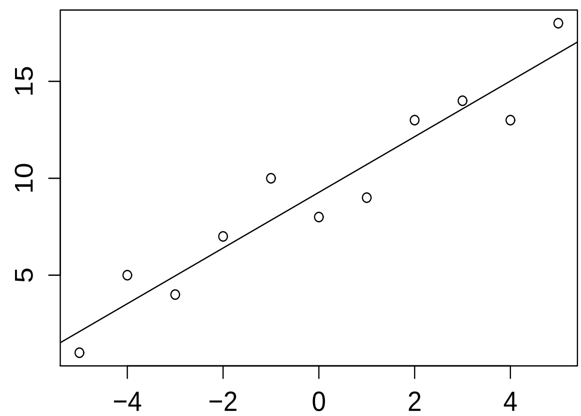
Temperatura (x)	-5	-4	-3	-2	-1	0	1	2	3	4	5
Rendimiento (y)	1	5	4	7	10	8	9	13	14	13	18

(a) Representa el diagrama de dispersión de los datos anteriores y calcula el coeficiente de correlación entre las dos variables. ¿Se puede admitir que existe una relación lineal aproximada entre ambas, es decir, $y_i \sim a + bx_i$?

(b) Calcula el término independiente y la pendiente de la recta de mínimos cuadrados.

(c) ¿Qué rendimiento predecirías para un nuevo proceso realizado a temperatura $x = 3{,}5$?

```
# Temperatura:
x = -5:5
# Rendimiento:
y = c(1,5,4,7,10,8,9,13,14,13,18)

# Diagrama de dispersion
plot(x,y)

# Coeficiente de correlacion
cor(x,y)

# Recta de regresion:
zz = lm(y~x)
abline(zz)

# Prediccion para temperatura x=3.5:
new <- data.frame(x = 3.5)
Prediccion = predict.lm(zz,new)
```



$\hat{y} = 9{,}27 + 1{,}44x \qquad r = 0{,}956 \qquad \hat{y}(3{,}5) = 9{,}27 + 1{,}44 \cdot 3{,}5 = 14{,}30$

Ejercicio 1.10: *¿Qué valor tiene que tomar x para que el coeficiente de correlación sea 1?*

a) $A = \{(1,1),(2,3),(2,3),(4,x)\}$

b) $B = \{(1,1),(2,3),(3,4),(4,x)\}$

Para que el coeficiente de correlación sea exactamente 1, los puntos tienen que estar en la misma recta. Buscamos el x que cumpla eso.

Apartado a)

$x = 6$

Apartado b)

Imposible (porque los 3 puntos dados no están alineados)

B.2. Tema 2 - Muestreo aleatorio

Ejercicio 2.1: *Se desea estimar el momento de orden 4, $\alpha_3 = \mathbb{E}\left(X^3\right)$ en una v.a. X con distribución exponencial de parámetro 2, es decir, la función de distribución de X es $F(t) = \mathbb{P}\{X \leq t\} = 1 - e^{-2t}$ para $t \geq 0$. Definir un estimador natural para α_3 y calcular su error cuadrático medio.*

Usando el criterio de *plugin*, podríamos definir el estimador

$$\hat{\alpha}_3 = \int_{\mathbb{R}} x^3 \, d\mathbb{F}_n(x)$$

Calculamos ahora el error cuadrático medio:

$$\text{ECM}(\hat{\alpha}_3) = \mathbb{E}\left((\hat{\alpha}_3 - \alpha_3)^2\right) = \mathbb{E}\left((\hat{\alpha}_3 - \mathbb{E}(\hat{\alpha}_3) + \mathbb{E}(\hat{\alpha}_3) - \alpha_3)^2\right) =$$

$$\mathbb{E}\left(\underbrace{(\hat{\alpha}_3 - \mathbb{E}(\hat{\alpha}_3))^2}_{(a)} + \underbrace{(\mathbb{E}(\hat{\alpha}_3) - \alpha_3)^2}_{(b)} + \underbrace{2 \cdot (\hat{\alpha}_3 - \mathbb{E}(\hat{\alpha}_3)) \cdot (\mathbb{E}(\hat{\alpha}_3) - \alpha_3)}_{(c)}\right)$$

Calculamos (b) que es el $\text{sesgo}^2(\hat{\alpha}_3)$:

$$\text{sesgo}(\hat{\alpha}_3) = \mathbb{E}(\hat{\alpha}_3) - \alpha_3 = \alpha_3 - \alpha_3 = 0$$

Como el sesgo es 0, tenemos que (c) es también 0.
Solo nos queda calcular $\mathbb{E}\left((a)\right)$, que es la varianza:

$$\mathbb{V}(\hat{\alpha}_3) = \mathbb{V}\left(\frac{1}{n}\sum X_i^3\right) = \frac{1}{n^2}\mathbb{V}\left(\sum X_i^3\right) = \frac{1}{n^2}\sum \mathbb{V}\left(X_i^3\right) = \frac{\mathbb{V}\left(X^3\right)}{n}$$

y, teniendo en cuenta el enunciado,

$$\mathbb{V}\left(X^3\right) = \mathbb{E}\left(X^6\right) - \mathbb{E}\left(X^3\right)^2 = \frac{6!}{2^6} - \left(\frac{3!}{2^3}\right)^2 = \frac{171}{16}$$

y por lo tanto

$$\text{ECM}(\hat{\alpha}_3) = \frac{171}{16n} = O\left(\frac{1}{n}\right) \xrightarrow[n\to\infty]{} 0$$

donde lo que más nos importa es la convergencia a cero, que indica que cuanto más muestras tenemos mejor será el estimador.

Ejercicio 2.2: *Supongamos que la muestra tiene tamaño $n = 50$ y que la distribución de las X_i es una $N(4, 1)$.*

a) *Obtener, utilizando la desigualdad de Chebichev, una cota superior para la probabilidad* $\mathbb{P}\left\{\left|\overline{X} - 4\right| > 0.3\right\}$.

b) *Calcula exactamente* $\mathbb{P}\left\{\left|\overline{X} - 4\right| > 0.3\right\}$ *utilizando la distribución de X_i.*

Apartado a)

Como $\mu = 4$, la desigualdad de Chebichev nos da una cota de

$$\mathbb{P}\left\{\left|\overline{X} - 4\right| > 0.3\right\} \leq \frac{\mathbb{V}\left(\overline{X}\right)}{0.3^2} = \frac{\mathbb{V}(X)}{n \cdot 0.3^2} \simeq 0.22$$

Apartado b)

Normalizamos

$$Z = \frac{\overline{X} - 4}{\frac{1}{\sqrt{50}}} \sim N(0, 1)$$

y calculamos.

$$\mathbb{P}\left\{\left|\overline{X} - 4\right| > 0.3\right\} = \mathbb{P}\left\{|Z| > \frac{0.3}{\frac{1}{\sqrt{50}}}\right\} = 2 \cdot \mathbb{P}\left\{Z > 2.12\right\} = 0.034$$

Ejercicio 2.3: *Utilizando R dibuja la función de densidad y la función de distribución de una v.a. con distribución beta de parámetros $a = 3$, $b = 6$.*
A continuación dibuja, sobrepuestas en cada uno de los gráficos, las aproximaciones a F y f obtenidas respectivamente mediante la función empírica y un estimador kernel.
Verificar empíricamente el grado de aproximación, en las estimaciones de F y f, que se obtiene mediante un experimento de simulación basado en 200 muestras de tamaño 20. Es decir, considerando, por ejemplo, la estimación de F, se trata de simular 200 muestras de tamaño 20; para cada una de ellas evaluar el error (medido en la norma del supremo) que se comete al aproximar F por $\mathbb{P}_n$. Por último, calcular el promedio de los 200 errores obtenidos. Análogamente para la estimación de f.

```
# Dibuja la función de densidad y la funcion de distribución de una v.a. con
# distribución beta de parametros a=3 y b=6.
```

```
a = 3
b = 6
n = 20
t = seq(0,1,0.01)
densidad = dbeta(t, a, b, ncp = 0, log = FALSE)
fndistrib = pbeta(t, a, b, ncp = 0, log = FALSE)
X = rbeta(n,a, b, ncp = 0)
kernelest = density(X,kernel="gaussian")
M = max(max(densidad),max(kernelest$y))
distremp = ecdf(X)

layout(matrix(1:2,2,1))
layout.show(2)

plot(t,densidad,type="l",lwd=2,col="tomato3",xlab="",ylab="",ylim=c(0,M),
     main="Densidad y estimador kernel",font.main=1,cex.main=1)
lines(kernelest,type="l",lwd=2,col="navyblue")
mtext("Distribución beta(3,6)",side=3,line=3,cex=1.5)
plot(t,fndistrib,type="l",lwd=2,col="tomato3",xlab="",ylab="",ylim=c(0,1),
     main="Función de distribución poblacional y empírica",font.main=1,cex.main=1)
lines(distremp,do.points=FALSE,lwd=2,col="navyblue")

# Verificar empiricamente el grado de aproximacion:
nMC = 200
Supremo1 = rep(0,nMC) ; Supremo2 = rep(0,nMC)
for (i in 1:nMC){
  XMC = rbeta(n,a, b, ncp = 0)
  kernelMC = density(XMC,kernel="gaussian")
  densidadMC = dbeta(kernelMC$x, a, b, ncp = 0, log = FALSE)
  Supremo1[i] = max(abs(kernelMC$y - densidadMC))
  distempMC = ecdf(XMC)
  Supremo2[i] = max(abs(distempMC(t) - fndistrib))
}
Error1 = mean(Supremo1)
Error2 = mean(Supremo2)
```

Ejercicio 2.4: *Denotemos por*

$$C_n = \int_{\mathbb{R}} \left(\mathbb{F}_n(t) - F(t)\right)^2 \, dF(t)$$

la llamada discrepancia de Cramer-Von Mises entre $\mathbb{F}_n$ *y* F.

a) *¿Converge a cero casi seguro esta discrepancia?*

b) *Calcular la distribución asintótica de la sucesión* $D_n = \sqrt{n}\left(\mathbb{F}_n(t) - F(t)\right)$ *para un valor fijo* $t \in \mathbb{R}$.

Apartado a)

$$C_n = \int_{\mathbb{R}} \left(\mathbb{F}_n(t) - F(t)\right)^2 \, dF(t) = \int_{\mathbb{R}} \left(\mathbb{F}_n(t) - F(t)\right)^2 f(t) \, dt$$

Tenemos que

$$\mathbb{F}_n(t) - F(t) \leq \sup_t \left|\mathbb{F}_n(t) - F(t)\right| = \|\mathbb{F}_n - F\|_\infty$$

entonces

$$\int_{\mathbb{R}} \left(\mathbb{F}_n(t) - F(t)\right)^2 f(t) \, dt \leq \|\mathbb{F}_n - F\|_\infty^2 \int_{\mathbb{R}} f(t) \, dt = \|\mathbb{F}_n - F\|_\infty^2$$

Finalmente, por el teorema de Glivenko-Cantelli (II.7) tenemos que

$$\|\mathbb{F}_n - F\|_\infty^2 \xrightarrow[n\to\infty]{c.s} 0 \quad \square$$

Apartado b)

Para calcular la distribución asintótica de

$$D_n = \sqrt{n}\left(\mathbb{F}_n(t) - F(t)\right)$$

usamos el Teorema Central del Límite (II.8). Necesitamos algo que se asemeje a una media muestral, y de hecho

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,t]}(X_i) = \frac{1}{n}\sum_{i=1}^n Y_i = \overline{Y}$$

Por otra parte, $Y = \mathbb{1}_{(-\infty,t]}(X)$ y por lo tanto

$$\mathbb{E}\left(Y\right) = \mathbb{E}\left(\mathbb{1}_{(-\infty,t]}(X)\right) = \mathbb{P}\left\{X \leq t\right\} = F(t)$$

Ya podemos aplicar el TCL, pero nos falta saber cuál es la desviación típica de Y. Como es una distribución de Bernoulli

$$\mathbb{V}(Y) = p(1-p) = F(t)(1 - F(t))$$

y por lo tanto aplicando el TCL

$$D_n \xrightarrow[n\to\infty]{d} N\left(0, \sqrt{F(t)(1-F(t))}\right)$$

Ejercicio 2.5: *Sea X una v.a. cuya función de densidad depende de un parámetro desconocido $\theta \in \mathbb{R}$, concretamente*

$$f(x;\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}$$

para $x \in \mathbb{R}$. Comprobar que θ coincide con la mediana y la moda de X pero que la media $\mathbb{E}(X)$ no está definida.

Diseñar un experimento de simulación en R, tomando algún valor concreto de θ, orientado a comprobar cómo se comportan la mediana muestral y la media muestral como estimadores de θ: mientras la mediana muestral se acerca al verdadero valor de θ al aumentar N, la media muestral oscila fuertemente y no se acerca a θ aunque se aumente el tamaño muestral n.

Viendo la función, vemos que es simétrica con respecto al eje $x = \theta$. Por lo tanto, el punto que deja a izquierda y derecha la misma probablidad, la mediana, es precisamente θ.

La moda es el valor máximo de la distribución,

$$f'(x;\theta) = \frac{1}{\pi}\frac{-2(x-\theta)}{(1+(x-\theta)^2)^2} = 0 \iff x = \theta$$

Y se ve que es un máximo porque es el punto en el que el signo de la derivada pasa de positivo a negativo.

Ejercicio 2.6: *Se extrae una muestra aleatoria de tamaño $n = 600$ de una v.a. cuya desviación típica es $\sigma = 3$. Calcular aproximadamente la probabilidad*

$$\mathbb{P}\left\{\left|\overline{X} - \mu\right| < 0.1\right\}$$

Tenemos 2 posibilidades: Tipificar o con Chebichev.

Según Chebichev, tenemos que

$$\mathbb{P}\left\{\left|X - \mathbb{E}(X)\right| > \varepsilon\right\} \leq \frac{\mathbb{V}(X)}{\varepsilon^2}$$

Tenemos que $\mu = \mathbb{E}\left(\overline{X}\right)$, tenemos que hallar $\mathbb{V}\left(\overline{X}\right)$:

$$\mathbb{V}\left(\overline{X}\right) = \frac{\mathbb{V}(X)}{n} = \frac{\sigma^2}{n} = 0.015$$

Y por lo tanto,

$$\mathbb{P}\left\{\left|\overline{X}-\mu\right| < 0.1\right\} = 1 - \underbrace{\mathbb{P}\left\{\left|\overline{X}-\mu\right| > 0.1\right\}}_{\leq 1.5} \geq -0.5$$

Que no es una aproximación muy buena. Así que pasamos a tipificar:

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

Entonces:

$$\mathbb{P}\left\{\left|\overline{X}-\mu\right| < 0.1\right\} = 1 - \mathbb{P}\left\{\left|\overline{X}-\mu\right| > 0.1\right\} =$$
$$= 1 - \mathbb{P}\left\{|Z| > \frac{0.1\sqrt{n}}{\sigma}\right\} = 1 - 2\cdot\mathbb{P}\left\{Z > \frac{0.1\sqrt{n}}{\sigma}\right\} = 0.582$$

(Recordemos que $\mathbb{V}\left(\overline{X}\right) = \frac{\mathbb{V}(X)}{n}$ y que $\mathbb{E}\left(\overline{X}\right) = \mu = \mathbb{E}\left(X\right)$)

Ejercicio 2.7: *Sea X una v.a con distribución absolutamente continua. Sea F la correspondiente función de distribución y $f = F'$ continua en todo punto la función de densidad. Para $r \in \{1,\dots,n\}$, denotemos por $X_{(r)}$ el r-simo estadístico ordenado de una muestra de tamaño n extraída de X. Calcular la función de distribución y la de densidad de la v.a. $X_{(r)}$.*

Por definición, la función de distribución es:

$$F_{X_{(r)}}(x) = \mathbb{P}\left\{X_{(r)} \leq x\right\}$$

que es la probabilidad que al menos r elementos de la muestra sean menores o iguales que x. Luego la probabilidad es igual a

$$\sum_{j=r}^{n} \mathbb{P}\left\{\text{exactamente j observaciones de la muestra}^{1}\text{son} \leq \text{x}\right\} =$$
$$= \sum_{j=r}^{n} \mathbb{P}\left\{B(n,F(x)) = j\right\} = \sum_{j=r}^{n} \begin{pmatrix} n \\ j \end{pmatrix} F(x)^j \left(1-F(x)\right)^{n-j}$$

Ahora sólo falta calcular la densidad de $X_{(r)}$, y la obtenemos derivando la función de distribución:

[1]que una observación sea exactemente $\leq x$ es una Bernouilli, y la suma de Bernouillis es la Binomial

$$f_{X_{(r)}}(x) =$$

$$= \sum_{j=r}^{n} \left(\binom{n}{j} j(F(x)^{j-1}(1-F(x))^{n-j} f(x) - (F(x))^{j}(n-j)(1-F(x))^{n-j-1} f(x) \right) =$$

$$= \sum_{j=r}^{n} \binom{n}{j} j(F(x)^{j-1}(1-F(x))^{n-j} f(x) - \sum_{j=r}^{n} \binom{n}{j} (F(x))^{j}(n-j)(1-F(x))^{n-j-1} f(x) =$$

$$= \binom{n}{r} r(F(x))^{r-1}(1-F(x))^{n-1} f(x) + \sum_{j=r+1}^{n} \binom{n}{j} j(F(x))^{j-1} f(x)(1-F(x))^{n-j}$$

$$- \sum_{j=r}^{n} \binom{n}{j} (n-j)(F(x))^{j}(1-F(x))^{n-j-1} f(x) =$$

$$n \binom{n-1}{r-1} (F(x))^{r-1}(1-F(x))^{n-r} f(x) + \sum_{l=r}^{n-1} n \binom{n-1}{l} (F(x))^{l}(1-F(x))^{n-l-1} f(x)$$

$$- \sum_{j=r}^{n-1} n \binom{n-1}{j} (F(x))^{j}(1-F(x))^{n-j-1} f(x)$$

Los dos últimos términos se cancelan y nos queda que:

$$f_{X_{(r)}}(x) = n \binom{n-1}{r-1} (F(x))^{r-1}(1-F(x))^{n-r} f(x)$$

Consideremos los dos casos particulares del mínimo y máximo de la muestra. Con el mínimo, $r = 1$ y entonces:

$$F_{X_{(1)}}(x) = \mathbb{P}\left\{X_{(1)} \leq x\right\} = \sum_{j=1}^{n} \binom{n}{j} (F(x))^{j}(1-F(x))^{n-j} =^{[2]} 1 - (1-F(x))^{n}$$

En el caso del máximo:

$$F_{X_{(n)}}(x) = \mathbb{P}\left\{X_{(n)} \leq x\right\} = \prod_{j=1}^{n} \mathbb{P}\left\{X_{(j)} \leq x\right\} =^{[3]} (F(x))^{n}$$

[2] $1 = 1^n = (1 - F(x) + F(x))^n = \sum_{j=0}^{n} \binom{n}{j} (F(x))^{j}(1-F(x))^{n-j}$

[3] $\forall j \; X_{(j)} \sim X \implies \mathbb{P}\{X \leq x\} = F(x)$

Ejercicio 2.8: *Sea $\hat{f}_n$ un estimador kernel de la densidad basado en un núcleo K que es una función de densidad con media finita. Comprobar que, en general, $\hat{f}_n(t)$ es un estimador sesgado de $f(t)$ en el sentido de que **no** se tiene $\mathbb{E}\left(\hat{f}_n(t)\right) = f(t)$ para todo t y para toda densidad f.*

Lo que buscamos es calcular el sesgo:

$$\text{sesgo}\left(\hat{f}_n(t)\right) = \mathbb{E}\left(\hat{f}_n(t)\right) - f(t) \tag{B.1}$$

$$\mathbb{E}\left(\hat{f}_n(t)\right) = \mathbb{E}\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t-X_i}{h}\right)\right) = \frac{1}{nh}\sum_{i=1}^{n}\mathbb{E}\left(K\left(\frac{t-X_i}{h}\right)\right) =$$
$$= \frac{1}{h}\mathbb{E}\left(K\left(\frac{t-X}{h}\right)\right) = \frac{1}{h}\int_{\mathbb{R}} K\left(\frac{t-x}{h}\right) f(x)\,dx = ...$$

Haciendo un cambio de variable $x = t - hz$, $dx = -h\,dz$, los límites se invierten,

$$... = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{t-x}{h}\right) f(x)\,d(x) = \frac{1}{h}\int_{\infty}^{-\infty} K(z)f(t{-}hz)(-h)\,dz = \int_{-\infty}^{\infty} K(z)f(t{-}hz)\,dz$$

Usando que K es función de densidad $\implies \int K = 1$, (B.1) nos queda

$$... = \int_{-\infty}^{\infty} K(z)f(t-hz)\,dz - \int_{-\infty}^{\infty} K(z)f(t)\,dz = \int_{-\infty}^{\infty} K(z)\left[f(t-hz) - f(t)\right]\,dz =$$
$$= hf'(t)\int_{-\infty}^{\infty} zK(z)\,dz + \frac{1}{2}h^2 f''(t)\int_{-\infty}^{\infty} z^2K(z)\,dz + \frac{1}{6}h^3 f'''(t)\int_{-\infty}^{\infty} z^3K(z)\,dz + \cdots$$

Al hacer el desarrollo de Taylor, como K es una función simétrica, las integrales con índice impar (con $z = 1, 3, \dots$) se anulan. Sin embargo, el segundo término no lo hace. Por lo tanto, el sesgo de un estimador kernel **no es nunca cero**.

