# *Inferencia Estadística*

# *Apuntes*

**Andrés Herrera Poyatos**
Universidad de Granada
andreshp9@gmail.com

# Índice

# 1. Familias de distribuciones

## 1.1. Distribuciones discretas

## 1.2. Distribuciones continuas

### 1.2.1. Distribución uniforme

La distribución uniforme asigna una credibilidad uniforme a todos los puntos de un intervalo $[a, b]$. Esto es, su función de densidad viene dada por

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} \text{ si } x \in [a,b], \\ 0 \text{ en otro caso.} \end{cases}$$

Claramente tenemos que $\int_{-\infty}^{\infty} f(x|a,b)dx = 1$. Además, podemos calcular fácilmente sus momentos como sigue (y, por tanto, también su varianza)

$$E[X^j] = \int_a^b \frac{x^j}{b-a}dx = \frac{b^{j+1} - a^{j+1}}{(b-a)(j+1)},$$

$$Var(X) = E[X^2] - E[X]^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a + b^2)}{4} = \frac{(b-a)^2}{12}.$$

### 1.2.2. Distribución normal

La distribución normal, también llamada distribución gaussiana, es la distribución más importante de la estadística. Esto se debe a sus numerosas aplicaciones en análisis de poblaciones y al teorema central del límite.

**Definición 1.** Sean $\mu \in \mathbb{R}$ y $\sigma^2 > 0$. Definimos la distribución $N(x|\mu, \sigma^2)$ como la distribución que tiene función de densidad

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, x \in \mathbb{R}.$$

La distribución normal está bien definida como consecuencia del siguiente lema.

**Lema 1.** *Sean $\mu \in \mathbb{R}$ y $\sigma > 0$. Tenemos que $\int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)}dx = \sqrt{2\pi}\sigma$.*

*Demostración.* En primer lugar, vamos a calcular la integral para $\mu = 0$ y $\sigma = 1$. La demostración consiste en reducir el problema en calcular una integral en dos variables. Para ello, elevamos al cuadrado y obtenemos

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2}dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-t^2/2}dt\right)\left(\int_{-\infty}^{\infty} e^{-s^2/2}ds\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(t^2+s^2)/2}dtds.$$

Resolvemos esta última integral mediante un cambio a polares

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(t^2+s^2)/2}dtds = \int_{-\pi}^{\pi}\left(\int_0^{\infty} \rho e^{-\rho^2/2}d\rho\right)d\theta = 2\pi \int_0^{\infty} \rho e^{-\rho^2/2}d\rho = 2\pi.$$

Por último, utilizamos el cambio de variable $y = (x-\mu)/\sigma$ para obtener

$$\int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)}dx = \int_{-\infty}^{\infty} \sigma e^{-y^2/2}dy = \sqrt{2\pi}\sigma. \quad \square$$

Nótese que si $X \sim N(x|\mu, \sigma^2)$, entonces $Y = (X - \mu)/\sigma$ sigue una distribución $N(x|0, 1)$.

**Proposición 2.** *La función característica de la distribución $N(x|\mu, \sigma^2)$ viene dada por $\varphi_X(t) = e^{it\mu - t^2\sigma^2/2}$.*

*Demostración.* En primer lugar, tenemos que

$$\varphi_X(t) = E[e^{itX}] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{itx-(x-\mu)^2/(2\sigma^2)}dx = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-((x-\mu)^2-2itx\sigma^2)/(2\sigma^2)}dx.$$

Completamos cuadrados como sigue

$$(x-\mu)^2 - 2itx\sigma^2 = (x - (it\sigma^2 + \mu))^2 + t^2\sigma^4 - 2it\sigma^2\mu.$$