El sesgo del estimador kernel depende de h (el parámetro de suavizado o *bandwith*) en potencias pares. Por eso, se toma de manera tal que $h \xrightarrow[n\to\infty]{} 0$ y entonces sesgo $\hat{f}_n(t) \xrightarrow[n\to\infty]{} 0$ pero manteniendo un equilibrio para que la varianza también sea pequeña y no tengamos picos en el histograma (ver sección I.1.3.1).

B.3. Tema 3 - Estimación puntual paramétrica

Ejercicio 3.1: *Sea X una v.a. con distribución exponencial de parámetro θ. Calcula la función de distribución, la de densidad y la función cuantílica de la v.a. $Y = \theta X^{1/3}$.*

$$\text{[4]Como X } \sim exp(\theta)\,(\theta > 0) \wedge Y = \theta X^{1/3} \iff X = (\frac{Y}{\theta})^3$$

$$f(x) = f((\frac{y}{\theta})^3) = \theta e^{-\theta(\frac{y}{\theta})^3} = \theta e^{-\frac{y^3}{\theta^2}} = f(y)$$

$$F(X) = F((\frac{Y}{\theta})^3) = \int f((\frac{y}{\theta})^3)(3\frac{y^2}{\theta^3})dy =$$

$$= \int 3\frac{y^2}{\theta^2}e^{-\frac{y^3}{\theta^2}}dy = -e^{-\frac{y^3}{\theta^2}} + C = F(Y), C \in \mathbb{R}$$

Finalmente, como $e^{-\frac{y^3}{\theta^2}}$ es creciente con valor máximo 0 y mínimo -1 $\implies$ C = 1.

La función cuantílica por definición es: $F^{-1}(p) = inf\{y \,/\, F(y) \geq p\}$, luego

$$F^{-1}(p) = \inf_{p\in[0,1]}\{y \,/\, 1 - e^{-\frac{y^3}{\theta^2}} \geq p\}$$

Ejercicio 3.2: *Supongamos que X mide el error cometido en la medición de una magnitud. X es una v.a. normal de media 0 y varianza θ.*

$$X \rightsquigarrow N(0, \sqrt{\theta}), \theta > 0, \Theta = (0, \infty)$$

Se desea estimar θ a partir de una muestra.

a) *Calcular el estimador de máxima verosimilitud T_n.*

b) *Probar que T_n es insesgado y eficiente.*

c) *Estudiar la distribución asintótica de T_n.*

Apartado a)

Buscamos el máximo de la función de verosimilitud

$$L_n(\theta; X_1, ..., X_n) = \prod_{i=1}^{n} f(x_i; \theta) = \frac{1}{(\sqrt{2\pi}\,\theta)^{\frac{n}{2}}}e^{-\frac{1}{2\theta}\sum x_i^2}$$

El máximo de la función de verosimilitud será también el máximo de la logverosimilitud

[4] Esta solución puede estar mal. Edu

$$logL_n(\theta) = \frac{n}{2} \cdot log(2\pi) - \frac{n}{2} log(\theta) - \frac{1}{2\theta} \sum x_i^2$$

Para ello derivamos e igualamos a 0.

$$\frac{\partial}{\partial \theta} logL_n(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum x_i^2 = 0$$

$$\frac{1}{2} \left(-\frac{n}{\theta} + \frac{\sum x_i^2}{\theta^2} \right) = 0 \implies T_n = e.m.v.(\theta) = \frac{1}{n} \sum x_i^2$$

Apartado b)

$\mathbb{E}_\theta (T_n) = \mathbb{E}_\theta \left(\frac{1}{n} \sum x_i^2 \right) = \mathbb{E}_\theta \left(X^2 \right) = \theta$

Nos tenemos que dar cuenta de que $\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2$. En este caso $\mathbb{E}(X) = \mu = 0$ por lo que $\mathbb{E}\left(X^2\right) = \theta$ por hipótesis. Vamos a calcular la información de fisher para comprobar si el estimador es eficiente o no.

$$logf(x;\theta) = \frac{-1}{2} log(2\pi) - \frac{1}{2} log(\theta) - \frac{1}{2\theta} X^2$$

Derivamos:

$$\frac{\partial}{\partial \theta} logf(x;\theta) = -\frac{1}{2\theta} + \frac{1}{2\theta^2} X^2$$

Elegimos derivar otra vez o elevar al cuadrado (2 alternativas para calcularlo).

En este caso vamos a elevar al cuadrado:

$$\frac{\partial}{\partial \theta} logf(X;\theta) = \frac{1}{4\theta^2} \left(1 + \frac{X^4}{\theta^2} - 2\frac{X^2}{\theta} \right)$$

Entonces la información de fisher será:

$$I(\theta) = \mathbb{E}_\theta \left(\frac{1}{4\theta^2} \left(1 + \frac{X^4}{\theta^2} - 2\frac{X^2}{\theta} \right) \right) = \frac{1}{4\theta^2} \left(1 + \frac{\mathbb{E}_\theta \left(X^4\right)}{\theta^2} - 2\frac{\mathbb{E}_\theta \left(X^2\right)}{\theta} \right)$$

Aplicamos por hipótesis: $\mathbb{E}_\theta \left(X^4\right) = 3\theta^2$

$$I(\theta) = \frac{1}{4\theta^2} \left(1 + \frac{3\theta^2}{\theta^2} - 2\frac{\theta}{\theta} \right) = \frac{1}{2\theta^2}$$

Vamos a calcular

$$\mathbb{V}_\theta (T_n) = \mathbb{V}_\theta \left(\frac{1}{n} \sum x_i^2 \right) = \frac{1}{n^2} \sum \mathbb{V}_\theta \left(x_i^2\right) = \frac{n}{n^2} \mathbb{V}_\theta \left(X^2\right) =$$

$$\frac{1}{n}\left(\mathbb{E}_\theta\left(X^4\right) - \mathbb{E}_\theta\left(X^2\right)\right) = \frac{1}{n}(3\theta^2 - \theta^2) = \frac{2\theta^2}{n} = \frac{1}{nI(\theta)}$$

Como la varianza coincide con la cota de Frécher-Cramer-Rao entonces podemos decir que es un estimador eficiente.

Los siguientes pasos para comprobar lo bueno que es el estimador son:

- T_n asintóticamente normal.
- T_n es consistente casi seguro.

Apartado c)

Vamos a estudiar la distribución asintótica:

$$\sqrt{n}(T_n - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma(\theta))$$

Llamando $Y_i = X_i^2 \implies \mathbb{E}_\theta\left(Y\right) = \mathbb{E}_\theta\left(X^2\right) = \theta$

Entonces por el TCL (Teorema Central del Límite):

$$\sqrt{n}(\hat{Y} - \mathbb{E}_\theta\left(Y\right)) \xrightarrow[n\to\infty]{d} N(0, \sqrt{\mathbb{V}\left(Y\right)})$$

Donde $\mathbb{V}\left(Y\right) = \mathbb{V}_\theta\left(X^2\right) = \mathbb{E}\left((X^2)^2\right) - \mathbb{E}\left(X^2\right)^2 = 3\theta^2 - \theta^2 = 2\theta^2$

Ejercicio 3.3: *Se dispone de un gran lote de piezas producidas en una cadena de montaje. Denotemos por p la proporción de piezas defectuosas en ese lote. Supongamos que se seleccionan al azar sucesivamente (con reemplazamiento) piezas del lote hasta que se encuentra una defectuosa. Sea X la variable aleatoria que indica el número de la extracción en la que aparece la primera pieza defectuosa.*

a) *Calcular $\mathbb{P}\{X = k\}$ para $k = 1, 2, \ldots$ Obtener el estimador de p por el método de los momentos, a partir de una muestra $X_1, \ldots, X_n$.*

b) *Obtener el estimador de p por el método de máxima verosimilitud. Calcular su distribución asintótica.*

Apartado a)

La probabilidad sigue una distribución geométrica de parámetro p:

$$\mathbb{P}\{X = k\} = (1-p)^{k-1}p$$

Apartado b)

Calculamos la función de verosimilitud:

$$L_n(p; x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{\sum\limits_{i=1}^{n} x_i - 1} p^n$$

Tomamos logaritmos

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

y derivando

$$\frac{\partial}{\partial p} \log L_n(p) = \frac{-1}{1-p} \sum_{i=1}^{n} (x_i - 1) + \frac{n}{p} = 0 \iff$$

$$\frac{1-\hat{p}}{\hat{p}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - 1) \iff 1 - \hat{p} = \hat{p}(\overline{x} - 1) \iff \text{emv}(p) = \hat{p} = \frac{1}{\overline{x}}$$

Vamos a calcular su distribución asintótica, aplicando el método delta.

Para ello observamos que tomando $g(x) = \frac{1}{x}$, tenemos que $g(\overline{x}) = \hat{p}$.

Comprobamos que $g(\mathbb{E}(X)) = \frac{1}{\mathbb{E}(X)} = \frac{1}{\frac{1}{p}} = p$

Luego por el método delta y aplicando el TCL:

$$\sqrt{n}(g(\overline{X}) - g(\mathbb{E}(X))) = \sqrt{n}(\hat{p} - p) \xrightarrow[n\to\infty]{d} N\left(0, \left|g'(\mathbb{E}(X))\right| \sqrt{\mathbb{V}(X)}\right)$$

Como $g'(x) = \frac{-1}{x^2}$, y $\mathbb{V}(X) = \frac{1-p}{p^2}$, entonces

$$N\left(0, \left|g'(\mathbb{E}(X))\right| \sqrt{\mathbb{V}(X)}\right) = N\left(0, \frac{1}{\frac{1}{p^2}} \frac{\sqrt{1-p}}{p}\right) = N\left(0, p\sqrt{1-p}\right)$$

Ejercicio 3.4: *Estudiar si es eficiente el estimador de máxima verosimilitud de una poisson.*

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

El cálculo del estimador de máxima verosimilitud se hizo en clase llegando a $\lambda = \overline{x}$ (III.1.2.1).

Para ver si es eficiente vemos si es su varianza es igual a la cota de FCR. Necesitamos la información de Fisher para comprobar eso.

Para calcular la información de fisher derivamos el logaritmo de la densidad

$$\log f(\lambda; x) = -\lambda + x \log \lambda - \log x!$$

$$\frac{\partial}{\partial \lambda} \log f(\lambda; x) = -1 + \frac{x}{\lambda} + 0$$

Para calcular la información de Fisher podemos volver a derivar o elevar al cuadrado. Elegimos volver a derivar

$$\frac{\partial^2}{\partial^2 \lambda} \log f(\lambda; x) = -\frac{x}{\lambda^2}$$

Entonces tenemos que

$$I(\lambda) = \mathbb{E}\left(-\frac{\partial^2}{\partial^2 \lambda} \log f(\lambda; x)\right) = \mathbb{E}\left(\frac{x}{\lambda^2}\right) = \frac{1}{\lambda^2}\mathbb{E}(X) = \frac{1}{\lambda}$$

La cota de FCR será entonces $\frac{1}{n\frac{1}{\lambda}} = \frac{\lambda}{n}$.

Calculamos la varianza:

$$\mathbb{V}(\lambda) = \mathbb{V}(\overline{x}) = \frac{\mathbb{V}(x)}{n} = \frac{\lambda}{n}$$

Como tenemos la igualdad podemos afirmar que **si** es un estimador eficiente.

Ejercicio 3.5: *Distribución de Rayleigh, cuya función de densidad es:*

$$f(x; \theta) = \frac{x}{\theta^2} e^{\frac{-x^2}{2\theta^2}} \mathbb{I}_{[0,\infty)}(x), \theta > 0$$

a) *Calcular el estimador de máxima verosimilitud (e.m.v.) y por el método de los momentos*

b) *Calcular la consistencia del e.m.v.*

c) *¿Son asintóticamente normales ambos estimadores?*

Apartado a)

$$L_n(\theta; x_1, ..., x_n) = \frac{x_1 \cdot ... \cdot x_n}{\theta^2} e^{\frac{-1}{2\theta^2} \sum_{i=1}^{n} x_i^2}$$

$$logL_n(\theta) = \sum logx_i - 2nlog\theta - \frac{1}{2\theta^2} \sum x_i^2$$

$$\frac{\partial\, l}{\partial\, o} gL_n(\theta) = \frac{1}{\theta}\left(-2n + \frac{1}{\theta^2} \sum x_i^2\right) = 0$$

$$\implies \hat{\theta}^2 = \frac{\sum x_i^2}{2n} \implies \hat{\theta} = emv(\theta) = \left(\frac{\sum x_i^2}{2n}\right)^{\frac{1}{2}}$$

Estimador razonable porque $\mathbb{E}\left(x^2\right) = \mathbb{V}\left(x\right) + \mathbb{E}\left(x\right)^2 = \left(\sqrt{\frac{\pi}{2}}\theta\right)^2 + \frac{4-\pi}{2}\theta^2 = 2\theta^2 \iff \theta^2 = \frac{1}{2}E(x^2)$

Buscamos ahora el estimador $\tilde{\theta}$ por el **método de los momentos**

$$\mathbb{E}_\theta\left(X\right) = \theta\sqrt{\frac{\pi}{2}} = \overline{X}$$

y entonces el estimador es

$$\tilde{\theta} = \overline{X}\sqrt{\frac{2}{\pi}}$$

Apartado b)

Consistencia: $\hat{\theta}^2 = \frac{1}{2}\overline{Y}, Y_i = X_i^2$

Por la ley fuerte de los grandes números (II.6) sabemos que:

$$\overline{Y} \xrightarrow[n\to\infty]{cs} E_\theta(Y) = E_\theta(X^2) = 2\theta^2$$

Vamos a aplicar el teorema de Slutsky.

Sea $g(x) = \sqrt{\frac{1}{2}x}$ definida sobre $[0,\infty)$.

Teorema de Slutsky (II.3) $\implies g\left(\overline{Y}\right) = \sqrt{\frac{1}{2}\frac{\sum x_i^2}{n}} \xrightarrow[n\to\infty]{c.s} g(E_\theta) = \sqrt{\frac{1}{2}\theta^2} = \theta \implies$ El e.m.v. de θ, $\hat{\theta}$ es consistente c.s.

Apartado c)

Queremos aplicar el método delta:

$$\sqrt{n}(\hat{\theta}-\theta) = \sqrt{n}\left(g\left(\overline{Y}\right) - g\left(E(Y)\right)\right) \xrightarrow[n\to\infty]{d} N(0, \left|g'(E(Y))\right|\sqrt{V(Y)}$$

$$E_\theta(Y) = E_\theta(X^2) = 2\theta^2$$
$$V_\theta(Y) = E(X^4) - E^2(X^2) = 8\theta^4 - 4\theta^4 = 4\theta^4$$

Entonces tenemos que $g'(E(Y)) = \dfrac{1}{2\sqrt{2E(Y)}} = \dfrac{1}{4\theta}$.

Con esta información completamos:

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow[n\to\infty]{d} N\left(0, \sqrt{\frac{1}{2\theta}}\right)$$

Buscamos ahora la convergencia asintótica del estimador por el método de los momentos:

$$\sqrt{n}(\tilde{\theta} - \theta) = \sqrt{n}\left(\overline{X}\frac{2}{\pi} - \mathbb{E}(X)\frac{2}{\pi}\right) = \sqrt{\frac{2}{\pi}}\sqrt{n}(\overline{X} - \mathbb{E}(X))$$

que, por el TCL (II.8)

$$\sqrt{\frac{2}{\pi}}\sqrt{n}(\overline{X} - \mathbb{E}(X)) \xrightarrow[n\to\infty]{d} \sqrt{\frac{2}{\pi}}N\left(0, \theta\sqrt{\frac{4-\pi}{2}}\right) = N\left(0, \theta\sqrt{\frac{4-\pi}{\pi}}\right)$$

y por lo tanto es efectivamente asintóticamente normal.

Ejercicio 3.6: *Se dice que una v.a. X tiene distribución Beta de parámetros a > 0 y*
b > 0 (y se denota $X \sim Beta(a, b)$) si su función de densidad es

$$f(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x).$$

siendo Γ la función gamma que aparece en la definición de la distribución del mismo nombre. Calcular el valor de $\mathbb{E}(X)$ y $\mathbb{V}(X)$.

Vamos a utilizar la siguiente propiedad de la gamma: $\Gamma(n+1) = n \cdot \Gamma(n)$. Empecemos con $\mathbb{E}(X)$:

$$\mathbb{E}(X) = \int_0^1 xf(x)dx = \int_0^1 x \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}dx = \tag{B.2}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)}\underbrace{\int_0^1 \frac{\Gamma(a+1+b)}{\Gamma(a+1)\Gamma(b)}x^{(a+1)-1}(1-x)^{b-1}dx}_{=1 \text{ porque es la función de densidad de una Beta(a + 1, b)}} =$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{a \cdot \Gamma(a)\Gamma(b)}{(a+b)\cdot\Gamma(a+b)} = \frac{a}{a+b}$$

Y ahora calcularemos la varianza:

$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 \tag{B.3}$$

$$\mathbb{E}\left(X^2\right) = \int_0^1 x^2 f(x)dx = \int_0^1 x^2 \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}dx = \tag{B.4}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+2+b)} \underbrace{\int_0^1 \frac{\Gamma(a+2+b)}{\Gamma(a+2)\Gamma(b)} x^{(a+2)-1}(1-x)^{b-1}dx}_{=1 \text{ porque es la función de densidad de una Beta(a + 2, b)}} =$$

$$= \frac{(a+1)a}{(a+b+1)(a+b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{(a+1)a}{(a+b+1)(a+b)}$$

Sustituimos en (B.3) lo obtenido en (B.2) y (B.4):

$$\mathbb{V}(X) = \frac{(a+1)a}{(a+b+1)(a+b)} - \left(\frac{a}{a+b}\right)^2 = \frac{(a+1)\cdot a \cdot (a+b) - a^2 \cdot (a+b+1)}{(a+b+1)(a+b)^2} =$$

$$= \frac{a^3 + a^2b + a^2 + ab - a^3 - a^2b - a^2}{(a+b+1)(a+b)^2} = \frac{ab}{(a+b+1)(a+b)^2}$$

Ejercicio 3.7:

Ver transparencias 36 y 37 del tema 3.