Esto sugiere utilizar el cambio de variable $g(y) = y + it\sigma^2$. Obtenemos

$$\begin{aligned}\sqrt{2\pi}\sigma\varphi_X(t) = \int_{-\infty}^{\infty} e^{-((x-\mu)^2-2itx\sigma^2)/(2\sigma^2)} &= e^{it\mu - t^2\sigma^2/2}\int_{-\infty}^{\infty} e^{-((x-(it\sigma^2+\mu))^2)/(2\sigma^2)}dx \\ &= e^{it\mu-t^2\sigma^2/2}\int_{-\infty}^{\infty} e^{-(y-\mu)^2)/(2\sigma^2)}dy = \sqrt{2\pi}\sigma e^{it\mu-t^2\sigma^2/2},\end{aligned}$$

como se quería. Nótese que a pesar de ser una integral de contorno compleja el cambio de variable es válido. En efecto, el cambio de variable es afín y la función a integrar es entera. Por tanto, utilizando el camino cerrado $g([0, \infty]) + [\infty, 0]$ se puede probar que el cambio es válido. □

Análogamente se puede probar el siguiente resultado.

**Proposición 3.** *La función generatriz de momentos de la distribución $N(x|\mu, \sigma^2)$ viene dada por $\varphi_X(t) = e^{t\mu - t^2\sigma^2/2}$.*

**Corolario 4.** *Los momentos de la distribución $N(x|\mu, \sigma^2)$ verifican la ecuación recurrente*

$$E[X^k] = -(k-1)\sigma^2 E[X^{k-2}] + (\mu - t\sigma^2)E[X^{k-1}], \quad k \geq 2.$$

*Demostración.* Sabemos que $E[X^k] = \varphi_X^{(k)}(t)$. Tenemos $\varphi_X^{(1)}(t) = (\mu - t\sigma^2)\varphi_X(t)$. Consecuentemente,

$$\varphi_X^{(2)}(t) = -\sigma^2\varphi_X(t) + (\mu - t\sigma^2)\varphi_X^{(1)}(t).$$

Por inducción se extiende el resultado fácilmente para $k \geq 2$. □

**Corolario 5.** *Si $X \sim N(x|\mu, \sigma^2)$, entonces $E[X] = \mu$ y $E[X^2] = \sigma^2 + \mu^2$. Consecuentemente, $Var(X) = \sigma^2$. Como consecuencia de este resultado al parámetro $\mu$ se le llama media y al parámetro $\sigma^2$ varianza.*

Podemos utilizar los dos corolarios anteriores para calcular los momentos de la distribución normal resolviendo una ecuación recurrente de segundo orden. Evidentemente, la fórmula obtenida será bastante larga. Sin embargo, esta ecuación se simplifica en el caso de los momentos centrados, como pone de manifiesto el siguiente resultado, que se puede demostrar fácilmente por inducción a partir del Corolario 4.

**Corolario 6.** *Si $X \sim N(x|0, \sigma^2)$, entonces*

$$E[X^k] = \begin{cases} 0 & \text{si } k \text{ es impar;} \\ (k-1)!!\sigma^k & \text{si } k \text{ es par;} \end{cases}$$

*donde $n!!$ denota al doble factorial, definido como el producto de los números desde 1 hasta $n$ con la misma paridad que $n$.*

La Figura 1 muestra la función de densidad de una distribución normal. Podemos ver que la densidad se concentra en torno a la media. De hecho, $P(|X-\mu| \geq 2\sigma) \approx 0{,}046$. Es más, $P(|X-\mu| \geq 3\sigma) \approx 0{,}03$.

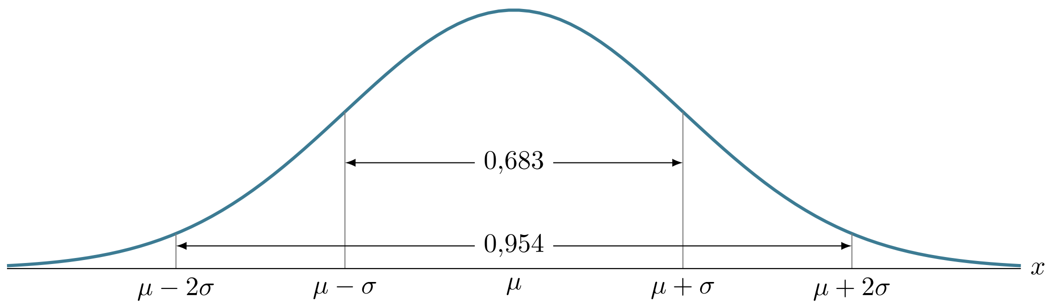


Figura 1: Función de densidad de una distribución normal.

**Proposición 7.** *Sean $X_1$ e $Y_2$ dos variables aleatorias independientes que siguen una distribución $N(x|\mu_1, \sigma_1^2)$ y $N(x|\mu_2, \sigma_2^2)$ respectivamente. Entonces $X+Y$ sigue una distribución $N(x|\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.*

*Demostración.* Basta darse cuenta de que $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{t(\mu_1+\mu_2)-t^2(\sigma_1^2+\sigma_2^2)/2}$ es la función característica asociada a la distribución $N(x|\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$. Recordemos que la función característica determina de forma unívoca a la distribución. □

El recíproco del resultado anterior también es cierto.

**Teorema 8** (Cramer)**.** *Sean $X$ e $Y$ dos variables aleatorias independientes. Si $X+Y$ es normal, entonces $X$ e $Y$ son normales.*

### 1.2.3. Distribución gamma

La famila de distribuciones gamma se encuentra definida sobre el intervalo $[0, \infty)$. En su definición entra en juego la famosa función gamma, de ahí su nombre.

**Definición 2.** Se define la función gamma como la aplicación $\Gamma : (0, \infty) \to (0, \infty)$ dada por

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

**Proposición 9.** *La función gamma está bien definida.*

*Demostración.* Sea $\alpha > 0$. Tenemos que probar que $\int_0^\infty t^{\alpha-1}e^{-t}dt < \infty$. Tomando $b > 0$, escribimos

$$\int_0^\infty t^{\alpha-1}e^{-t}dt = \int_0^b t^{\alpha-1}e^{-t}dt + \int_b^\infty t^{\alpha-1}e^{-t}dt.$$

Sabemos que la función $t^{\alpha-1}$ tiene a $t^\alpha/\alpha$ como primitiva y, por tanto, es integrable en $[0, b]$. Puesto que $t^{\alpha-1}e^{-t} \leq t^{\alpha-1}$, obtenemos que $t^{\alpha-1}e^{-t}$ es integrable en $[0, b]$. Por otro lado tenemos que

$$\lim_{t\to\infty} \frac{t^{\alpha-1}e^{-t}}{e^{-t/2}} = 0.$$

Consecuentemente, para cierto $b > 0$ se verifica $t^{\alpha-1}e^{-t} \leq e^{-t/2}$ para todo $t \geq b$. Puesto que $e^{-t/2}$ es integrable en $[b, \infty)$, deducimos que $t^{\alpha-1}e^{-t}$ también lo es, lo que termina la demostración. □

**Proposición 10** (Propiedades de la función gamma)**.** *Sea $\alpha > 0$. Se verifica:*

*a)* $\Gamma(1) = 1$*;*

*b)* $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$*;*

*c)* $\Gamma(n + 1) = n!$ *para cualquier* $n \in \mathbb{N}$*;*

*d) (Fórmula de reflexión de Euler) si* $0 < \alpha < 1$*, entonces* $\Gamma(\alpha)\Gamma(1 - \alpha) = \frac{\pi}{\sin(\alpha\pi)}$*;*

*e)* $\Gamma(1/2) = \sqrt{\pi}$*;*

*f)* $\Gamma(\alpha) = \beta^\alpha \int_0^\infty t^{\alpha-1}e^{-\beta t}dt$ *para todo* $\beta > 0$*.*

*Demostración.*

a) Es fácil ver que $\int_0^\infty e^{-t}dt = 1$.