Ejercicio 3.8: *Sea $X \sim N(\mu, \sqrt{\theta})$. Estamos interesados en la estimación de θ basados en muestras $X_1, \dots, X_n$ de tamaño n. Calcular la cota de Fréchet-Cramer-Rao (III.7) para estimadores insesgados.*

La cota FCR es

$$\frac{1}{nI(\theta)}$$

Podíamos calcular la información de Fisher como

$$I(\theta) = \mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right)$$

Usaremos la segunda expresión. Calculamos primero el logaritmo:

$$\log f(X;\theta) = \frac{-1}{2}\log 2\pi - \frac{1}{2}\log\theta - \frac{1}{2\theta}(x-\mu)^2$$

y derivamos dos veces

$$\frac{\partial}{\partial\,\theta}\log f(X;\theta) = \log f(X;\theta) = -\frac{1}{2\theta} + \frac{1}{2\theta^2}(x-\mu)^2$$
$$\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = \frac{1}{2\theta^2} - \frac{2}{2\theta^3}(x-\mu)^2 = \frac{1}{\theta^2}\left(\frac{1}{2} - \frac{1}{\theta}(x-\mu)^2\right)$$

Calculamos ahora la esperanza:

$$\mathbb{E}\left(\frac{1}{\theta^2}\left(\frac{1}{2} - \frac{1}{\theta}(x-\mu)^2\right)\right) = -\frac{1}{\theta^2}\left(\frac{1}{2} - \frac{1}{\theta}\underbrace{\mathbb{E}\left(X-\mu\right)^2}_{\theta}\right) = \frac{1}{2\theta^2}$$

y por lo tanto la cota FCR vale $\dfrac{2\theta^2}{n}$, el valor mínimo.

Ejercicio 3.9: *Sea $X_1,\dots,X_n$ una muestra de una v.a. con función de densidad*

$$f(x;\theta) = \theta x^{\theta-1}$$

Sea

$$T_n(X_1,\dots,X_n) = \frac{-1}{n}\sum_{i=1}^{n}\log X_i$$

a) *Probar que*

$$\mathbb{E}_\theta\left(T_n\right) = \frac{1}{\theta};\ \mathbb{V}_\theta\left(T_n\right) = \frac{1}{n\theta^2}$$

b) *¿Es eficiente T_n como estimador de $\frac{1}{\theta}$?*

Apartado a)

Aplicamos que la esperanza de la media muestral de una variable es la esperanza de la variable. En este caso nuestra variable es $logX$.

$$\mathbb{E}_\theta\left(T_n\right) = -\mathbb{E}_\theta\left(\log X\right) = -\int_0^1 \log x\theta x^{\theta-1}\,dx = \frac{1}{\theta}$$

Calculamos ahora la varianza (aplicando $\mathbb{V}\left(\overline{X}\right) = \dfrac{\mathbb{V}\left(X\right)}{n}$).

$$\mathbb{V}_\theta\left(T_n\right) = \frac{\mathbb{V}_\theta\left(\log X\right)}{n} =$$
$$= \mathbb{E}_\theta\left(\log^2 X\right) - \mathbb{E}_\theta\left(\log X\right)^2 = \frac{1}{\theta^2}$$

Ejercicio 3.10: *El número de fallos que se producen anualmente en cierto mecanismo es una v.a. con distribución de Poisson de parámetro θ. El valor de θ no es conocido exactamente, pero se tiene cierta información a priori que permite considerarlo como una v.a. con distribución $\gamma(a,p)$ (a y p son conocidos). Si $x_1, \dots, x_n$ son observaciones observaciones independientes de la variable aleatoria "número de fallos", calcular la distribución a posteriori y obtener, a partir de ella, un estimador puntual para θ.*

Sea $X \equiv$ número de fallos anuales $\sim Poisson(\theta)$, $\theta > 0$.
Su función de densidad es

$$f(x|\theta) = \mathbb{P}_\theta\{X = x\} = e^{-\theta}\frac{\theta^x}{x!}\ ,\ \ x = 1, 2, 3, ...$$

Y el prior es

$$\pi(\theta) = \frac{a^p}{\Gamma(p)} e^{-a\cdot\theta}\theta^{p-1} \ \text{ con } \theta > 0, a > 0, p > 0$$

Entonces

$$\pi(\theta|x_1, \dots, x_n) \approx f(x_1, \dots, x_n|)\pi(\theta) = \left(\prod_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!}\right)\frac{a^p}{\Gamma(p)} e^{-a\cdot\theta}\theta^{p-1} \sim$$

$$\sim e^{-\theta(n+a)}\theta^{(\sum^n x_i + p)-1} \sim \Gamma(a+n, \sum^n x_i + p)$$

Luego el estimador Bayes de θ es

$$\mathbb{E}\left(\pi(\theta|x_1, \dots, x_n)\right) = \frac{\sum^n x_i + p}{a+n} = \frac{\sum^n x_i}{a+n} + \frac{p}{a+n} = \overline{x}\cdot\underbrace{\frac{n}{a+n}}_{\xrightarrow[n\to\infty]{}1} + \frac{p}{a}\cdot\underbrace{\frac{a}{a+n}}_{\xrightarrow[n\to\infty]{}0} \Rightarrow \hat{\theta} = \overline{x}$$

Ejercicio 3.11: *(Este ejercicio es del parcial del año pasado)*

$X \rightsquigarrow Unif[0,\theta]$ Con

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x \leq \theta \\ 0 & x \notin [0,\theta] \end{cases}$$

Vamos a calcular la función de distribución:

$$F_\theta(x) = \mathbb{P}_\theta\{X \leq x\} = \int_{-infty}^{x} f_\theta(t)dt = \int_0^x \frac{1}{\theta}dt = \frac{x}{\theta}\ si 0 \leq x \leq \theta$$

$$F_\theta = \begin{cases} \frac{x}{\theta} & 0 \leq x \leq \theta \\ 0 & x \notin [0,\theta] \end{cases}$$

Nos piden dibujar las funciones.

Vamos a calcular

$$L_n(\theta; x_i) = \prod_{i=1}^{n} f_\theta(x_i) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \forall x_i \in [0,\theta] \\ 0 & \exists x_i \notin [0,\theta] \end{cases}$$

Calculamos la $logL_n$ que nos piden dibujarla:

$$logL_n(\theta) = \begin{cases} -nlog(\theta) & si\ max(\{x_i\}) \leq \theta \\ 0 & si\ no \end{cases}$$

Dibujoo!

$$\hat{\theta_n} = e.m.v.(\theta) = max\left(L_n(\theta)\right)$$

También vale tomando el logaritmo:

$$\hat{\theta}_n = e.m.v.(\theta) = arg\ maslogL_n(\theta) = max\{x_i\}$$

porque

$$logLn(\theta) = \begin{cases} -nlog(\theta) & max\{x_i\} \leq \theta \\ -\infty & si\ no \end{cases}$$

B.4. Tema 4 - Intervalos de confianza

Ejercicio 4.1 y 2:

a) *Representa un estimador de la función de densidad de la v.a. X = cantidad de contaminación por mercurio (en p.p.m.) en los peces capturados en los ríos norteamericanos Lumber y Wacamaw (ver fichero Datos-mercurio.txt). Comparar esta densidad estimada con la densidad normal de igual media y desviación típica (representada en la misma gráfica). En vista de las dos funciones dirías que la función de densidad de X es aproximadamente normal?*

b) *Obtener un intervalo de confianza de nivel 0.95 para la media de X.*

c) *Se puede considerar fiable este intervalo a pesar de la posible no-normalidad de X?*

d) *Qué tamaño muestral habrá que tomar para estimar la contaminación media con un error máximo de 0.06?*

Solucionado por Amparo, descargable aquí.

Ejercicio 4.3:

a) *Representa en un mismo gráfico las densidades de las distribuciones χ^2_k con k = 4,8,20,30.*

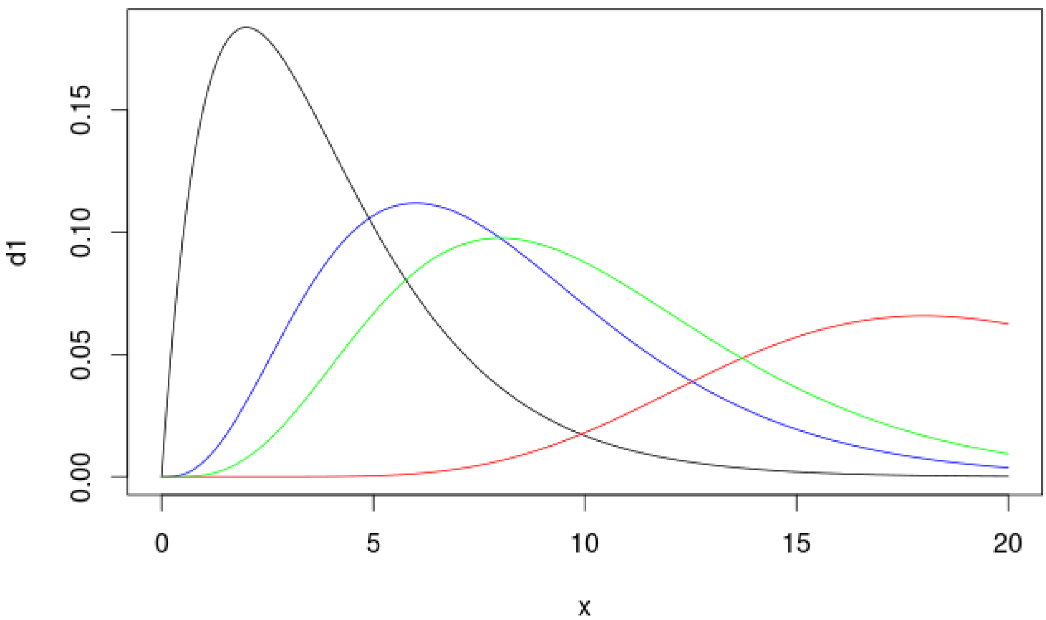
b) $X \sim \gamma(5, 10)$. *Calcular* $\mathbb{P}\{X \leq 3\}$

c) *Sea* $Y \sim \chi^2_{200}$. *Calcular* $\mathbb{P}\{Y \leq 3\}$

Apartado a)

El código R utilizado para generar las gráficas es:

```
> x = seq(0,20,length.out=1000)
> d1=dchisq(x,df=4)
> d2=dchisq(x,df=8)
> d3=dchisq(x,df=10)
> d4=dchisq(x,df=20)
> plot(x,d1,type='l')
> lines(x,d2,type='l',col='blue')
> lines(x,d3,type='l',col='green')
> lines(x,d4,type='l',col='red')
```



Apartado b)

Vamos a usar el resultado visto en clase: Si $X \sim \gamma(a, p)$ entonces tenemos que

$$X \sim c \cdot \gamma(\frac{a}{c}, p) \implies c \cdot X \sim \gamma(\frac{a}{c}, p)$$

$$\text{Como } \gamma\left(\frac{a}{c}, p\right) \sim \chi\left(\frac{1}{2}, \frac{k}{2}\right) \text{ y } a = 5, p = 10$$

Tenemos que $c = 10$, luego:

$$\mathbb{P}\left\{10X \leq 30\right\} = \mathbb{P}\left\{\chi^2_{20} \leq 30\right\}$$

Tenemos varias opciontes. Una de ellas es ir a R y calcularlo con el comando

$$pchisq(30, 20) = 0.93$$

Y la otra es irse a las tablas y vemos que $\mathbb{P}\{\chi^2_{20} \leq 30\} \simeq 1 - \frac{0.1+0.05}{2} = 0.93$, ya que en las tablas estamos entre 28.4 y 31.4.

Apartado c)

Sea $Y \sim \chi^2_{200}$

Podemos hacerlo en R directamente y nos da $\mathbb{P}\left\{Y \leq 3\right\} = 10^{-141}$

A mano, aplicamos el T.C.L, que dice:

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow[n\to\infty]{d} N(0, \sigma)$$

Entonces tenemos: $\overline{X} \sim N\left(\mathbb{E}(X), \sqrt{\frac{\mathbb{V}(X)}{n}}\right)$

Donde[5] $\mathbb{E}(X) = \mathbb{E}(Z^2) = \mathbb{V}(Z) = 1$ y $\mathbb{V}(X) = \mathbb{V}(Z^2) = \mathbb{V}(\chi_1^2) = 2$

Con lo que:

$$\overline{X} \sim N\left(1, \sqrt{\frac{2}{200}}\right) = N\left(1, \frac{1}{10}\right)$$

Sustituyendo y estandarizando:

$$\mathbb{P}\left\{\overline{X} \leq \frac{3}{20}\right\} \simeq \mathbb{P}\left\{Z \leq \frac{\frac{3}{200} - 1}{\frac{1}{10}}\right\} = \mathbb{P}\{Z \leq -9.85\} = 3 \cdot 10^{-23}$$

Una diferencia bastante distinta a lo que decía R. Tras un debate entre Miguel y Amparo de 10 minutos no se ha llegado a ninguna conclusión.

Ejercicio 4.4:

a) *Utilizando el fichero Datos-lipidos.txt, estima, mediante un intervalo de confianza de nivel 0.95, la proporción de pacientes que tienen una concentración de colesterol superior o igual a 220 mg/dl. ¿Qué tamaço muestral habrá que usar para tener una probabilidad aproximada de 0.95 de no cometer un error mayor que 0.01 en la estimación de esta proporción?*

b)

Solucionado por Amparo, descargable aqui

Ejercicio 4.5: *Sea una v.a. con función de densidad* $f(x;\theta) = \theta x^{-(\theta+1)} \mathbb{1}_{[1,\infty)}$

a) *Obtener el e.m.v.*

b) *Obtener su distribución asintótica*

c) *Calcular la cantidad pivotal aproximada y, a partir de ella, un intervalo de confianza de nivel aproximada* $1 - \alpha$ *para* θ

Apartado a)

$$\frac{\partial\, logL(\theta)}{\partial\, \theta} = 0 \implies e.m.v.(\theta) = \frac{1}{\overline{Y}}$$

donde $Y = logX_i$

Apartado b)

[5] Recuerda: $\mathbb{V}(\chi_k^2) = 2 * k$

Posibles caminos:

a) $\hat{\theta} \xrightarrow[n\to\infty]{d}$¿?

b) $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} N\left(0, ?\right)$

La primera opción es algo difusa y la segunda es mucho más concreta y mejor.

Tenemos que examinar la expresión $\sqrt{n}(\hat{\theta}-\theta)$ Tenemos 2 posibilidades con las que calcular este tipo de cosas (T.C.L.) y método delta (que es el que emplearemos a continuación)

$$\mu = \mathbb{E}(X); \sigma = \mathbb{V}(X)$$
$$\sqrt{n}\left(g(\overline{X}) - g(u)\right) \xrightarrow[n\to\infty]{} N(0, \left|g'(u)\right| \sigma)$$

Aplicando el método delta:

$$\sqrt{n}(\hat{\theta}-\theta) = \sqrt{n}\left(g(\overline{y}) - g(\mathbb{E}(Y))\right) \xrightarrow[n\to\infty]{d} N\left(0, \underbrace{\left|g'\left(\frac{1}{\theta}\right)\right|}_{\theta^2}\sqrt{\mathbb{V}(Y)}\right) = N(0,\theta)$$

Peeero... hay que tener cuidado con que $\theta = g(\mathbb{E}(Y))$ porque sino no podemos aplicar el método delta.

$$\mathbb{V}(Y) = \mathbb{E}\left(Y^2\right) - \mathbb{E}\left({}^2Y\right) = \underbrace{\int_1^\infty (log\, x)^2\theta x^{-(\theta+1)}dx}_{\frac{2}{\theta^2}} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

Apartado c)

La cantidad pivotal es un estadístico que depende de la muestra y del parámetro desconocido (del que estamos calculando el intervalo) y cuya distribución, al menos asintóticamente) es totalmente conocida.

En el apartado b) hemos encontrado la distribución asintótica para poder construir la cantidad pivotal.

Tipificamos el resultado anterior para evitar que la distribución dependa del parámetro desconocido.

$$\frac{1}{\theta}\sqrt{n}(\hat{\theta}-\theta) = \sqrt{n}\left(\frac{\hat{\theta}}{\theta} - 1\right) = \mathbb{Q}(\theta; X_1, ..., X_N)$$

Esta es nuestra cantidad pivotal, que depende de la muestra (por el $\hat{\theta}$) y depende del parámetro.

$$1-\alpha = \mathbb{P} = \{q_1(\alpha) \leq \mathbb{Q}(\theta; X_1, ..., X_N) \leq q_2(\alpha)\}$$

Tras despejar obtenemos

$$IC_{1-\alpha}(\theta) = \left(\frac{\hat{\theta}}{1+\frac{1}{\sqrt{n}}z_{\alpha/2}}, \frac{\hat{\theta}}{1-\frac{1}{\sqrt{n}}z_{\alpha/2}}\right)$$

Ejercicio 4.6: *Sea $X_1, \dots, X_n$ una muestra de una v.a. uniforme en el interalo $[0, \theta]$ con $0 < \theta < 1$. Obtener una cantidad pivotal para θ a partir del emv. Usando esta cantidad pivotal construye un intervalo de confianza para θ de nivel prefijado $1 - \alpha$.*

El e.m.v es

$$emv(\theta) = \hat{\theta} = \max X_i$$

La cantidad pivotal para $\theta = Q(\theta; X_1, \dots, X_n)$

$$F_{X_{(n)}}(x) = \mathbb{P}\left\{\hat{\theta}_n \leq x\right\} = \mathbb{P}\left\{X_{(n)} \leq x\right\} = \prod_{i=1}^{n} \mathbb{P}\left\{X_i \leq x\right\} = \begin{cases} 0 & x < 0 \\ \left(\frac{x}{\theta}\right)^n & 0 \leq x \leq \theta \\ 1 & x > 1 \end{cases}$$

Tomo $Q(\theta; X_1, \dots, X_n) = \dfrac{X_{(n)}}{\theta} = \dfrac{\hat{\theta}}{n}$, que es válido como cantidad pivotal porque

$$\mathbb{P}\left\{Q \leq x\right\} = \mathbb{P}\left\{\frac{X_{(n)}}{\theta} \leq \ x\right\} = \begin{cases} 0 & x < 0 \\ x^n & 0 \leq x \leq \theta \\ 1 & x > 1 \end{cases}$$

Tenemos que elegir dos valores q_1, q_2 de tal forma que

$$1 - \alpha \ = \mathbb{P}\left\{q_1(\alpha) \leq Q(\theta; X_1, \dots, X_n) \leq q_2(\alpha)\right\}$$

¿Cómo elegirlos? Queremos buscar que la longitud del intervalo de confianza $IC_{1-\alpha}(\theta) = \left(\dfrac{\hat{\theta}_n}{q_2}, \dfrac{\hat{\theta}_n}{q_1}\right)$ sea mínima. Calculamos esa longitud:

$$\text{len IC} = \hat{\theta}_n \left(\frac{1}{q_1} - \frac{1}{q_2}\right) = \hat{\theta}_n \left(\frac{q_2 - q_1}{q_1 q_2}\right)$$

Es decir, tenemos que buscar que $q_1 - q_2$ sea más pequeño y además tienen que ser lo mayores posible. Por lo tanto, la elección óptima es

$$q_2 = 1, \ q_1 = \alpha^{1/n}$$

Ejercicio 4.7: *Construye tres intervalos de confianza asintóticos diferentes para el parámetro λ de una distribución de Poisson usando los tres métodos siguientes:*

a) *Utiliza el comportamiento asintótico de la media muestral, estima de forma consistente la varianza y aplica el teorema de Slutsky.*

b) *Igual que el anterior, pero sin estimar la varianza*

c) *Aplicando el método delta para estabilizar la varianza, es decir, buscando una función g tal que $\sqrt{n}(g(\overline{X}) - g(\lambda)) \xrightarrow[n\to\infty]{d} N(0,1)$.*

Apartado a)

El TCL (II.8) nos dice que

$$\sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\lambda}} \xrightarrow[n\to\infty]{d} N(0,1)$$

Entonces tenemos que

$$1 - \alpha = \mathbb{P}\left\{ -z_{\alpha/2} \leq \sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\lambda}} \leq z_{\alpha/2} \right\} \tag{B.5}$$

Sustituyo λ en el denominador por una estimación consistente $\hat{\lambda} \xrightarrow[n\to\infty]{P,c.s} \lambda$:

$$\sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\hat{\lambda}}} \xrightarrow[n\to\infty]{d} N(0,1)$$

Como sabemos que $\lambda = \mathbb{E}(X)$, tomamos la media muestral como el estimador: $\hat{\lambda} = \overline{X}$. La convergencia nos queda entonces como

$$\sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\overline{X}}} \xrightarrow[n\to\infty]{d} N(0,1)$$

y por lo tanto tomamos $\sqrt{n}\frac{\overline{X} - \lambda}{\sqrt{\overline{X}}}$ como nuestra cantidad pivotal. Despejamos ahora en (B.5):

$$\mathbb{P}\left\{ \overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}} \leq \lambda \leq \overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}} \right\}$$

Apartado b)

Partimos de nuevo de (B.5), pero no tenemos que estimar λ. Esta ecuación es equivalente a

$$\mathbb{P}\left\{n\frac{(\overline{X}-\lambda)^2}{\lambda} \leq z_{\alpha/2}^2\right\}$$

De ahí sólo tenemos que despejar λ para hallar nuestro intervalo de confianza.