b) Integrando por partes obtemos

$$\Gamma(\alpha + 1) = \int_0^\infty t^\alpha e^{-t}dt = \Big[ - e^{-t}t^\alpha \Big]_0^\infty + \int_0^\infty xt^{\alpha-1}e^{-t}dt = \alpha \int_0^\infty t^{\alpha-1}e^{-t}dt = \alpha\Gamma(\alpha).$$

c) Es consecuencia directa de los apartados a) y b).

d) Se obtiene utilizando definiciones alternativas de la función gamma tras extenderla a $\mathbb{C} \setminus \mathbb{Z}_0^-$. Para más información véase [1]. No desarrollamos esta demostración pues solo la necesitamos para el siguiente apartado.

e) Se obtiene al evaluar la fórmula de reflexión en $\alpha = 1/2$.

f) Se obtiene realizando el cambio de variable $t = \beta s$. □

**Definición 3.** Sean $\alpha, \beta > 0$. Definimos la distribución $Gamma(x|\alpha, \beta)$ como la distribución que tiene función de densidad

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

El parámetro $\alpha$ se conoce como parámetro de forma ya que influencia la forma de la distribución, como muestra el siguiente resultado.

**Proposición 11.** *La función de densidad de la distribución* $Gamma(\alpha, \beta)$ *verifica las siguientes propiedades:*

- *Si* $0 < \alpha < 1$*, entonces* $f(x|\alpha, \beta)$ *es decreciente y* $f(x) \to \infty$ *para* $x \to 0$*.*
- *Si* $\alpha = 1$*, entonces* $f(x|\alpha, \beta)$ *es decreciente con* $f(0) = 1$*.*
- *Si* $\alpha > 1$*, entonces* $f(x|\alpha, \beta)$ *crece en* $[0, (\alpha - 1)/\beta]$ *y decrece en* $[(\alpha - 1)/\beta, \infty]$*.*
- *Si* $0 < \alpha \leq 1$*, entonces* $f(x|\alpha, \beta)$ *es convexa.*
- *Si* $1 < \alpha \leq 2$*, entonces* $f(x|\alpha, \beta)$ *es cóncava en* $[0, (\alpha - 1 + \sqrt{\alpha - 1})/\beta]$ *y convexa en* $[(\alpha - 1 + \sqrt{\alpha - 1})/\beta, \infty]$*.*
- *Si* $2 < \alpha$*, entonces* $f(x|\alpha, \beta)$ *es cóncava en* $[(\alpha - 1 - \sqrt{\alpha - 1})/\beta, (\alpha - 1 + \sqrt{\alpha - 1})/\beta]$ *y convexa en* $[0, (\alpha - 1 - \sqrt{\alpha - 1})/\beta]$ *y* $[(\alpha - 1 + \sqrt{\alpha - 1})/\beta, \infty]$*.*

*Demostración.* Los resultados se obtienen mediante las herramientas habituales del cálculo. Basta estudiar la derivada primera y la derivada segunda

$$f'(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-2} e^{-\beta x}[(\alpha-1) - \beta x];$$
$$f''(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-3} e^{-\beta x}\left[(\alpha-1)(\alpha-2) - 2\beta(\alpha-1)x + \beta^2 x^2\right]. \qquad \square$$

La Figura 2 muestra la función de densidad de la distribución gamma para distintos valores de $\alpha$. El parámetro $\beta$ se denomina parámetro de escala debido a su influencia en la escala de la función de densidad. La Figura 3 muestra la función de densidad de la distribución gamma para distintos valores de $\beta$.

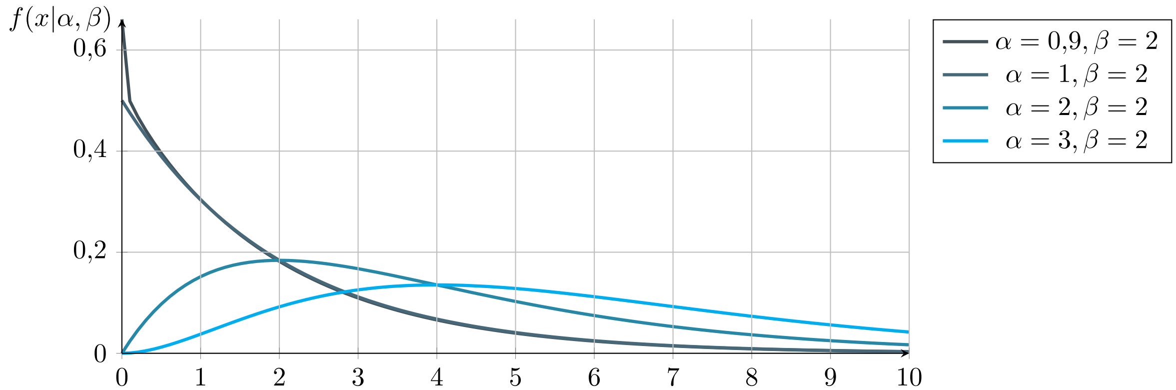


Figura 2: Densidad de la distribución gamma con distintos valores de $\alpha$.

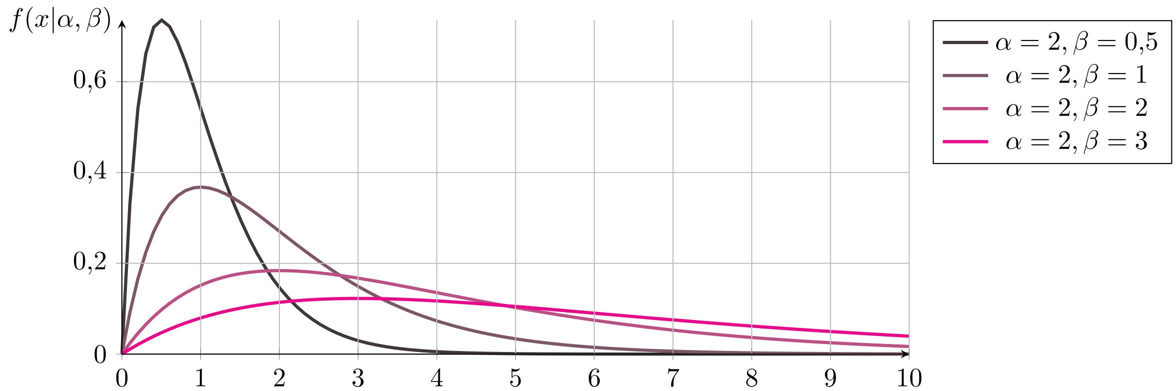


Figura 3: Densidad de la distribución gamma con distintos valores de $\beta$.

**Proposición 12.** *La función característica de la distribución Gamma*$(x|\alpha,\beta)$ *viene dada por* $\varphi_X(t) = \left(\frac{\beta}{\beta - it}\right)^\alpha$.

*Demostración.* Basta utilizar el cambio de variable $g(y) = y/(\beta - it)$ como sigue

$$E[e^{itX}] = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\beta-it)x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)(\beta-it)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy = \left(\frac{\beta}{\beta-it}\right)^\alpha.$$

Nótese que a pesar de ser una integral de contorno compleja el cambio de variable afín es válido como se comentó en la Proposición 2. □

**Corolario 13.** *El momento k-ésimo de la distribución Gamma$(x|\alpha,\beta)$ es $\alpha(\alpha+1)\dots(\alpha+k-1)/\beta^k$.*

*Demostración.* Tenemos que $i^k E[X^k] = \varphi_X^{(k)}(t) = i^k\alpha(\alpha+1)\dots(\alpha+k-1)/\beta^k$. □

**Proposición 14.** *La función generatriz de momentos de la distribución Gamma$(x|\alpha,\beta)$ viene dada por $\varphi_X(t) = \left(\frac{\beta}{\beta-t}\right)^\alpha$.*