Apartado c)

Tenemos que buscar que se satisfaga la ecuación

$$\sqrt{n}(g(\overline{X}) - g(\lambda)) \xrightarrow[n\to\infty]{d} N(0,1)$$

Sin embargo, el método delta (III.1.1.3) nos dice algo distinto:

$$\sqrt{n}(g(\overline{X}) - g(\lambda)) \xrightarrow[n\to\infty]{d} N(0, \left|g'(\mu)\right| \sqrt{\mathbb{V}(X)})$$

Entonces tenemos que

$$\left|g'(\lambda)\right| \sqrt{\lambda} = 1 \implies g'(\lambda) = \frac{1}{\sqrt{\lambda}}$$

e integrando vemos que $g(\lambda) = 2\sqrt{\lambda}$.

Ejercicio 4.8:

a) *Se desea evaluar aproximadamente, por el método de Montecarlo, la integral*

$$p = \int_0^1 f(x)\,dx$$

de una función continua $f : [0,1] \longmapsto [0,1]$*. Para ello se generan 500 observaciones independientes* (X_i, Y_i) *con* $i = 1, \dots, 500$ *con distribución uniforme en el cuadrado* $[0,1]\times[0,1]$ *y se estima* p *mediante*

$$\hat{p} = \sum_{i=1}^{500} \frac{Z_i}{500}$$

donde la v.a. Z_i *vale 1 si* $Y_i \leq f(X_i)$ *y* 0 *en caso contrario. ¿Qué distribución tienen las* Z_i*? Suponiendo que, en una muestra concreta hemos obtenido* $\sum_{i=1}^{500} z_i = 255$*, obtener un intervalo de confianza de nivel* 0.99 *para la correspondiente estimación de* p*.*

Apartado a)

La v.a. sigue una distribución de Bernoulli, de tal forma que

$$\mathbb{P}\{Z=1\} = \mathbb{P}\left\{Y \leq f(X)\right\} \tag{B.6}$$

La distribución de densidad de la v.a. (X_i, Y_i) es

$$f(x,y) = \begin{cases} 1 & (x,y) \in [0,1] \times [0,1] \\ 0 & \text{en otro caso} \end{cases}$$

Aplicando esto en $(B.6)$

$$\mathbb{P}\left\{Z = 1\right\} = \mathbb{P}\left\{(X,Y) \in \{(x,y) \;/\; y \leq f(x)\;\}\right\} = \int_0^1 \int_0^{f(x)} dy\, dx = \int_0^1 f(x)\, dx = p$$

y llegamos a la forma de estimar la integral que queríamos.

Vamos a contruir el intervalo de confianza de nuvel 0.99.

$$IC_{0.99}(p) = \left(\overline{z} \pm Z_{0.005}\sqrt{\frac{\overline{z}(1-\overline{z})}{500}}\right) = \left(\hat{p} \pm 2575\sqrt{\frac{\hat{p}(1-\hat{p})}{500}})\right) = (0.45 \pm 0.057)$$

Apartado b)

En este caso sabemos el valor de

$$p = \int_0^1 x^2 dx = \frac{1}{3}$$

Buscamos un n que cumpla:

$$z_{0.005}\sqrt{\frac{\frac{1}{3} \cdot \frac{2}{3}}{n}} \implies n > 14734.72$$

Ejercicio 4.9: *Sea X una v.a. con distribución normal de media μ y varianza θ. Estamos interesados en la estimación de θ basados en muestras $X_1, ..., X_n$. Si s^2 denota la cuasivarianza muestral, calcular $\mathbb{V}\left(s^2\right)$ y compararla con la cota de Fréchet-Cramer-Rao obtenida en la relación 3 de problemas.*

Comentarios previos: Sabemos que s^2 es un estimador insesgado de

$$\mathbb{V}\left(X\right) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Vamos a calcular $\mathbb{V}\left(s^2\right)$

Posibilidades:

- Aunque es un poco largo

$$\mathbb{V}\left(s^2\right) = \mathbb{E}\left(s^4\right) - \left[\mathbb{E}\left(s^2\right)\right]^2$$

- Si $X \sim N(\mu, \sigma)$ entonces
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Vamos a utilizar la segunda opción[6] y que $\mathbb{V}\left(\chi^2_{n-1}\right) = 2(n-1)$:

$$\mathbb{V}\left(s^2\right) = \mathbb{V}\left(\frac{n-1}{\sigma^2}s^2 \cdot \frac{\sigma^2}{n-1}\right) = \frac{\sigma^4}{(n-1)^2}\mathbb{V}\left(\frac{n-1}{\sigma^2}s^2\right) =$$
$$= \frac{\sigma^4}{(n-1)^2}\mathbb{V}\left(\chi^2_{n-1}\right) \overset{\sigma^2=\theta}{=} \frac{\theta^2}{(n-1)^2}2(n-1) = \frac{2\theta^2}{n-1}$$

s^2 por lo tanto no es eficiente $\left(\text{porque la Cota de FCR es: } \frac{2\theta}{n}\right)$. Por ser θ la varianza de una $N(\mu, \sigma)$, cuya cota de FCR se calcula en el problema 8H3.

[6] ver (III.2.3.1)

B.5. Tema 5 - Contraste de hipótesis

B.5.1. Hoja 5A

Ejercicio 5.1: *En octubre de 2007 el periódico The New York Times realizó un muestreo en 20 restaurantes y tiendas de Nueva York con objeto de analizar la variable X, que representa el contenido en ppm de metilmercurio en el sushi de atún que se pone a la venta. La media y la cuasi-desviación típica muestrales obtenidas con estas 20 observaciones de X fueron $\overline{x} = 0.794$, $s = 0.2953$. Supongamos que X tiene distribución aproximadamente normal.*

a) *¿Proporcionan estos datos suficiente evidencia estadística a nivel 0.05 a favor de la hipótesis de que la concentración media de metilmercurio en las raciones de sushi de atún en la población considerada es superior a 0.6 ppm? El p-valor, ¿es menor o mayor que 0.01?*

b) *Obtener, a partir de estos datos, un intervalo de confianza de nivel 0.95 para la concentración media de metilmercurio μ en toda la población. Calcular el mínimo tamaño muestral mínimo que habría que utilizar para, con una probabilidad de 0.95, estimar la concentración media de metilmercurio con un error máximo de 0.06 ppm.*

Apartado a)

Empezamos definiendo la hipótesis nula, que será que $\mu \leq 0.6$ ya que queremos una evidencia muy fuerte para rechazar que la concentración suba del nivel mínimo.

Tenemos el siguiente contraste a nivel $\alpha = 0.05$:

$$\begin{aligned} H_0 : &\quad \mu \leq 0.6 \\ H_1 : &\quad \mu > 0.6 \end{aligned}$$

La región de rechazo en este caso es

$$R = \{T > t_{19;\alpha}\}$$

donde

$$T = \frac{\overline{x} - 0.6}{0.2953/\sqrt{20}} = 2.938$$

Por otra parte, $t_{19;\alpha} = 1.729$. Se cumple la condición de la región de rechazo, por lo tanto rechazamos H_0. El p-valor del contraste tendrá que ser menor entonces que 0.05.

Para saber si el p-valor es menor que 0.01 calculamos $t_{19;0.01} = 2.53$. Como sigue siendo menor que T, seguimos rechazando H_0 y por lo tanto el p-valor del contraste será menor que 0.01.

Si quisiésemos obtener el p-valor concreto del contraste, buscaríamos el valor de α tal que $t_{19;\alpha} = 2.938$. En R, obtendríamos este valor con la orden

```
> pt(2.938, 19, lower.tail=FALSE)
[1] 0.004221168
```

El p-valor es por lo tanto 0.004. Esto quiere decir que la probabilidad de obtener la muestra que hemos conseguido suponiendo que H_0 sea cierta (esto es, suponiendo que la media de ppm de metilmercurio en el atún es menor que 0.6) es extremadamente baja, y o bien hemos obtenido una muestra muy, muy extraña o H_0 es falsa. Por lo tanto, lo razonable sería rechazar la hipótesis nula y decir que, de media, la concentración de metilmercurio es mayor que 0.6.

Apartado b)

El intervalo de confianza sería

$$IC_{0.95}(\mu) = \left(\overline{x} \mp t_{n-1;\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = (0.656, 0.932)$$

Como además $0.6 \notin IC_{0.95}(\mu)$, rechazaríamos $H_0 : \mu = 0.06$ a nivel $\alpha = 0.05$.

Para hallar el tamaño muestral mínimo buscamos que

$$IC_{0.95}(\mu) = (\overline{x} \mp 0.06)$$

Despejando, tenemos que resolver

$$t_{n-1;0.025} \frac{s}{\sqrt{n}} < 0.06$$

Como no conocemos s, lo sustituimos por una aproximación, la cuasivarianza muestral de los 20 restaurantes que teníamos al principio. Además, intuimos que n va a ser grande y por lo tanto t se aproximaría a una distribución normal $Z = N(0,1)$, y por lo tanto

$$t_{n-1;0.025} \approx z_{0.025} = 1.96$$

y entonces $n > 93$.

Otra forma de aproximar el $t_{n-1;0.025}$ sería sustituirlo por t_1 ya que a menos grados de libertad, menor peso tienen las colas, luego $t_{n-1;0.025} < t_{1;0.025}$.

Despejando obtenemos que n > 3910.

Finalmente, otra forma de aproximarlo sería tomar $n - 1 = 20$, ya que sabemos que el n va a ser mayor que 20. Con esta aproximación obtenemos que n > 105.

Ejercicio 5.2:

En vista de los datos `Datos-mercurio.txt` ¿hay suficiente evidencia estadística para afirmar que el nivel medio de contaminación por mercurio en los dos ríos es diferente? Contrastar la hipótesis de igualdad de varianzas.
Indicar, en cada caso, las suposiciones previas necesarias para garantizar la validez de los procedimientos empleados.

Suponemos que X = Nivel de contaminación por mercurio en un pez (de la especie *large mouth bass*) elegido al azar en el río Lumber e Y = Nivel de contaminación por mercurio en un pez (de la misma especie) del río Wacamaw son v.a. independientes y siguen una distribución normal: $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$.
Contrastemos primero la hipótesis de igualdad de varianzas a nivel α:

$$\begin{aligned} H_0 : &\quad \sigma_1 = \sigma_2 \\ H_1 : &\quad \sigma_1 \neq \sigma_2. \end{aligned} \tag{1}$$

La región de rechazo es $R = \{s_1^2/s_2^2 \notin [F_{n_1-1;n_2-1;1-\alpha/2}, F_{n_1-1;n_2-1;\alpha/2}]\}$.

```
X = read.table('Datos-mercurio.txt')
ContHg = X$V5
Rio = X$V1
ContHgL = ContHg[Rio==0]
ContHgW = ContHg[Rio==1]
s2L = var(ContHgL)
s2W = var(ContHgW)
s2L/s2W
[1] 0.6119333
alpha = 0.1
n1 = length(ContHgL)
n2 = length(ContHgW)
c(qf(alpha/2,n1-1,n2-1),qf(alpha/2,n1-1,n2-1,lower.tail=F))
[1] 0.690974 1.430908
```

Por tanto, a nivel $\alpha = 0{,}1$ no podemos considerar las varianzas iguales.

```
alpha = 0.05
c(qf(alpha/2,n1-1,n2-1),qf(alpha/2,n1-1,n2-1,lower.tail=F))
[1] 0.6432225 1.5328961
```

A nivel $\alpha = 0{,}05$ tampoco.
Entonces la región de rechazo del contraste

$$\begin{aligned} H_0 : &\quad \mu_1 = \mu_2 \\ H_1 : &\quad \mu_1 \neq \mu_2 \end{aligned} \tag{2}$$

a nivel de significación α es

$$R = \left\{ |\bar{x} - \bar{y}| \geq t_{f;\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\},$$

donde $f = 169$ es el entero más próximo a

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} = 168{,}57. \tag{3}$$

Como $|\bar{x}-\bar{y}| = 0{,}198$ y $t_{169;0{,}025}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = 0{,}223$, no tenemos suficiente evidencia estadística para rechazar $H_0 : \mu_1 = \mu_2$.

Con R podemos hacer *t-tests* (contrastes en los que el estadístico del contraste sigue una distribución t) de la siguiente manera:

```
t.test(ContHg ~ Rio, alternative = "two.sided", mu = 0, paired = FALSE, var.equal
    = FALSE, conf.level = 0.95)
```

o equivalentemente

```
t.test(ContHgL, ContHgW, alternative = "two.sided", mu = 0, paired = FALSE, var.
    equal = FALSE, conf.level = 0.95)
```

Obtenemos como resultado

```
        Welch Two Sample t-test

data:  ContHgL and ContHgW
t = -1.7547, df = 168.57, p-value = 0.08114
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.42150363 0.02481087
sample estimates:
mean of x mean of y
 1.078082 1.276429
```

El valor `t` es el del estadístico del contraste

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}},$$

y `df` es el valor de la expresión (3). El intervalo de confianza es $\mathrm{IC}_{0{,}95}(\mu_1-\mu_2)$.

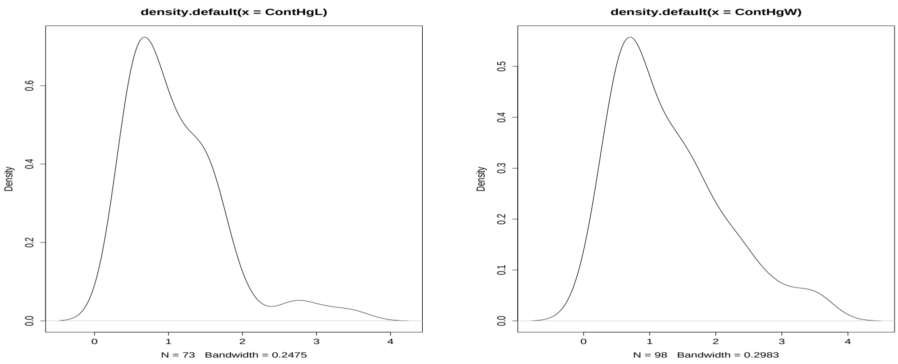
Con `t.test` también podemos hacer contrastes para una sola muestra (es decir, contrastes acerca de la media de una $N(\mu,\sigma)$ con σ desconocido). Por ejemplo, si quisiéramos contrastar $H_0 = \mu_1 \geq 1$ frente a $H_1 : \mu_1 < 1$ escribiríamos:

```
t.test(ContHgL, alternative = "less", mu = 1, conf.level = 0.95)
```

Y para hacer el contraste (1) de igualdad de varianzas

```
> var.test(ContHgL, ContHgW, ratio = 1, alternative = "two.sided", conf.level =
    0.95)

        F test to compare two variances

data:  ContHgL and ContHgW
F = 0.6119, num df = 72, denom df = 97, p-value = 0.0294
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3992008 0.9513555
sample estimates:
ratio of variances
        0.6119333
```

Otra posibilidad para hacer el contraste (2) sin suponer normalidad de X e Y (ver figura)



es utilizar que, por el TCL, $\bar{X} \overset{\text{aprox}}{\sim} N(\mu_1, \sigma_1/\sqrt{n_1})$ e $\bar{Y} \overset{\text{aprox}}{\sim} N(\mu_2, \sigma_2/\sqrt{n_2})$. Si X e Y son independientes entonces

$$\bar{X} - \bar{Y} \overset{\text{aprox}}{\sim} N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

y, por el Teorema de Slustky, si $H_0 : \mu_1 = \mu_2$ es cierta entonces

$$\bar{X} - \bar{Y} \overset{\text{aprox}}{\sim} N\left(0, \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right).$$

A nivel $\alpha = 0{,}05$ no podemos rechazar la hipótesis nula (2) porque $|\bar{x} - \bar{y}| = 0{,}198 < z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0{,}222$, pero sí podemos rechazar a nivel $\alpha = 0{,}1$.
Observemos que, como el tamaño muestral es grande, las regiones de rechazo suponiendo normalidad y utilizando la aproximación del TCL son prácticamente iguales.

Ejercicio 5.3:

Con objeto de averiguar si la estatura de las personas disminuye significativamente a lo largo de la jornada se seleccionaron al azar diez mujeres de la misma edad de las que se midió su estatura (en cm.) por la mañana al levantarse (X_i) y por la noche antes de acostarse (Y_i). Se obtuvieron los siguientes resultados:

X_i	169.7	168.5	165.9	177.8	179.6	168.9	169.2	167.9	181.8	163.3
Y_i	168.2	166.4	166.7	177.2	177.9	168.0	169.5	166.7	182.5	161.1

¿Proporcionan estos datos suficiente evidencia estadística, al nivel 0.05, a favor de la hipótesis de que la estatura disminuye a lo largo de la jornada?

Definimos $D = X - Y$, la variación que experimenta la estatura (en cm.) de una mujer entre el momento de levantarse y el de acostarse. Suponemos que $D \sim N(\mu, \sigma)$ con μ y σ desconocidos. A nivel de significación $\alpha = 0{,}05$, queremos contrastar

$$\begin{aligned} H_0 :&\ \mu \leq 0 \quad \text{(la estatura no disminuye a lo largo del día)} \\ H_1 :&\ \mu > 0 \quad \text{(la estatura disminuye a lo largo del día).} \end{aligned}$$

La región de rechazo de este test es

$$R = \left\{ \bar{d} > t_{n-1;\alpha} \frac{s_d}{\sqrt{n}} \right\},$$

donde $\bar{d} = 0{,}84$ y $s_d = 1{,}11$ son la media y cuasidesviación típica de los valores observados de D:

d_i	1.5	2.1	-0.8	0.6	1.7	0.9	-0.3	1.2	-0.7	2.2

Como $t_{n-1;\alpha} \frac{s_d}{\sqrt{n}} = 1{,}833 \frac{1{,}11}{\sqrt{10}} = 0{,}64 < \bar{d}$, sí hay suficiente evidencia estadística, a nivel $\alpha = 0{,}05$, para rechazar H_0.