*Demostración.* La demostración es análoga a la dada en la Proposición 12. □

**Corolario 15.** *La distribución Gamma$(x|\alpha,\beta)$ tiene media $\alpha/\beta$ y varianza $\alpha/\beta^2$.*

**Proposición 16.** *Sea $n \geq 1$. Consideremos $X_1,\dots,X_n$ variables aleatorias independientes tales que $X_j$ sigue una distribución Gamma$(x|\alpha_i,\beta)$. Entonces, $\sum_{i=1}^n X_j$ sigue una distribución Gamma$(x|\sum_{i=1}^n \alpha_i,\beta)$.*

*Demostración.* En primer lugar, calculamos la función característica de $\sum_{i=1}^n X_j$ como sigue

$$E[e^{i\sum X_j}] = E[\prod e^{iX_j}] = \prod E[e^{iX_j}] = \left(\frac{\beta}{\beta - it}\right)^{\sum \alpha_j},$$

donde se ha utilizado que la esperanza del producto de dos variables aleatorias independientes es el producto de las esperanzas. Por último, nótese que la función característica de la variable $\sum X_j$ es la función característica de $Gamma(x|\sum_{i=1}^n \alpha_i,\beta)$. El hecho de que la función característica de una distribución la determina de forma unívoca finaliza la prueba. □

**Proposición 17.** *Sea $X \sim N(x|0,1)$. La variable aleatoria $Y = X^2$ sigue una distribución Gamma$(y,1/2,1/2)$.*

*Demostración.* Sean $F$ y $G$ las funciones de distribución de las variables $X$ e $Y$ respectivamente. Tenemos que $G(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$. Derivando, obtenemos

$$G'(y) = \frac{F'(\sqrt{y}) + F'(-\sqrt{y})}{2\sqrt{y}} = \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-y/2} = \frac{(1/2)^{1/2}}{\Gamma(1/2)}y^{-1/2}e^{-y/2}.$$

Por último, basta darse cuenta de que $G'(y)$ es la función de densidad de $Gamma(y,1/2,1/2)$. □

### 1.2.4. Distribución beta

La famila de distribuciones beta se encuentra definida sobre el intervalo $(0,1)$. En su definición entra en juego la denominada función beta, de ahí su nombre.

**Definición 4.** Se define la función beta como la aplicación $\beta : (0,\infty)\times(0,\infty) \to (0,\infty)$ dada por

$$\beta(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

**Proposición 18.** *Para cada $x,y>0$ se tiene que $\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \beta(x,y)$. Como consecuencia, la función beta está bien definida.*

*Demostración.* En primer lugar escribimos $\Gamma(x)\Gamma(y)$ como una integral doble

$$\Gamma(x)\Gamma(y) = \int_0^\infty e^{-u}u^{x-1}\,du\int_0^\infty e^{-v}v^{y-1}\,dv = \int_0^\infty\int_0^\infty e^{-u-v}u^{x-1}v^{y-1}\,du\,dv.$$

La expresión anterior nos sugiere utilizar el cambio de variable $(u, v) = J(t, s) = (st, (1 - t)s)$. Nótese que $|J(t, s)| = s$. Aplicamos el cambio a continuación

$$\Gamma(x)\Gamma(y) = \int_0^\infty \left( \int_0^1 e^{-s}(st)^{x-1}(s(1-t))^{y-1}|J(t, s)|\, dt \right) ds$$
$$= \int_0^\infty e^{-s} s^{x+y-2} s \left( \int_0^1 t^{x-1}(1-t)^{y-1}\, dt \right) ds = \Gamma(x + y)\beta(x, y) \qquad \square$$

En la práctica siempre se utiliza la función gamma para evaluar la función beta. Ya podemos definir la distribución beta.