Ejercicio 5.4: *Los niveles en sangre de una hormona denominada FSH están asociados con la fertilidad femenina. Las mujeres que tienen un nivel de FSH "alto" (superior a 10 IU/L) tienen en general más dificultad para concebir que aquellas que tienen niveles bajos de FSH. En un estudio realizado recientemente, se analizó la posible relación entre el grupo sanguíneo y la fertilidad. Para ello se midieron los niveles de FSH en una muestra de 254 mujeres en edad fértil con grupo sanguíneo "O" y resultó que 43 de ellas tenían niveles altos de FSH y, por tanto, podrían tener dificultades para concebir. En otra muestra, independiente de la anterior, de 309 mujeres cuyo grupo sanguíneo no es O, resultó que 27 tenían niveles altos de FSH.*

a) *¿Proporcionan estos datos suficiente evidencia estadística, al nivel 0.05, a favor de la hipótesis de que las mujeres con grupo sanguíneo 0 tienen más dificultades para concebir que las que tienen otro grupo sanguíneo?*

b) *Calcular el tamaño muestral necesario para, con probabilidad 0.95, estimar en la población de mujeres del grupo 0 el porcentaje de las que tienen un nivel alto de FSH, con un error máximo de 2 puntos.*

Consideramos la v.a. X que vale 1 si una mujer del grupo 0 tiene nivel alto de FSH y 0 si no, y que sigue una distribución de Bernoulli con probabilidad p_1. Análogamente, definimos la v.a. Y que vale 1 si una mujer del grupo no 0 tiene nivel alto de FSH y 0 si no, y que sigue una distribución de Bernoulli con probabilidad p_2.

Tenemos que

$$\sum_{i=1}^{254} x_i = 43$$
$$\sum_{i=1}^{309} y_i = 27$$

Apartado a)

Primero tenemos que definir la hipótesis nula:

$$H_0 : p_1 \leq p_2$$

es decir, que las mujeres con grupo 0 no tienen más dificultad para concebir. Tomamos esto como la hipótesis nula porque es la que aceptamos por defecto, y queremos una evidencia muy fuerte para poder decir que es falsa.

Para construir la región de rechazo, usamos la región del formulario para comparación de proporciones. Usando el TCL, tenemos que si $p_1 = p_2 = p$ entonces tanto $\overline{X}$ como $\overline{Y}$ van a seguir una distribución normal con $n_i = n_1$ o n_2 según sea X ó Y

$$N\left(p, \sqrt{\frac{p(1-p)}{n_i}}\right)$$

y por lo tanto el estadístico del contraste es

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

siendo $\overline{p}$ un estimador puntual de p, y que se calcula como

$$\overline{p} = \frac{\sum x_i + \sum y_i}{n_1 + n_2} = \frac{n_1\overline{x} + n_2\overline{y}}{n_1 + n_2}$$

La región de rechazo es

$$R = \left\{ \overline{x} - \overline{y} > z_{0.05}\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \right\} \equiv \{0.0819 > 0.0460\ \}$$

y por lo tanto rechazamos la hipótesis nula al nivel $\alpha = 0.05$.

Calculamos ahora el p-valor para tener más datos sobre la hipótesis:

$$\text{p-valor } = \mathbb{P}\left\{N(0,1) > z\right\}\ = \cdots$$

Apartado b)

Necesitamos un intervalo de confianza

$$IC_{0.95}(p_1) = \left(\overline{x} \pm z_0.025\sqrt{\frac{\overline{x}(1-\overline{x}}{n_1}} \right)$$

donde $z_0.025\sqrt{\frac{\overline{x}(1-\overline{x}}{n_1}}$ es el error cometido al estimar p_1 con el IC, y que tiene que ser menor que 0.02. Como no tenemos el valor de $\overline{x}$, lo sustituimos por el valor de la media muestral obtenido en la anterior medición, de tal forma que tenemos que $n_1 \geq 1351$ para obtener la confianza requerida.

Si quisiésemos ser más conservadores, sustituiríamos $\overline{x}$ por el valor máximo que podemos obtener, aunque en este caso saldría un tamaño muestral mucho más grande.

Ejercicio 5.5: *El gasto telefónico medio bimensual en una muestra de 10 usuarios elegidos al azar en una ciudad ha resultado ser 90 euros y la cuasidesviación típica 11 euros. En otra ciudad se ha tomado, de modo independiente, otra muestra de 12 usuarios y los valores obtenidos para la media y la cuasidesviación típica muestrales han sido, respectivamente, 80 y 10.*

a) *¿Proporcionan estos datos suficiente evidencia estadística, al nivel 0.05, a favor de la hipótesis de que el gasto medio en la primera ciudad es más alto* *que el gasto medio en la segunda? Suponer que las varianzas de las variables que indican los gastos telefónicos en ambas ciudades son iguales. Indicar claramente las restantes suposiciones necesarias para garantizar la validez del procedimiento empleado.*

b) *El p-valor ¿es mayor o menor que 0.01? Razonar la respuesta.*

Apartado a)

Definimos las dos variables aleatorias que tenemos: X es el gasto medio bimensual en la primera ciudad, y Y el gasto en la segunda. Tomamos las esperanzas y varianzas:

$$\mathbb{E}(X) = \mu_1, \; \mathbb{V}(X) = \sigma_1^2$$
$$\mathbb{E}(Y) = \mu_2, \; \mathbb{V}(Y) = \sigma_2^2$$

Definimos la hipótesis nula: $H_0 : \mu_1 \leq \mu_2$, es decir, que el gasto medio en la primera ciudad no es mayor que en la segunda.

Tenemos que suponer que X e Y son normales para poder definir bien el estadístico del contraste. Si usásemos cualquier otra distribución el estadístico del contraste toma una distribución mucho más complicada que no podríamos determinar correctamente. También suponemos que son independientes.

La región de rechazo es

$$R = \left\{ \overline{x} - \overline{y} > t_{n_1+n_2-2,\alpha} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

Calculando, tenemos que

$$\overline{x} - \overline{y} = 10 \text{ y } s_p^2 = 109.45$$
$$R = \{10 > 7.73 \}$$

y por lo tanto rechazamos la hipótesis nula.

Apartado b)

Calculamos la región de rechazo para $\alpha = 0.01$:

$$R = \{10 > 11.32\}$$

y por lo tanto para nivel 0.01 no hay evidencia para rechazar H_0. Entonces, el p-valor es mayor que 0.01.

Ejercicio 5.6: *Se realiza un experimento para comparar los incrementos en los niveles plasmáticos de insulina producidos por la ingesta de carne y de pescado. Para ello se midieron los incrementos (medido esn picomoles por litro) producidos en la concentración de insulina en la sangre de 6 voluntarios, 90 minutos después de comer un bistec de 250 gramos. Dos días más tarde se realizó de nuevo el experimento con las mismas 6 personas, después de consumir un filete de pescado. En la tabla se observan los resultados:*

Persona	*1*	*2*	*3*	*4*	*5*	*6*
Resultados con la carne:	*109*	*106*	*111*	*105*	*110*	*108*
Resultados con el pescado:	*100*	*95*	*105*	*106*	*80*	*88*

a) *Proporcionan estos datos suficiente estadística a nivel significación 0.05 para afirmar que el incremento medio...?*

Apartado a)

1) Definir las variables:

- X nivel de insulina en 1 voluntario tras la ingesta de carne. Llamamos a $\mathbb{E}(X) = \mu_1$
- Y nivel de insulina en **el mismo** voluntario tras la ingesta de carne. $\mathbb{E}(Y) = \mu_2$

Tenemos que las variables no son independientes (porque son muestras tomadas de los mismo voluntarios). A este tipo de datos le llamamos datos emparejados

2) Definir las hipótesis

- $H_0 : \mu_1 \leq \mu_2$
- $H_1 : \mu_1 > \mu_2$

3) Como tenemos datos emparejados, podemos trabajar más facilmente con la diferencia, es decir, definimos $D = X - Y$ y definimos el contraste (siendo $\mathbb{E}(D) = \mu$)

- $_0 : \mu \leq 0$
- $H_1 : \mu > 0$

Que es un contraste equivalente.

Además tenemos que $D \sim N(\mu, \sigma)$

Suponer que la diferencia es una normal es el procedimiento estándar para datos emparejados. (nos la jugamos, es una hipótesis del problema, que puede ser más o menos razonable. En este caso, lo único que de momento sabemos hacer es suponer que es normal (si no fuera normal, tendríamos que aplciar el TCL (para lo que necesitamos n grande) y con este tamaño muestral (6) no podríamos aplicarlo)

Mirando en la tabla de regiones de rechazo tenemos:

$$R = \left\{ \overline{d} > t_{n-1;\alpha} \frac{s_d}{\sqrt{n}} \right\}$$

Donde $\dfrac{\overline{d}}{s_d/\sqrt{n}}$ es el estadístico del contraste, que sigue una t_{n-1}.

De los datos extraemos $\overline{d} = 12.5; s_d = 10.97$.

Para $\alpha = 0.05$ calculamos el cuantil correspondiente de la t de Student. Para $\alpha = 0.05$ es 9.02.

De aquí deducimos que sí hay evidencia para rechazar la hipótesis nula (porque $\dfrac{\overline{d}}{s_d/\sqrt{n}} > 9.02$).

Apartado b)

Tomando $\alpha = 0.01$ no se cumple la condición de rechazo, no pudiendo negar entonces la hipótesis nula.

Apartado c)

Es el típico ejercicio mecánico de extraer el tamaño muestral.

Ejercicio 5.7: *Se ha comprobado que la probabilidad de curación espontánea (sin medicación alguna) de cierta enfermedad es de 0.4. Un laboratorio ha obtenido un nuevo medicamento para tratar la enfermedad y quiere demostrar que es eficaz. Para ello, se aplica el tratamiento a 100 pacientes que sufren la enfermedad en silencio y se observa cuántos de ellos se leen este texto.*

a) *Si se han curado 50 personas de las 100. ¿puede afirmarse que el medicamento es eficaz a nivel $\alpha = 0.05$? Calcula el p-valor del contraste.*

b) *¿Cuántas personas de las 100 deberían curarse como mínimo para poder afirmar al nivel $\alpha = 0.001$ que el tratamiento es eficaz?*

c) *Supongamos que la probabilidad de curación con el tratamiento fuese realmente de 0.5 y que se realiza el test de nivel 0.05 con 100 personas. ¿Cuál sería la probabilidad de error, es decir, la probabilidad de rechazar el medicamento como inútil?*

Apartado a)

Sea $X \sim Bernouilli(p)$, luego $\sum_i^{100} x_i$ = "número de pacientes que se curan". Tenemos el siguiente contraste a nivel $\alpha = 0.05$:

$$\begin{aligned} H_0 &: \quad p \leq 0.4 \\ H_1 &: \quad p > 0.4 \end{aligned}$$

La región de rechazo es

$$R = \{z = \frac{\overline{x} - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{100}}} > z_{0.05}\}$$

Como $\frac{\overline{x}-0.4}{\sqrt{\frac{0.4\cdot0.6}{100}}} = 2.041 > 1.645 \implies$ hay evidencia muestral para afirmar que el medicamento es eficaz a nivel $\alpha = 0.05$ (rechazo H_0).

El pvalor se calcula así

```
> pnorm(2.041, lower.tail=FALSE)
[1] 0.02062541
```

Luego a nivel $\alpha = 0.01 <$ p-valor, no habría suficiente evidencia muestral para rechazar la hipótesis nula.

Apartado b)

$$\frac{\overline{x} - 0.4}{\sqrt{\frac{0.4 \cdot 0.6}{100}}} > z_{0.001} \implies \sum^{100} x_i > 55.1$$

Apartado c)

Como $p = 0.5 \to H_1$ es cierta $\implies$ solo puede cometerse el error de tipo II. Luego

$$\mathbb{P}_{p=0.5}\left\{\text{error tipo II}\right\} = \mathbb{P}_{p=0.5}\left\{\text{aceptar } H_0\right\} = 1 - \mathbb{P}_{p=0.5}\left\{R\right\} = 1 - \beta_n(0.5) =$$

$$= 1 - \mathbb{P}\left\{\frac{\overline{X} - 0.4}{\sqrt{\frac{0.4\cdot0.6}{100}}} > z_{0.05}\right\} = 1 - \mathbb{P}\left\{\overline{X} > 0.4 + z_{0.05}\sqrt{\frac{0.4 \cdot 0.6}{100}}\right\} \overset{X \approx N(p,\sqrt{\frac{p\cdot(1-p)}{n}}) = N(0.5, 0.05)}{=}$$

$$= 1 - \mathbb{P}\left\{Z = \frac{\overline{X} - 0.5}{0.05} > \frac{0.4 + z_{0.05} \cdot \sqrt{\frac{0.4\cdot0.6}{100}} - 0.5}{0.05}\right\} =$$

$$= 1 - \mathbb{P}\left\{Z > -0.388\right\} = 1 - (1 - \mathbb{P}\left\{Z > 0.388\right\}) = 0.35$$

Ejercicio 5.8:

a) *Supongamos que en una determinada población de referencia, formada por adultos sanos, el nivel en sangre de la enzima hepática GGT (gamma-glutamil-transpeptidasa) sigue aproximadamente una distribución normal con media poblacional $42IU/L$ y desviación típica poblacional 13. Calcular aproximadamente el porcentaje de personas en la población que tienen un nivel de GGT superior a 80.*

b) *Supongamos ahora que se selecciona una muestra de 61 personas en otra población formada por bebedores habituales no diagnosticados de alcoholismo y se obtiene una media muestra de 58 IU/L con una desviación típica de 21. ¿Hay suficiente evidencia estadística, al nivel 0.05, para afirmar que la concentración media de GGT en la población de bebedores es mayor que 42?*

Apartado a)

Sea $X \sim N(42, 13)$,

$$\mathbb{P}\{X > 80\} = \mathbb{P}\left\{\frac{X-42}{13} > \frac{80-42}{13}\right\} = 0.0017$$

Apartado b)

Sea Y el nivel de GGT en sangre

Tenemos el siguiente contraste a nivel $\alpha = 0.05$:

$$\begin{aligned} H_0 : &\quad \mu \leq 42 \\ H_1 : &\quad \mu > 42 \end{aligned}$$

La región de rechazo es

$$R = \{z = \frac{\overline{y} - 42}{s/\sqrt{61}} > Z_{0.05}\}$$

Y por lo tanto rechazamos H_0 ya que $5.95 > 1.645$. Podemos calcular el p-valor de la siguiente manera

$$\text{p-valor} = \mathbb{P}\left\{N(0,1) > 5.95\right\} = 7 * 10^{-8}$$

Con lo cual, es muy razonable rechazar la hipótesis nula.

B.5.2. Hoja 5B

Ejercicio 5.1: *Tenemos una* $X \sim \exp(\theta)$. *Queremos contrastar para* $\alpha = 0.01$ *las dos siguientes hipótesis:* $H_0 : \theta = 5$ *frente a* $H_1 : \theta = \theta_1$, *siendo* $\theta_1 > 5$ *un valor prefijado.*

a) *Obtener la región crítica del test UMP.*

b) *Calcular la probabilidad de error de tipo II en este test.*

c) *Supongamos que para una determinada muestra, se obtiene* $\sum_{i=1}^{5} x_i = 5$. *¿Qué decisión habría que adoptar si se utiliza el test construido en a)?*

Apartado a)

Primero comprobamos la propiedad de CVM[7]:

$$\frac{f_n(x_1,\dots,x_n;\theta_1)}{f_n(x_1,\dots,x_n;5)} = \left(\frac{\theta_1}{5}\right)^n e^{-(\theta_1-5)\sum^n x_i}$$

Efectivamente, la función es monótona.

Por tanto, la región de rechazo del test UMP es, por el lema de Neyman-Pearson (IV.1), la siguiente:

$$R^* = \left\{ \left(\frac{\theta_1}{5}\right)^n e^{(-\theta_1-5)\sum^n x_i > k_\alpha} \right\}$$

Ya que una vez fijado θ_1 lo que determina la cota superior es el sumatorio, tenemos que

$$R^* = \left\{ \sum x_i < c_\alpha \right\} \text{ tal que } \mathbb{P}_{\theta=5}\left\{R^*\right\} = \alpha$$

Como $X \sim \exp(\theta) = \gamma(\theta, 1)$ y las X_i son v.a.i., tenemos que

$$\sum X_i \sim \gamma(\theta, n)$$

y entonces

$$\mathbb{P}_{\theta=5}\left\{R^*\right\} = \alpha = \mathbb{P}\left\{\gamma(5,n) < c_\alpha\right\}$$

De esta forma, c_α es el cuantil α de la distribución $\gamma(5,n)$:

$$c_\alpha = q_{5;n}(\alpha)$$

Finalmente, como $\alpha = 0.01$, entonces

[7] Ejemplo típico de aplicar el lema de Neyman-Pearson (IV.1).

$$R^* = \left\{ \sum^n x_i < q_{5;n}(0.01) \right\}$$

Apartado b)

Calculamos el error de tipo II, (IV.1)

$$\mathbb{P}_{\theta_1}\left\{R^c\right\} = 1 - \mathbb{P}_{\theta_1}\left\{R\right\} = 1 - \mathbb{P}_{\theta_1}\left\{\sum x_i < q_{5;n}(0,01)\right\} = 1 - \mathbb{P}\left\{\gamma(\theta_1, n) < q_{5;n}(0,01)\right\}$$

Usando las propiedades de la distribución gamma, tenemos que

$$\gamma(\theta_1, n) = \gamma\left(\frac{\theta_1}{5}5, n\right) = \frac{5}{\theta_1}\gamma(5, n)$$

y entonces

$$\mathbb{P}_{\theta_1}\left\{R^c\right\} = 1 - \mathbb{P}\left\{\gamma(5,n) < \frac{\theta_1}{5} q_{5;n}(0,01)\right\} = \mathbb{P}\left\{\gamma(5,n) \geq \frac{\theta_1}{5} q_{5;n}(0,01)\right\} \xrightarrow[\theta_1 \to \infty]{} 0$$

Concretamente tenemos que

$$\mathbb{P}_{\theta_1}\left\{R^c\right\} = 1 - 0.01 - \mathbb{P}\left\{q_{5;n}(0.01) \leq \gamma(5,n) < \frac{\theta_1}{5} q_{5;n}(0,01)\right\} = 0.99 - O(1 - \frac{\theta_1}{5})$$

Lo que quiere decir que la probabilidad de error de tipo II se hace arbitrariamente cercana a 1 - α cerca de θ_1.

Apartado c)

Nuestra muestra nos da una estimación puntual de $\overline{x} = 1$.
Bajo la hipótesis nula, la media de la población debería ser $\frac{1}{5}$, ya que $\mathbb{E}\left(X\right) = \frac{1}{\theta}$.
Bajo la hipótesis alternativa, la media debería ser $< \frac{1}{5}$.

Intuitivamente, no tenemos evidencia muestral en contra de H_0. Comprobémoslo ahora calculando la región de rechazo: tenemos que calcular el cuantil de la distribución Gamma:

$$q_{5,5}(0.01) = 0.2558 \;\; \to 5 \not< 0.2558$$

Luego no hay evidencia muestral para rechazar la hipótesis nula, tal y como habíamos intuido.

Ejercicio 5.2:

En una piscifactoría se desea contrastar la hipótesis nula de que el porcentaje de peces adultos que miden menos de 20 cm es como máximo del 10 % . Para ello, se toma una muestra de 6 peces y se rechaza H_0 si se encuentra más de uno con longitud inferior a 20 cm.

***a)** ¿Cuál es el nivel de significación de este contraste?*

***b)** Calcula la potencia del contraste si en realidad hay un 20 % de peces que miden menos de 20 cm.*

Sea $X \sim Bernouilli(p)$ tal que

$$X = \begin{cases} 1 & \text{si un pez adulto de la piscifactoría mide menos de 20cm} \\ 0 & \text{en otro caso} \end{cases}$$

Tenemos pues el siguiente contraste a nivel α:

$$H_0: \ p \leq 0.1$$
$$H_1: \ p > 0.1$$

Nos dicen que

$$R = \left\{ \sum^6 x_i > 1 \right\} = \left\{ \sum^6 x_i \geq 2 \right\}$$

Nótese que $\sum^6 x_i$ es una binomial (6, p).

Apartado a)

Tamaño del test = $\text{máx}\, \mathbb{P}\left\{\text{error tipo I}\right\} = \text{máx}_{p\leq 0.1} \mathbb{P}_p \left\{R\right\} \leq \alpha$.

Tenemos que maximizar la siguiente expresión:

$$\beta(p) = \mathbb{P}_p\left\{R\right\} = \mathbb{P}_p \left\{ \sum^6 X_i \geq 2 \right\} = 1 - \mathbb{P}_p \left\{ \sum^6 X_i = 0 \right\} - \mathbb{P}_p \left\{ \sum^6 X_i = 1 \right\} =$$
$$= 1 - (1-p)^6 - 6 * (1-p)^5 p = 1 - (1-p)^5(1+5p)$$

Notese que hay 6 formas de obtener un 1 y cinco 0s. Bien, derivemos:

$$\beta'(p) = -5(1-p)^4 * (-1) * (1+5p) - (1-p)^5 * (5) =$$
$$= (1-p)^4(5 + 25p - 5 - 5p) = (1-p)^4(20p) > 0 \ \forall p \in (0,1) \Rightarrow \beta \text{ es creciente en } (0,1)$$

Luego

$$\max_{p\leq 0.1} \mathbb{P}_p \left\{R\right\} = \max_{p \leq 0.1} \beta(p) = \beta(0.1) = 1 - 0.9^5 * (1 + 5 * 0.1) = 0.1143$$

Nótese que como β es monótona creciente, alcanza su máximo en el extremo del intervalo.

Apartado b)

Simplemente debemos calcular el valor de la función de potencia. Es decir, sustituir:

$$\beta(0.2) = 1 - (1 - 0.2)^5(1 + 5 * 0.2) = 0.3446$$

Ejercicio 5.3: *El error que se comete en la medición de una magnitud es una v.a. X cuya función de densidad es*

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}$$

siendo $\theta > 0$ un parámetro que se desea estimar. Obtener el test uniformemente más potente de nivel α para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$

OBSERVACIÓN: los errores de los aparatos de medición se suponen que siguen una N(0, σ).

Sea $X \sim N(0, \sqrt{\theta})$, $\theta > 0$, donde X es "el error cometido por el aparato de medición".

Tenemos que comprobar primero que el cociente de verosimilitudes es monótono. Para ello tomamos $\theta_1 < \theta_2$ y calculamos la razón de verosimilitudes:

$$\frac{f(x_1, \ldots, x_n; \theta_2)}{f(x_1, \ldots, x_n; \theta_1)} = \left(\frac{\theta_1}{\theta_2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right)\sum x_i^2}$$

que sí es una función creciente[8] de $T_n = \sum x_i^2$. Por lo tanto esta es una familia paramétrica CVM (ver definición IV.6). Aplicando el teorema (IV.2)

$$R = \{T_n > k_\alpha\} \,/\, \mathbb{P}_{\theta_0}\{R\} = \alpha = \mathbb{P}_{\theta_0}\left\{\sum^n X_i^2 > k_\alpha\right\}$$

¿Cómo resolvemos la expresión de k_α? Tomamos

$$k_\alpha = \theta_0 \chi^2_{n;\alpha}$$

Por lo que

$$R = \mathbb{P}_{\theta_0}\left\{\sum^n X_i^2 > \theta_0 \chi^2_{n;\alpha}\right\}$$

[8] Porque el exponente de la exponencial es siempre positivo

Ejercicio 5.4:

Sea $X_1, \ldots, X_n$ una muestra de una población con función de densidad

$$f(x, \theta) = e^{-(x-\theta)} \mathbb{1}_{[\theta,\infty)}(x)$$

siendo $\theta > 0$. Calcula el test de razón de verosimilitudes, de nivel α, para contrastar $H_0 : \theta \leq \theta_0$ frente a $H_1 : \theta > \theta_0$.

Solución: La función de verosimilitud es

$$f_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta) = e^{-\sum_{i=1}^{n}(x_i-\theta)} \mathbb{1}_{[\theta,\infty)}(x_{(1)}) = \begin{cases} e^{n\theta - \sum_{i=1}^{n} x_i} & \text{si } \theta \leq x_{(1)}, \\ 0 & \text{si } \theta > x_{(1)}, \end{cases} \tag{1}$$

donde hemos utilizado que $x_{(1)} := \min_{1\leq i\leq n} x_i$ y que

$$\begin{aligned} \prod_{i=1}^{n} \mathbb{1}_{[\theta,\infty)}(x_i) &= \begin{cases} 1 & \text{si } \theta \leq x_i \text{ para todo } 1 \leq i \leq n \\ 0 & \text{si no} \end{cases} \\ &= \mathbb{1}_{[\theta,\infty)} \left(\min_{1\leq i\leq n} (x_i) \right). \end{aligned}$$

Como la verosimilitud (1) es una función creciente en θ, el estimador de máxima verosimilitud (e.m.v.) de θ es $\hat{\theta} = x_{(1)}$.
Calculemos la razón de verosimilitudes

$$\Lambda_n = \frac{\sup_{\theta\in\Theta_0} f_n(x_1, \ldots, x_n; \theta)}{\sup_{\theta\in\Theta} f_n(x_1, \ldots, x_n; \theta)}.$$

Observemos que $\Theta_0 = (0, \theta_0]$ y $\Theta = (0, \infty)$. Como la verosimilitud es creciente en θ, tenemos que

$$\sup_{\theta\in\Theta_0} f_n(x_1, \ldots, x_n; \theta) = e^{n\theta_0 - \sum_{i=1}^{n} x_i}.$$

Como el supremo del denominador de Λ_n se alcanza cuando θ es igual al e.m.v., tenemos que

$$\sup_{\theta\in\Theta} f_n(x_1, \ldots, x_n; \theta) = e^{n\hat{\theta} - \sum_{i=1}^{n} x_i}.$$

Por tanto, $\Lambda_n = e^{n(\theta_0 - x_{(1)})}$.
El test de razón de verosimilitudes es el que tiene como región crítica o de rechazo

$$R = \{(x_1, \ldots, x_1) : \Lambda_n < k_\alpha\} \tag{2}$$

donde k_α se elige de manera que el tamaño del test (la máxima probabilidad de error de tipo I)

$$\max_{\theta\leq\theta_0} \mathbb{P}(R)$$

sea igual a α. Ahora bien, observemos que

$$\Lambda_n < k_\alpha \Leftrightarrow e^{n(\theta_0 - x_{(1)})} < k_\alpha \Leftrightarrow n(\theta_0 - x_{(1)}) < \log(k_\alpha) \Leftrightarrow x_{(1)} > \theta_0 - \frac{1}{n}\log(k_\alpha) =: \theta_0 + c_\alpha.$$

Por tanto, la región de rechazo (2) equivale a la región

$$R = \{(x_1, \ldots, x_1) : x_{(1)} > \theta_0 + c_\alpha\}$$

donde c_α es tal que

$$\max_{\theta\leq\theta_0} \mathbb{P}(R) = \max_{\theta\leq\theta_0} \mathbb{P}_\theta\{X_{(1)} > \theta_0 + c_\alpha\} = \alpha.$$

1

Para completar la expresión de la región de rechazo, determinemos c_α. Observemos que

$$\begin{aligned}
\mathbb{P}_\theta\{X_{(1)} > \theta_0 + c_\alpha\} &= \mathbb{P}_\theta\{\min_{1\le i\le n} X_i > \theta_0 + c_\alpha\} \\
&= \mathbb{P}_\theta\{X_1 > \theta_0 + c_\alpha, X_2 > \theta_0 + c_\alpha, \ldots, X_n > \theta_0 + c_\alpha\} \\
&= \prod_{i=1}^{n} \mathbb{P}_\theta\{X_i > \theta_0 + c_\alpha\} = (\mathbb{P}_\theta\{X > \theta_0 + c_\alpha\})^n \\
&= e^{-n(\theta_0 + c_\alpha - \theta)}, \qquad (3)
\end{aligned}$$

donde hemos usado que

$$\mathbb{P}_\theta\{X > \theta_0 + c_\alpha\} = \int_{\theta_0+c_\alpha}^{\infty} e^{-(x-\theta)} dx = e^{-(\theta_0+c_\alpha-\theta)}.$$

Como la función (3) es creciente en θ tenemos que

$$\alpha = \max_{\theta\le\theta_0} e^{-n(\theta_0+c_\alpha-\theta)} = e^{-n(\theta_0+c_\alpha-\theta_0)} = e^{-nc_\alpha} \Leftrightarrow c_\alpha = -\frac{1}{n}\log\alpha.$$

Observemos que, como $\alpha \in (0, 1)$, se cumple que $c_\alpha > 0$.
Así pues, finalmente la expresión de la región crítica del test de razón de verosimilitudes para el contraste del enunciado es

$$R = \left\{(x_1, \ldots, x_1) : x_{(1)} > \theta_0 - \frac{1}{n}\log\alpha\right\}.$$

Intuitivamente es una región crítica razonable, pues rechazamos que $\theta \le \theta_0$ cuando la menor de las observaciones de la muestra está demasiado alejada de estos valores de θ. Recordemos que, para un θ fijo el soporte de la densidad $f(\cdot;\theta)$ es precisamente el intervalo $[\theta, \infty)$ y $f(x, \theta)$ es decreciente en x. Así que esperamos que, al muestrear de $f(\cdot, \theta)$, salgan observaciones "justo a la derecha"de θ. Es decir, si estoy contrastando $H_0 : \theta < 3$ y todas las observaciones de la muestra son mucho mayores que $\theta_0 = 3$, intuimos que H_0 es falsa.

Ejercicio 5.5: *Sea $X_1, \dots, X_{16}$ una muestra de tamaño 16 de una población normal de esperanza μ y varianza $\sigma^2 = 1$. Se desea contrastar $H_0 : \mu = 0$ frente a $H_1 : \mu \neq 0$.*

a) *Calcula la región crítica del contraste de razón de verosimilitudes de nivel $\alpha = 0.05$. ¿Qué decisión se toma a nivel $\alpha = 0.05$ si con 16 datos se ha obtenido una media muestral $\overline{x} = 1$?*

b) *Para el contraste anterior, ¿cuál es el valor de la función de potencia evaluada en $\mu = 0.75$?*

Apartado a)

Sea X $\sim N(\mu, 1)$, calculamos la función de verosimilitud:

$$f(x_1, \dots, x_n; \mu) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum(x_i-\mu)^2}$$

Nuestro espacio paramétrico es

$$\Theta_0 = \{\mu = 0\}$$
$$\Theta = \mathbb{R}$$

Entonces el cociente es[9]

$$\Lambda_n = \frac{f(x_1, \dots, x_n; 0)}{f(x_1, \dots, x_n; \overline{x})} = e^{-\frac{1}{2}\left(\sum x_i^2 - \sum(x_i - \overline{x})^2\right)} = e^{-\frac{1}{2}n\overline{x}^2}$$

Y la región de rechazo es

$$R = \{\Lambda_n < k_\alpha\} \text{ donde } k_\alpha \text{ es tal que } \mathbb{P}_{\mu=0}\{R\} = \alpha$$

La región de rechazo se puede expresar (utilizando (IV.4.3.1)) de forma equivalente

$$R = \{-2 \log \Lambda_n > c_\alpha\} = \{n\overline{x}^2 > c_\alpha\}$$

con c_α cumpliendo la misma condición que k_α. Es decir

$$\alpha = \mathbb{P}_{\mu=0}\left\{n\overline{X}^2 > c_\alpha\right\}$$

Sabemos que la distribución de una media de normales es también una normal, luego bajo $H_0 : \mu = 0$, $\overline{X} \sim N(0, \frac{1}{\sqrt{n}})$.
De la misma forma $\sqrt{n}\,\overline{X}^2 \sim N(0, 1)$ y finalmente $n\overline{x}^2 \sim \chi_1^2$. Entonces

$$R = \{n\overline{x}^2 > \chi^2_{1;\alpha}\}$$

[9]recuerda que el EMV$(\mu) = \overline{x}$

A nivel $\alpha = 0.05$, como n = 16 y $\overline{x} = 1$, tenemos que

$$R = \{16 > 3.84\ \}$$

y por lo tanto, hay evidencia muestral para rechazar la hipótesis nula.

Apartado b)

Tenemos que

$$\beta_n(\mu = 0.75) = \mathbb{P}_{\mu=0.75}\left\{\text{rechazar}\, H_0\right\} = \mathbb{P}_{\mu=0.75}\left\{R\right\} = \mathbb{P}_{\mu=0.75}\left\{n\overline{X}^2 > \chi^2_{1;\alpha}\right\}$$

Evaluando

$$n\overline{X}^2 = n(\overline{X} - 0.75 + 0.75)^2 = \underbrace{n(\overline{X} - 0.75)^2}_{\chi^2_1} + 0.75^2 + \underbrace{2(\overline{X} - 0.75)\cdot 0.75}_{\text{nueva v.a.} \to \text{problemón}}$$

Amparo observa que esto nos complica la vida, así que toma otro camino:

$$\mathbb{P}_{\mu=0.75}\left\{n\overline{X}^2 > \chi^2_{1;\alpha}\right\} = 1 - \mathbb{P}_{\mu=0.75}\left\{n\overline{X}^2 \leq \chi^2_{1;\alpha}\right\} = 1 - \mathbb{P}_{\mu=0.75}\left\{\left|\frac{\overline{X}}{1/\sqrt{n}}\right| \leq (\chi^2_{1;\alpha})^{1/2}\right\} =$$

$$\overset{\overline{X}\sim N(0.75, \frac{1}{\sqrt{n}})}{=} 1 - \mathbb{P}_{\mu=0.75}\left\{-(\chi^2_{1;\alpha})^{1/2} \leq \frac{\overline{X}}{1/\sqrt{n}} \leq (\chi^2_{1;\alpha})^{1/2}\right\} = \ldots$$

Con lo que solo nos queda estandarizar y resolver

$$\ldots = 1 - \mathbb{P}_{\mu=0.75}\left\{-(3.84)^{1/2} \leq \frac{\overline{X}}{1/\sqrt{n}} \leq (3.84)^{1/2}\right\} \overset{Z\sim\frac{\overline{X}-\mu}{\sigma},\ \sigma=\frac{1}{\sqrt{n}},\ \mu=0.75}{=}$$

$$= 1 - \mathbb{P}_{\mu=0.75}\left\{\frac{-1.96\frac{1}{\sqrt{n}} - 0.75}{1/\sqrt{n}} \leq Z \leq \frac{1.96\frac{1}{\sqrt{n}} - 0.75}{1/\sqrt{n}}\right\} \overset{n=16}{=}$$

$$= 1 - \mathbb{P}\left\{-4.96 \leq Z \leq -1.04\right\} \overset{*}{=} 1 - (\underbrace{\mathbb{P}\left\{Z > 1.04\right\}}_{\approx 0.15} - \underbrace{\mathbb{P}\left\{Z > 4.96\right\}}_{\approx 0})\approx 0.85$$

(*) Aquí utilizo que la normal es simétrica para poder calcular esa probabilidad con las tablas que tenemos.

Apéndice C

Exámenes

1. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una variable discreta X con función de probabilidad

$$\mathbb{P}_\theta(X = x) = f(x;\theta) = \frac{\theta(-\log\theta)^x}{x!}, \; x = 0, 1, 2, \ldots, \; \theta \in (0, 1). \tag{1}$$

Indicación: $\mathbb{E}_\theta(X) = -\log\theta = V_\theta(X)$.

(a) Calcular el estimador $\hat{\theta}_n$ de θ por el método de máxima verosimilitud, probar que es asintóticamente normal y obtener la distribución asintótica.

(b) Definir la "cantidad de información de Fisher", $I(\theta)$, y explicar muy brevemente su importancia en la teoría estadística. Calcular el valor de $I(\theta)$ para el modelo (1).

(c) Probar que $\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \leq -\log\theta/n$. Deducir de aquí que $n^{1/3}(\hat{\theta}_n - \theta) \xrightarrow{P} 0$.

[3 p.]

2. Una marca de detergente concentrado vende su producto en paquetes cuyo contenido nominal medio es 800 gramos. Se seleccionan al azar 20 paquetes y se obtiene para ellos un contenido medio de 793 gr. con una cuasi-desviación típica de 15. ¿Hay suficiente evidencia estadística, al nivel 0.05, para afirmar que la empresa fabricante vende su producto con un peso medio menor que el valor nominal 800? Indicar si el p-valor del correspondiente contraste es mayor o menor que 0.01. Explicar claramente las suposiciones que se necesiten para garantizar la validez del procedimiento que se utilice. **[2 p.]**

3. Se considera el problema de contrastar $H_0 : \mu \leq 0$ frente a $H_1 : \mu > 0$ a partir de una muestra de tamaño 100 de una $N(\mu, \sigma)$ (con σ conocida). Se utiliza para ello el test cuya región crítica es $R = \{(x_1, \ldots, x_n) : \bar{x} > z_\alpha \frac{\sigma}{\sqrt{n}}\}$, donde $\alpha \in (0, 1)$ es el nivel de significación elegido.

Demostrar que la función de potencia $\beta_\alpha(\mu)$ de este test es una función monótona creciente en $[0, \infty)$. Calcular $\beta_{0.05}(1)$ en el caso en que $\sigma = 2$. **[1.5 p.]**

4. Sea X una v.a. con distribución geométrica de parámetro θ. Esto significa que X es discreta con $\mathbb{P}_\theta(X = x) = (1-\theta)^x\theta$, $x = 0, 1, 2, \ldots$. Se desea estimar θ a partir de una muestra de tamaño n de X, usando la metodología bayesiana. Para ello se supone que la distribución a priori de θ es Beta de parámetros $\alpha > 0$ y $\beta > 0$, es decir que la función de densidad a priori es $\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{I}_{[0,1]}(\theta)$ (la correspondiente media es $\alpha/(\alpha+\beta)$). Calcular el estimador Bayes de θ y estudiar su consistencia casi segura.
Indicación: $\mathbb{E}_\theta(X) = \frac{1-\theta}{\theta}$. **[2.5 p.]**

5. En el directorio de trabajo del R tenemos un fichero con 1000 datos (en dos columnas de 500) llamado `datos.txt`. Redactar un código que realice las siguientes o operaciones:

(a) Leer el fichero `datos.txt`.

(b) Definir un vector llamado x con los valores de la primera columna y otro llamado y con los de la segunda.

(c) Dibujar en un mismo gráfico los dos diagramas de caja de x e y.

(d) Obtener la ecuación de la recta de mínimos cuadrados de y respecto a x (es decir, y debe ser la “variable respuesta”).

[1 p.]

Estadística I

Soluciones a los problemas del examen 10 de enero de 2013

1) (a) El estimador de máxima verosimilitud se obtiene maximizando en θ la función de verosimilitud

$$L_n(X_1,\dots,X_n;\theta) = \frac{\theta^n(-\log\theta)^{\sum X_i}}{X_1!\dots X_n!}.$$

Para maximizar en θ tomamos logaritmos y calculamos la derivada

$$\frac{\partial}{\partial\theta}\log L_n(X_1,\dots,X_n;\theta) = \frac{n}{\theta} + \frac{1}{\theta\log\theta}\sum_{i=1}^{n} X_i$$

La única solución de la ecuación $\frac{\partial}{\partial\theta}\log L_n(X_1,\dots,X_n;\theta) = 0$ es

$$\hat{\theta}_n = e^{-\bar{X}}.$$

Este valor corresponde a un máximo porque la derivada segunda en $\hat{\theta}_n$ es negativa.

Aplicando el TCL sabemos que $\sqrt{n}(\bar{X} - \mathbb{E}_\theta(X)) \xrightarrow{d} N(0,\sigma(\theta))$, siendo $\sigma(\theta) = V_\theta(X)^{1/2}$. Como $\hat{\theta}_n = g(\bar{X})$ con $g(u) = e^{-u}$ y esta función es derivable con derivada continua, podemos aplicar el método delta para obtener

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{d} N(0,|g'(\mu)|\sigma(\theta)),$$

denotando $\mu = \mathbb{E}_\theta(X)$. En definitiva, hemos obtenido

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,\theta\sqrt{-\log\theta}),$$

(observemos que $-\log\theta > 0$ porque $\theta \in (0,1)$).

(b)

$$I(\theta) = \mathbb{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right)$$

Tenemos

$$\frac{\partial}{\partial\theta}\log f(X;\theta) = \frac{1}{\theta} + \frac{1}{\theta\log\theta}X$$

$$\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = \frac{-1}{\theta^2} - \frac{X}{\theta^2\log\theta} - \frac{X}{\theta^2\log^2\theta}$$

$$I(\theta) = \mathbb{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right) = \frac{1}{\theta^2} + \frac{\mathbb{E}_\theta(X)}{\theta^2\log\theta} + \frac{\mathbb{E}_\theta(X)}{\theta^2\log^2\theta} = -\frac{1}{\theta^2\log\theta}$$

La cantidad $I(\theta)$ es importante por varios motivos: bajo ciertas condiciones se verifica $V_\theta(T_n) \geq 1/(nI(\theta))$ (cota de Fréchet-Cramer-Rao) para estimadores insesgados T_n de θ. También (bajo condiciones de regularidad) el estimador de máxima verosimilitud $\hat{\theta}_n$ verifica

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, 1/\sqrt{I(\theta)}),$$

de manera que $I(\theta)^{-1}$ es la varianza de la distribución asintótica. En efecto, obsérvese que en este caso $I(\theta)^{-1}$ coincide con la varianza de la distribución límite obtenida en el apartado anterior.

(c) Si denotamos $g(u) = e^{-u}$,

$$(\hat{\theta}_n - \theta)^2 = (g(\bar{X}) - g(-\log\theta))^2 \leq (\bar{X} - (-\log\theta))^2. \qquad (*)$$

Para obtener esta desigualdad hemos usado el Teorema del Valor Medio, junto con el hecho de que $|g'(u)| = |e^{-u}| \leq 1$ para $u \geq 0$.

Tomando esperanzas,

$$\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2] \leq \mathbb{E}_\theta[(\bar{X} - (-\log\theta))^2] = V_\theta(\bar{X}) = \frac{V_\theta(X)}{n} = \frac{-\log\theta}{n}.$$

Por tanto, dado cualquier $\epsilon > 0$,

$$\mathbb{P}_\theta\{n^{1/3}|\hat{\theta}_n - \theta| > \epsilon\} = \mathbb{P}_\theta\{n^{2/3}(\hat{\theta}_n - \theta)^2 > \epsilon^2\} \leq \text{ (usando la desigualdad de Markov)}$$
$$\leq \frac{n^{2/3}\mathbb{E}_\theta[(\hat{\theta}_n - \theta)^2]}{\epsilon^2} \overset{\text{(usando (*))}}{\leq} \frac{n^{2/3}(-\log\theta)}{n\epsilon^2} = \frac{-\log\theta}{n^{1/3}\epsilon^2} \to 0,$$

lo que demuestra $n^{1/3}(\hat{\theta}_n - \theta) \xrightarrow{P} 0$.

2) Supongamos que la v.a. X = "contenido de un paquete elegido al azar" tiene distribución $N(\mu, \sigma)$. Queremos contrastar

$$H_0 : \mu \geq 800 \text{ frente a } H_1 : \mu < 800.$$

Tenemos una muestra de $n = 20$ observaciones independientes de la v.a. X para la cual $\bar{x} = 793$, $s = 15$. La región crítica del test usual de nivel α para este problema es

$$\bar{x} - 800 < t_{n-1;1-\alpha} \frac{s}{\sqrt{n}}.$$

En este caso, $\bar{x} - 800 = -7$, $t_{n-1;1-\alpha} = t_{19,0.95} = -t_{19;0.05} = -1.729$, $s/\sqrt{n} = 15/\sqrt{20} = 3.354102$.

Por tanto,

$$t_{19;0.05} \frac{s}{\sqrt{n}} = -1.729 \cdot 3.354102 = -5.799242$$

Como $-7 < -5.799241$, se concluye que se ha encontrado suficiente evidencia estadística, al nivel 0.05, para aceptar H_1.

Si consideramos el nivel $\alpha = 0.01$, se tiene $t_{19;0.01} \frac{s}{\sqrt{n}} = -2.539 \cdot 3.354102 = -8.516065$.

Por tanto, al nivel 0.01 NO se ha encontrado evidencia estadística suficiente a favor de H_1.

Se concluye que el p-valor debe de ser mayor que 0.01 ya que el p-valor es el ínfimo de los valores del nivel de significación para los cuales se rechaza la hipótesis nula.

3) La función de potencia es la probabilidad de rechazar la hipótesis nula:

$$\beta_\alpha(\mu) = \mathbb{P}_\mu(R) = \mathbb{P}_\mu \left\{ \bar{X} > z_\alpha \frac{\sigma}{\sqrt{n}} \right\} = \mathbb{P}_\mu \left\{ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \left(z_\alpha \frac{\sigma}{\sqrt{n}} - \mu \right) \Big/ \frac{\sigma}{\sqrt{n}} \right\} = \mathbb{P} \left\{ Z > z_\alpha - 10 \frac{\mu}{\sigma} \right\}$$
$$= 1 - \Phi \left(z_\alpha - 10 \frac{\mu}{\sigma} \right)$$

donde Z es una v.a. $N(0,1)$ y Φ su función de distribución. Aquí hemos usado que, como $X \sim N(\mu, \sigma)$, tenemos que $\bar{X} \sim N\left(\mu, \frac{\sigma}{10}\right)$. Se concluye que $\beta_\alpha(\mu)$ es estrictamente creciente porque Φ lo es (ya que la densidad normal es estrictamente positiva en todo $\mathbb{R}$).

Si $\alpha = 0.05$, $\mu = 1$ y $\sigma = 2$, obtenemos

$$\beta_{0.05}(1) = \mathbb{P} \left\{ Z > z_{0.05} - 10 \frac{1}{2} \right\} = \mathbb{P}\{Z > 1.645 - 5\} \simeq 1.$$

4) La densidad $\pi(\theta|x_1, \ldots, x_n)$ de la distribución a posteriori de θ es proporcional a

$$f(x_1, \ldots, x_n|\theta)\, \pi(\theta),$$

donde

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \mathbb{P}_\theta\{X = x_i\} = \prod_{i=1}^n (1-\theta)^{x_i} \theta = \theta^n (1-\theta)^{\sum_{i=1}^n x_i}$$

es la función de verosimilitud de la muestra. Como

$$f(x_1,\ldots,x_n|\theta)\,\pi(\theta) = \theta^n(1-\theta)^{\sum_{i=1}^n x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\mathbb{I}_{[0,1]}(\theta)$$

y $\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)$ es simplemente una constante de proporcionalidad (no depende de θ), tenemos que $\pi(\theta|x_1,\ldots,x_n)$ es proporcional a

$$\theta^{n+\alpha-1}\,(1-\theta)^{(\sum_{i=1}^n x_i)+\beta-1}\mathbb{I}_{[0,1]}(\theta),$$

que corresponde a una beta de parámetros $n+\alpha$ y $(\sum_{i=1}^n x_i)+\beta$. El estimador Bayes de θ es la esperanza de esta distribución a posteriori:

$$T_n(x_1,\ldots,x_n) = \frac{n+\alpha}{n+\alpha+\sum_{i=1}^n x_i+\beta}.$$

Para probar la consistencia c.s. de T_n, reescribimos el estimador así

$$T_n(x_1,\ldots,x_n) = \frac{1+\frac{\alpha}{n}}{1+\bar{x}+\frac{\alpha+\beta}{n}}.$$

Recordemos que $T_n \xrightarrow[n\to\infty]{\text{c.s.}} \theta$ si y sólo si $1 = \mathbb{P}_\theta\{w : T_n(X_1(w),\ldots,X_n(w)) \xrightarrow[n\to\infty]{} \theta\}$, es decir, si la probabilidad del conjunto en el que se da la convergencia de T_n a θ es uno. Por la ley fuerte de los grandes números sabemos que $\bar{X} \xrightarrow[n\to\infty]{\text{c.s.}} \mathbb{E}_\theta(X) = \dfrac{1+\theta}{\theta}$. Además el denominador de T_n siempre será mayor o igual que 1. Por tanto, para todo w salvo en un conjunto de probabilidad 0, se cumple que

$$T_n(X_1(w),\ldots,X_n(w)) = \frac{1+\frac{\alpha}{n}}{1+\bar{X}(w)+\frac{\alpha+\beta}{n}} \xrightarrow[n\to\infty]{} \frac{1}{1+\frac{1+\theta}{\theta}} = \theta,$$

es decir, $T_n \xrightarrow[n\to\infty]{\text{c.s.}} \theta$.

5)

```
xx<-read.table('datos.txt')
x<-xx$V1
y<-xx$V2
boxplot(x,y)
lm(y ~ x)
```

Estadística I — Examen

Grado en Matemáticas / Doble Grado Matemáticas-Ing. Informática. 14 de junio de 2013

1. La v.a. $X =$ "ingresos (en miles de euros) de un habitante elegido al azar en una cierta ciudad" sigue una distribución de Pareto dada por la siguiente densidad:

$$f(x;\theta) = 3\theta^3 x^{-4}, \quad 0 < \theta < x < \infty.$$

Sea $X_1, \ldots, X_n$ una muestra aleatoria de X. Un economista sugiere que $T_n = \text{mín}(X_1, \ldots, X_n)$ es un posible estimador de θ. ¿Es T_n un estimador consistente de θ? ¿Es insesgado? En caso negativo, calcula el sesgo. **[2 p.]**

2. Sea X una v.a. con distribución Beta$(\theta, 1)$ cuya función de densidad es $f(x;\theta) = \theta x^{\theta-1}\mathbb{I}_{(0,1)}(x)$, para $\theta > 0$.

(a) Calcula la cantidad de información de Fisher $I(\theta)$. Explica brevemente por qué es importante esta cantidad.

(b) Calcula el estimador de máxima verosimilitud de θ (basado en muestras de tamaño n), demuestra que es asintóticamente normal e identifica completamente su distribución asintótica.

(c) Calcula el estimador de θ por el método de los momentos e identifica completamente su distribución asintótica. Demuestra que la correspondiente varianza asintótica es mayor que la obtenida en el apartado (b).

(d) Supongamos ahora que se desea contrastar $H_0 : \theta = 1$ frente a $H_1 : \theta = 2$ a partir de una muestra de tamaño 2, X_1, X_2. Para ello se usa el test de región crítica

$$R = \{(x_1, x_2) : 4x_1x_2 \geq 3\}.$$

Calcula el nivel de significación de este test y la probabilidad de error de tipo 2.

[4 p.]

3. En una encuesta realizada a una muestra aleatoria de 1500 personas, el 43 % de los encuestados se mostraba de acuerdo con endurecer la ley antitabaco.

(a) Calcula el intervalo de confianza de nivel 0.95 para la proporción p de personas en la población que están de acuerdo con endurecer la ley.

(b) Según los resultados obtenidos, ¿existe evidencia estadística suficiente para afirmar que la mayoría de los ciudadanos se opone a endurecer la ley? Para responder a la pregunta, calcula aproximadamente el p-valor del test e interpreta el resultado.

[3p.]

4. Supongamos que se tiene en el directorio de trabajo un fichero llamado `datos` que consiste en una matriz de 200 filas y 10 columnas. Cada fila es una muestra aleatoria de tamaño 10 de la distribución $N(2,1)$. Redacta un código en `R` que calcule las medias y las medianas muestrales de esas 200 muestras, las almacene en dos vectores llamados `medias` y `medianas`, respectivamente, y aproxime los errores cuadráticos medios de ambos estimadores del valor del parámetro $\theta = 2$. **[1 p.]**

Información de posible interés sobre distribuciones:

- **Distribución normal,** $N(\mu, \sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

  Función de densidad: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$, $x \in \mathbb{R}$.

  Momentos: $\mathbb{E}(X) = \mu$, $\quad V(X) = \sigma^2$.

- **Distribución gamma,** $\gamma(a, p)$, $a > 0$, $p > 0$. Cuando $p = 1$ se denomina "distribución exponencial de parámetro a".

  Función de densidad: $f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1}$, para $x > 0$. Aquí $\Gamma(p)$ denota la llamada "función gamma", $\Gamma(p) = \int_0^\infty e^{-x} x^{p-1} dx$ que verifica $\Gamma(p+1) = p\Gamma(p)$ para $p > 0$.

  Momentos: $\mathbb{E}(X) = \frac{p}{a}$, $\quad V(X) = \frac{p}{a^2}$.

- **Distribución beta,** $\text{Beta}(a, b)$, $a > 0$, $b > 0$.

  Función de densidad: $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$, para $x \in (0,1)$.

  Momentos: $\mathbb{E}(X) = \frac{a}{a+b}$, $\; V(X) = \frac{ab}{(a+b+1)(a+b)^2}$.

1. Determinemos primero $\mathbb{E}(T_n)$. Para ello seguimos los siguientes pasos:

$$\mathbb{P}\{X > t\} = \int_t^\infty f(x)\,dx = \int_t^\infty 3\theta^3 x^{-4}\,dx = \frac{\theta^3}{t^3} \quad \text{si } t > \theta$$

$$\mathbb{P}\{T_n > t\} = \mathbb{P}\{\text{mín}(X_1,\ldots,X_n) > t\} = \prod_{i=1}^n P\{X_i > t\} = \frac{\theta^{3n}}{t^{3n}}$$

$$\text{Función de distribución de } T_n\text{: } F_{T_n}(t) = \mathbb{P}\{T_n \le t\} = 1 - \mathbb{P}\{T_n > t\} = 1 - \frac{\theta^{3n}}{t^{3n}}$$

$$\text{Densidad de } T_n\text{: } f_{T_n}(t) = \frac{d}{dt}F_{T_n}(t) = 3n\theta^{3n}\frac{1}{t^{3n+1}} \quad \text{si } t > \theta$$

$$\mathbb{E}(T_n) = \int_\theta^\infty t\, f_{T_n}(t)\,dt = \theta\frac{3n}{3n-1}$$

Como $\mathbb{E}(T_n) \ne \theta$ el estimador T_n está sesgado y su sesgo es

$$\text{Sesgo}(T_n) = \mathbb{E}(T_n) - \theta = \theta\frac{1}{3n-1}.$$

Observemos que $\text{Sesgo}(T_n) \to 0$ cuando $n \to \infty$, así que T_n es asintóticamente insesgado.

Probemos ahora que T_n es consistente en probabilidad: dado $\epsilon > 0$, por la desigualdad de Markov tenemos que

$$\mathbb{P}\{|T_n - \theta| > \epsilon\} \le \frac{\mathbb{E}|T_n - \theta|}{\epsilon} = \frac{\theta}{\epsilon}\frac{1}{3n-1} \to 0 \quad \text{cuando } n \to \infty.$$

2. **a)** La cantidad de información de Fisher, $I(\theta)$, aparece en

- la cota inferior de Fréchet-Cramer-Rao ($\mathbb{V}(T_n) \ge 1/(n\,I(\theta))$ para la varianza de un estimador insesgado, T_n, de θ;
- la varianza asintótica, $1/I(\theta)$, de los estimadores de máxima verosimilitud (bajo ciertas condiciones de regularidad).

Para la distribución Beta(θ,1) tenemos que

$$I(\theta) = \mathbb{E}_\theta\left[-\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right] = \frac{1}{\theta^2},$$

donde hemos utilizado que

$$\log f(x;\theta) = \log\theta + (\theta-1)\log x \quad \text{ y } \quad \frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{1}{\theta} + \log x.$$

b) La función de verosimilitud es

$$L_n(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1}.$$

Para calcular el estimador de máxima verosimilitud (e.m.v.) de θ

$$\hat{\theta}_{\text{MV}} = -\frac{n}{\sum_{i=1}^{n} \log x_i}$$

basta calcular el punto de máximo del logaritmo de la verosimilitud:

$$\log L_n(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i \quad \Rightarrow \quad \frac{d}{d\theta} \log L_n(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log x_i = 0$$

$$\Rightarrow \quad \frac{d^2}{d\theta^2} \log L_n(\theta) = -\frac{n}{\theta^2} < 0$$

Para obtener la distribución asintótica del e.m.v podemos aplicar el teorema sobre la eficiencia asintótica de los e.m.v.:

$$\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta) \xrightarrow{d} N\left(0, \frac{1}{\sqrt{I(\theta_0)}}\right) = N(0, \theta).$$

Aplicando el método delta obtenemos el mismo resultado. Primero observemos que $\hat{\theta}_{\text{MV}} = g(\bar{Y})$, siendo $g(y) = -1/y$ e $Y = \log X$. Por el TCL sabemos que

$$\sqrt{n}(\bar{Y} - \mathbb{E}Y) \xrightarrow{d} N(0, \mathbb{V}^{1/2}(Y)) = N\left(0, \frac{1}{\theta}\right),$$

donde hemos utilizado que

$$\mathbb{E}Y = \int_0^1 (\log x) \theta x^{\theta-1} dx = -\frac{1}{\theta} \quad \text{y} \quad \mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}^2(Y) = \frac{1}{\theta^2}.$$

Ahora aplicamos el método delta:

$$\sqrt{n}(\hat{\theta}_{\text{MV}} - \theta) = \sqrt{n}(g(\bar{Y}) - g(\mathbb{E}Y)) \xrightarrow{d} N(0, |g'(\mathbb{E}Y)| \mathbb{V}^{1/2}(Y)) = N(0, \theta).$$

c) Para obtener el estimador de θ por el método de los momentos igualamos los momentos poblacional y muestral de orden 1 de X:

$$\mathbb{E}X = \int_0^1 x \theta x^{\theta-1} dx = \frac{\theta}{\theta+1} = \bar{X} \Rightarrow \hat{\theta}_{\text{MOM}} = \frac{\bar{X}}{1 - \bar{X}}.$$

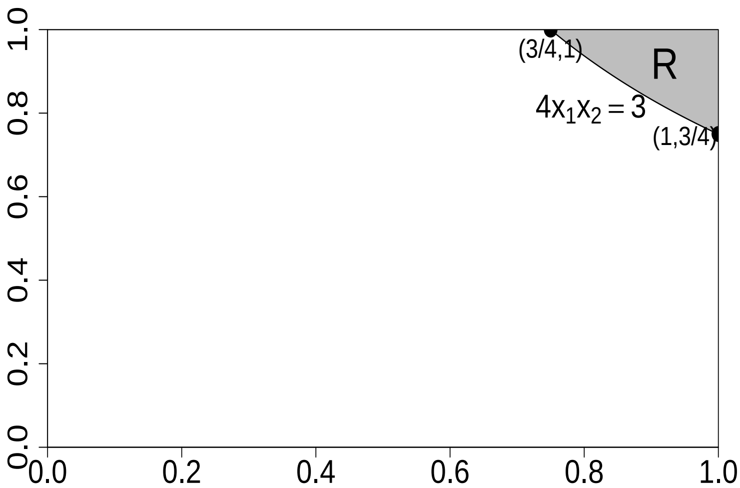
Para determinar la distribución asintótica del estimador aplicamos de nuevo el método delta: $\hat{\theta}_{\text{MOM}} = g(\bar{X})$ y $\theta = g(\mathbb{E}X)$, siendo $g(x) = x/(1-x)$. Por tanto,

$$\sqrt{n}(\hat{\theta}_{\text{MOM}} - \theta) = \sqrt{n}(g(\bar{X}) - g(\mathbb{E}X)) \xrightarrow{d} N(0, |g'(\mathbb{E}X)| \mathbb{V}^{1/2}(X)) = N\left(0, \frac{\theta^{1/2}(\theta+1)}{(\theta+2)^{1/2}}\right)$$

Es fácil comprobar que la varianza asintótica de $\hat{\theta}_{\text{MV}}$ es menor que la de $\hat{\theta}_{\text{MOM}}$:

$$\theta^2 < \frac{\theta(\theta+1)^2}{\theta+2} \Leftrightarrow \theta^2(\theta+2) \leq \theta(\theta+1)^2, \text{lo cual se cumple } \forall \theta > 0.$$

d) Observemos que la región de rechazo del contraste es la que aparece sombreada en la siguiente figura, es decir, $R = \{(x_1, x_2) \in \mathbb{R}^2 : 3/4 \leq x_1 \leq 1, 3/(4x_1) \leq x_2 \leq 1\}$



El nivel de significación o tamaño del test es la máxima probabilidad de error de tipo I, es decir, la máxima probabilidad de rechazar H_0 siendo cierta. En este caso, como la hipótesis nula es puntual, el nivel de significación es simplemente la probabilidad de rechazar H_0 cuando $\theta = 1$. Observemos que, si $\theta = 1$, entonces la función de densidad de X es $f(x; 1) = 1$ si $0 \leq x \leq 1$ y la función de densidad de la muestra X_1, X_2 es $f_{\theta=1}(x_1, x_2) = f(x_1; 1)f(x_2; 1) = 1$ si $0 \leq x_1, x_2 \leq 1$, que corresponde a una distribución uniforme en el cuadrado unidad. Por tanto,

$$\begin{aligned}
\mathbb{P}_{\theta=1}(R) \quad &= \mathbb{P}_{\theta=1}\left\{\frac{3}{4} \leq X_1 \leq 1, \frac{3}{4X_1} \leq X_2 \leq 1\right\} \\
&= \int_{3/4}^{1} \int_{3/(4x_1)}^{1} dx_2\, dx_1 = \frac{1}{4}\left(1 + 3\log\left(\frac{3}{4}\right)\right) \simeq 0.0342
\end{aligned}$$

La probabilidad de error de tipo 2 es la probabilidad de aceptar H_0 siendo falsa, es decir, $1 - \mathbb{P}_{\theta=2}(R)$. Para $\theta = 2$, la función de densidad de la muestra X_1, X_2 es $f_{\theta=1}(x_1, x_2) = f(x_1; 2)f(x_2; 2) = 4x_1x_2$, si $0 \leq x_1, x_2 \leq 1$. Luego

$$\mathbb{P}_{\theta=2}(R) = \int_{3/4}^{1} \int_{3/(4x_1)}^{1} 4x_1x_2\, dx_2\, dx_1 = \frac{1}{8}\left(\frac{7}{2} + 9\log\left(\frac{3}{4}\right)\right) \simeq 0.1139$$

y, en consecuencia, $1 - \mathbb{P}_{\theta=2}(R) \simeq 0.8861$.

3. Sea

$$X = \begin{cases} 1 & \text{si un encuestado es partidario de endurecer la ley} \\ 0 & \text{si no} \end{cases}$$

que sigue una distribución de Bernoulli(p) con $0 < p < 1$. Se ha tomado una muestra $x_1, \ldots, x_{1500}$ que ha proporcionado el dato $\bar{x} = 0.43$.

a)

$$\mathrm{IC}_{0.95}(p) = \left[0.43 \mp 1.96\sqrt{\frac{0.43(1 - 0.43)}{1500}}\right] = [0.043 \mp 0.025] = [0.405, 0.455]$$

b) Planteamos el contraste

$$\begin{array}{ll} H_0: & p \geq 0.5 \\ H_1: & p < 0.5 \quad \text{(la mayoría de los ciudadanos se opone a endurecer la ley),} \end{array}$$

cuya región de rechazo es

$$R = \left\{ \bar{x} - 0.5 < z_{1-\alpha}\sqrt{\frac{0.5^2}{n}} \right\} = \{z < z_{1-\alpha}\} = \{-z > z_\alpha\},$$

siendo

$$z = \frac{\bar{x} - 0.5}{\sqrt{0.5^2/n}} = -5.42$$

el estadístico del contraste. El p-valor del contraste es la probabilidad de que una $N(0,1)$ sea mayor que 5.42. Con la información de la tabla ($\mathbb{P}\{Z > 3.99\} = 0.0010$) llegamos a la conclusión de que el p-valor es menor que 0.0010. Utilizando R (`pnorm(-5.42)`) obtenemos que el p-valor es 2.979952e-08: es razonable rechazar la hipótesis nula.

4.

```
medias = apply(datos,1,mean)
medianas = apply(datos,1,median)
ECMmedia = (mean(medias)-2)^2 + var(medias)
# Se ha usado que ECM(T)=Sesgo^2(T)+V(T)
ECMmediana = (mean(medianas)-2)^2 + var(medianas)
```

Otro código alternativo (sin usar la función `apply` sino un `for`, y utilizando directamente la definición de ECM: $ECM_\theta(T) = \mathbb{E}[(T-\theta)^2]$), sería

```
medias<-rep(0,200)
medianas<-rep(0,200)
for (i in 1:200){medias[i]<-mean(datos[i,])}
for (i in 1:200){medianas[i]<-median(datos[i,])}
ECMmedia<-mean((medias-2)^2)
ECMmediana<-mean((medianas-2)^2)
```

ESTADÍSTICA I (2013-2014)
Grado en Matemáticas / Doble grado Ing. Informática/Matemáticas
Examen final, 18 de enero de 2014

Nombre: ______________________________

Grupo: ______________________

1. Se desea comparar la concentración observada de tiol (mM) en el lisado sanguíneo de dos grupos de voluntarios, siendo el primer grupo "normal" (X) y padeciendo el segundo grupo de artritis reumatoide (Y). Para ello se analizan los datos con R de la siguiente manera

```
> X = c(1.84, 1.92, 1.94, 1.92, 1.85, 1.91, 2.07)
> Y = c(2.81, 4.06, 3.62, 3.27, 3.40, 3.76)
> t.test(X,Y,alternative="two.sided",mu=0,paired=FALSE,var.equal=FALSE)

Welch Two Sample t-test

data:  X and Y
t = -8.759, df = 5.263, p-value = 0.0002473
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of x mean of y
 1.921429  3.486667
```

a) **(1 punto)** ¿Qué contraste se está haciendo? Especificar las hipótesis necesarias para garantizar la validez del método empleado. ¿Qué conclusiones se obtienen acerca del contraste?

b) **(1 punto)** Calcular un intervalo de confianza al 95 % para la diferencia de concentraciones medias de tiol entre los dos grupos. ¿Qué relación hay entre este intervalo y el contraste de (a)?

2. Sea

$$f(x;\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0,$$

la función de densidad de una v.a. X con distribución beta de parámetros θ y 1.

a) **(1.5 puntos)** Consideremos el contraste de hipótesis

$$\begin{array}{ll} H_0: & \theta = 1 \\ H_1: & \theta = 2. \end{array}$$

Dada una muestra X_1 de tamaño $n = 1$ de X, determina la región de rechazo del test más potente con nivel de significación α. Para $\alpha = 0.05$ calcula la función de potencia de ese test. Indicación: si $X \sim \text{beta}(\theta, 1)$, entonces $Y = -\log(X)$ sigue una distribución exponencial de parámetro θ, es decir, la densidad de Y es $g(y) = \theta e^{-\theta y}$, $y > 0$.

b) **(1.5 puntos)** A nivel de significación α, ¿cuál sería la región de rechazo del test de razón de verosimilitudes para el siguiente contraste?:

$$\begin{array}{ll} H_0: & \theta = 1 \\ H_1: & \theta \neq 1 \end{array}$$

Empleando la tabla de la χ^2, ¿hay evidencia para rechazar H_0 a nivel $\alpha = 0.05$ si, para una muestra de tamaño $n = 50$, hemos obtenido $\sum_{i=1}^{50} \log(x_i) = -19.342$?

3. Sea $\beta > 0$ un número conocido. Sea $x_1, \ldots, x_n$ una muestra de una variable aleatoria X con distribución Weibull de función de densidad

$$f(x;\theta) = \theta\,\beta\,x^{\beta-1}e^{-\theta\,x^{\beta}}, \qquad x > 0, \quad \theta > 0.$$

a) **(0.5 puntos)** Calcular el estimador de θ por el método de los momentos.

b) **(1 punto)** Calcular el estimador de máxima verosimilitud (e.m.v.) de θ.

c) **(1 punto)** Determinar la cantidad de información de Fisher $I(\theta)$.

d) **(2 puntos)** Estudiar la consistencia y la normalidad asintótica del e.m.v. determinado en (b).

e) **(0.5 puntos)** Define el concepto de estimador eficiente. Estudia la eficiencia del e.m.v. de θ determinado en (b).

Indicación: Para cualquier entero positivo m, $\mathbb{E}(X^m) = \dfrac{1}{\theta^{m/\beta}}\Gamma\left(1 + \dfrac{m}{\beta}\right)$, donde $\Gamma(t) = \displaystyle\int_0^\infty x^{t-1}e^{-x}dx$ es la función gamma, y $\Gamma(n) = (n-1)!$ si n es un entero positivo.

ESTADÍSTICA I (2013-2014)
Grado en Matemáticas / Doble grado Ing. Informática/Matemáticas
Examen final, 18 de enero de 2014. SOLUCIONES

1. a) Se supone que $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ independientes, con $\sigma_1 \neq \sigma_2$. El contraste es $H_0 : \mu_1 = \mu_2$ frente a $H_1 : \mu_1 \neq \mu_2$. La región de rechazo de este contraste es

$$R = \left\{ |\bar{x} - \bar{y}| > t_{f;\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right\} = \{|t| > t_{f;\alpha/2}\},$$

donde

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = -8.759$$

es el estadístico del contraste y $f = 5$ es el entero más próximo a 5.263 (los grados de libertad, `df`). Según la salida de R, el p-valor del contraste es 0.0002473. Por tanto, es razonable rechazar la hipótesis nula. Concluimos que la concentración esperada de tiol es distinta en el grupo normal y en el grupo con artritis reumatoide.

b) Bajo las mismas hipótesis que en (1a), el intervalo pedido es

$$\text{IC}_{95\,\%}(\mu_1 - \mu_2) = \left(\bar{x} - \bar{y} \mp t_{5,0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right).$$

Como $\bar{x} = 1.921429$, $\bar{y} = 3.486667$ y $t = -8.759$, tenemos que $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{\bar{x}-\bar{y}}{t} = 0.1787$. Por tanto,

$$\text{IC}_{95\,\%}(\mu_1 - \mu_2) = (-1.565238 \mp 2.571 \cdot 0.1787) = (-2.024676, -1.105800).$$

El intervalo no contiene al 0, luego rechazamos la hipótesis nula simple $H_0 : \mu_1 = \mu_2$ al nivel $\alpha = 0.05$, pues la región de rechazo R de (a) equivale a rechazar H_0 cuando $0 \notin \text{IC}_{1-\alpha}(\mu_1 - \mu_2)$.

2. a) Por el lema de Neyman-Pearson, el test más potente es el que tiene región de rechazo

$$R = \left\{ \frac{f_n(x_1, \ldots, x_n; \theta = 2)}{f_n(x_1, \ldots, x_n; \theta = 1)} > k_\alpha \right\},$$

donde k_α se elige de tal manera que $\mathbb{P}_{\theta=1}(R) = \alpha$ y $f_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^n \left(\prod_{i=1}^n x_i\right)^{\theta-1}$, si $0 \leq x_1, \ldots, x_n \leq 1$, es la función de verosimilitud de la muestra. Como

$$\frac{f_n(x_1, \ldots, x_n; \theta = 2)}{f_n(x_1, \ldots, x_n; \theta = 1)} = 2^n \prod_{i=1}^n x_i,$$

tenemos que $R = \{2^n \prod_{i=1}^n x_i > k_\alpha\}$. Si $n = 1$, entonces $R = \{2X_1 > k_\alpha\} = \{-\log X_1 < c_\alpha\}$, donde c_α es una constante tal que

$$\alpha = \mathbb{P}_{\theta=1}(R) = 1 - \frac{k_\alpha}{2}. \tag{1}$$

En la última igualdad de (1) se ha utilizado que, si $\theta = 1$, X_1 sigue una distribución uniforme en [0,1]. Despejando en (1) obtenemos $k_\alpha = 2(1-\alpha)$ (también se podía utilizar la indicación del enunciado para obtener $c = -\log(1-\alpha)$). Por tanto, si $n = 1$, $R = \{X_1 > 1 - \alpha\}$.
Si $\alpha = 0.05$, entonces $R = \{X_1 > 0.95\}$. La función de potencia es la probabilidad de rechazar la hipótesis nula: $\beta(\theta) = \mathbb{P}_\theta(R) = \mathbb{P}_\theta\{-\log X_1 < -\log 0.95\} = 1 - 0.95^\theta$. Si $\theta = 1$, obviamente $\beta(1) = 0.05$. Si $\theta = 2$, $\beta(2) = 0.0975$.

b) El estadístico del contraste de razón de verosimilitudes para el contraste propuesto es

$$\Lambda_n = \frac{f_n(X_1, \ldots, X_n; \theta = 1)}{f_n(X_1, \ldots, X_n; \hat{\theta})} = \frac{1}{\hat{\theta}^n} \left(\prod_{i=1}^n x_i \right)^{1-\hat{\theta}},$$

siendo $\hat{\theta} = -n / \sum_{i=1}^n \log X_i$ el estimador de máxima verosimilitud (e.m.v.) de θ. La región de rechazo de un test con nivel aproximado α es $R = \{-2 \log \Lambda_n > \chi^2_{1;\alpha}\}$. Es sencillo comprobar que $-2 \log \Lambda_n = 2n(\log \hat{\theta} + \frac{1}{\hat{\theta}} - 1)$. Si $\alpha = 0.05$, $n = 50$ y $\sum_{i=1}^{50} \log(x_i) = -19.342$, entonces $\hat{\theta} = 2.59$, $-2 \log \Lambda_n = 33.66$ y $\chi^2_{1;0.05} = 3.84$, luego rechazamos la hipótesis nula.

3. a) El estimador de los momentos

$$\tilde{\theta} = \left(\frac{1}{\bar{X}} \Gamma \left(1 + \frac{1}{\beta} \right) \right)^\beta$$

se obtiene igualando los momentos poblacionales y muestrales de orden 1:

$$\mathbb{E}X = \frac{1}{\theta^{1/\beta}} \Gamma \left(1 + \frac{1}{\beta} \right) = \bar{X}.$$

b) Función de verosimilitud: $L(\theta; x_1, \ldots, x_n) = \theta^n \beta^n \left(\prod_{i=1}^n x_i \right)^{\beta - 1} e^{-\theta \sum_{i=1}^n x_i^\beta}$

Función de logverosimilitud: $\log L(\theta) = n \log \theta + \log \left(\beta^n \left(\prod_{i=1}^n x_i \right)^{\beta-1} \right) - \theta \sum_{i=1}^n x_i^\beta$

Para hallar el punto de máximo de la logverosimilitud:

$$\frac{\partial}{\partial \theta} \log L(\theta) = \frac{n}{\theta} - \sum_{i=1}^n x_i^\beta = 0,$$

de donde obtenemos que $\hat{\theta} = \text{e.m.v.}(\theta) = n / \sum_{i=1}^n X_i^\beta$.

c)

$$I(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log(f(X; \theta)) \right)^2 = -\mathbb{E}_\theta \left(\frac{\partial^2}{\partial^2 \theta} \log(f(X; \theta)) \right) = \frac{1}{\theta^2}$$

d) Observemos que $\hat{\theta} = \dfrac{n}{\sum_{i=1}^n X_i^\beta} = \dfrac{1}{\bar{Y}}$, donde $Y_1, \ldots, Y_n$ es una muestra de la v.a. $Y = X^\beta$.

Por la ley fuerte de los grandes números, sabemos que $\bar{Y} \xrightarrow{\text{c.s.}} \mathbb{E}(Y) = \mathbb{E}(X^\beta) = \dfrac{1}{\theta}$. Sea $g(x) = 1/x$. Por el teorema de la aplicación continua, $\hat{\theta} = g(\bar{Y}) \xrightarrow{\text{c.s.}} g(\mathbb{E}(Y)) = \theta$. Por lo tanto, el e.m.v. de θ es consistente c.s.

Para demostrar la normalidad asintótica de $\hat{\theta}$ utilizamos el método delta:

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n} \left(\frac{1}{\bar{Y}} - \frac{1}{\mathbb{E}Y} \right) = \sqrt{n}(g(\bar{Y}) - g(\mathbb{E}Y)) \xrightarrow[n \to \infty]{\text{d}} N(0, |g'(\mathbb{E}Y)| \sqrt{\mathbb{V}(Y)}) = N(0, \theta).$$

En la última igualdad hemos utilizado que $\mathbb{V}(Y) = \mathbb{E}(X^{2\beta}) - \mathbb{E}^2(X^\beta) = 1/\theta^2$.

e) Un estimador T_n de θ es eficiente si es insesgado ($\mathbb{E}(T_n) = \hat{\theta}$) y su varianza alcanza la cota de Fréchet-Cramer-Rao: $\mathbb{V}(T_n) = \dfrac{1}{nI(\theta)} = \dfrac{\theta^2}{n}$. El e.m.v. de θ no es necesariamente insesgado ($\mathbb{E}(1/X) \neq 1/\mathbb{E}(X)$) y, por tanto, no podemos decir si es eficiente, pero sí es asintóticamente eficiente porque $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n \to \infty]{\text{d}} N(0, 1/\sqrt{I(\theta)})$.

Índice alfabético

Asintóticamente
- insesgado, 20
- normal, 25

Box-plot, 4

Cantidad
- pivotal, 48

Coeficiente
- de asimetría, 4
- de correlación lineal, 10
- de Pearson, 10

Condición
- de Borel-Cantelli, 24

Consistencia
- casi segura, 23
- en probabilidad, 23
- fuerte, 23

Convergencia
- casi segura, 13
- débil, 12
- en distribución, 12
- en probabilidad, 13

Cota
- de Fréchet-Cramér-Rao, 35

Covarianza muestral, 9

Cuantil, 4
- muestral, 21
- poblacional, 21

Cuartil, 4

Cuasivarianza
- muestral, 20

Datos emparejados, 129

Desigualdad
- de Chebichev, 14
- de Jensen, 31
- de Markov, 14

Desviación
- típica, 4

Diagrama
- de dispersión, 8

Distribución, 11
- F de Fisher, 55
- χ^2, 48
- t de Student, 49
- a posteriori, 44
- a priori, 43

Ecuación
- de verosimilitud, 26

Error
- de tipo I, 50
- de tipo II, 50
- estándar, 18
- típico, 18

Espacio
- paramétrico, 22

Esperanza, 11

Estadístico, 18
- de Kolmogorov-Smirnov, 15
- de contraste, 53
- de orden, 21
- del contraste de razón de verosimilitudes, 60

Estimador, 22
- Bayes, 44
- centrado, 19
- de máxima verosimilitud, 26
- eficiente, 37
- insesgado, 19, 23
- núcleo, 6
- por el método de los momentos, 42

Familia
- conjugada, 46
- paramétrica CVM, 59

Función
- cuantílica, 21
- de distribución, 11

de distribución empírica, 15
de potencia, 50
de verosimilitud, 26
indicatriz, 6

Histograma, 5

Información
de Fisher, 35
Intervalo
de confianza, 46
Invarianza del EMV, 30

Límite
inferior, 5
superior, 5
Lema
de Fischer-Cochran, 49
de Neyman-Pearson, 58
Ley
de los grandes números, 15

Método
delta, 25
Media, 3
de una distribución, 11
muestral, 18
poblacional, 18
Mediana, 3
Momento, 12
Muestra, 11
homocedástica, 54

Nivel
de significación, 51
Normalidad
asintótica, 25

p-valor del contraste, 52

Rango
intercuartílico, 4
Recta de regresión, 9
Región
creíble, 49
Regresión lineal
coeficiente de, 9
Residuo, 9

Skewness, 4
Soporte, 31
Sucesión
consistente, 57

Tamaño
de un test, 51
Teorema
central del límite, 19
de Bayes, 44
de cambio de espacio de integración, 12
de Glivenko-Cantelli, 15
de la aplicación continua, 23
de Slutsky, 14
MV1, 31
MV2, 33
MV3, 37
Test
óptimo, 58, 60
Bayesiano, 64
de bondad de ajuste, 61
de cociente de verosimilitudes, 60
insesgado, 57
UMP, 57

Varianza, 3
combinada, 55
muestral, 20
residual, 9
Ventana móvil, 6