**Definición 5.** Sean $p, q > 0$. Definimos la distribución $beta(x|p, q)$ como la distribución que tiene función de densidad

$$f(x|p, q) = \frac{1}{\beta(p, q)} x^{p-1}(1 - x)^{q-1}, 0 < x < 1.$$

Claramente, la función de densidad integra 1. Esta distribución asigna probabilidad 1 al intervalo $(0, 1)$. Por ello, es útil en modelos de proporciones. Las Figuras 4 y 5 muestran la distribución beta cambiando los valores $p$ y $q$ respectivamente. Podemos observar que las funciones de densidad de $beta(x|p, q)$ y $beta(x|q, p)$ son simétricas respecto del punto $1/2$. Esto se puede demostrar fácilmente a partir de la definición. La Figura 6 muestra la distribución beta con iguales valores de $p$ y $q$. Vemos que las densidades son simétricas en el eje $x = 1/2$, hecho que también puede demostrarse fácilmente a partir de la definición.

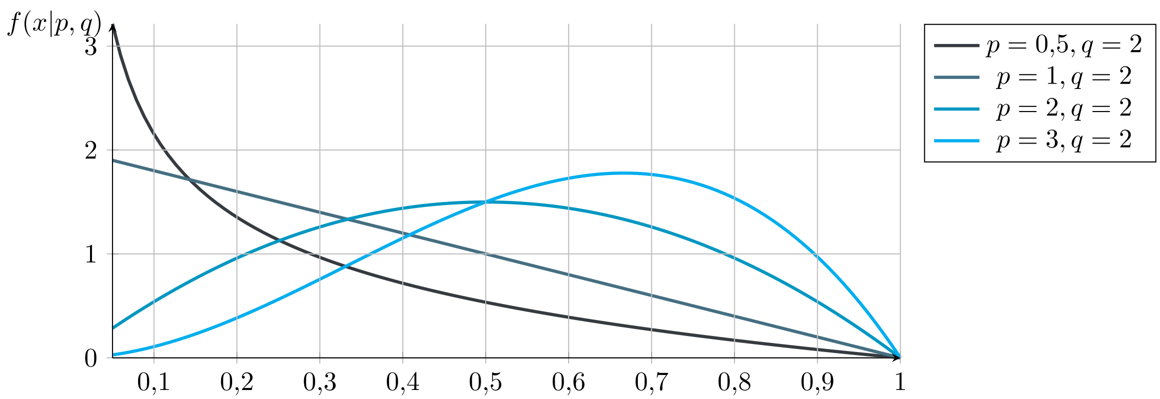


Figura 4: Densidad de la distribución beta con distintos valores de $p$.

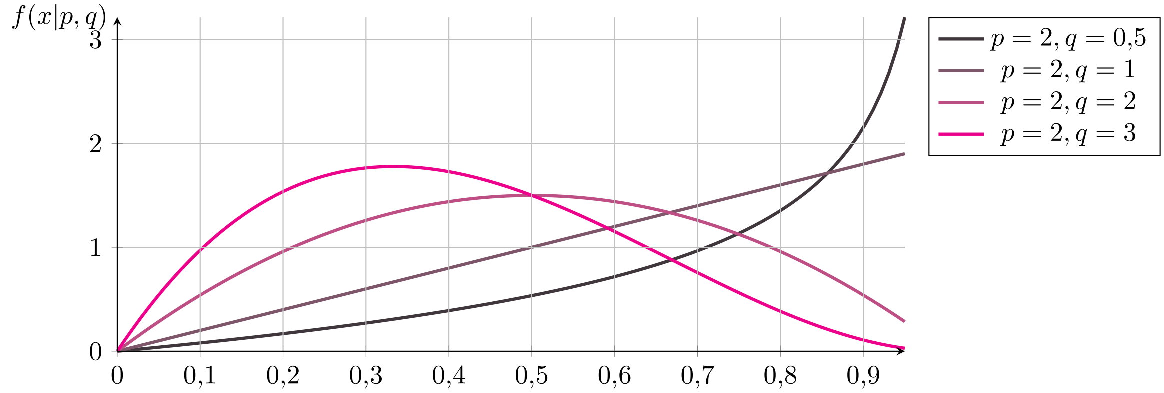


Figura 5: Densidad de la distribución beta con distintos valores de $q$.

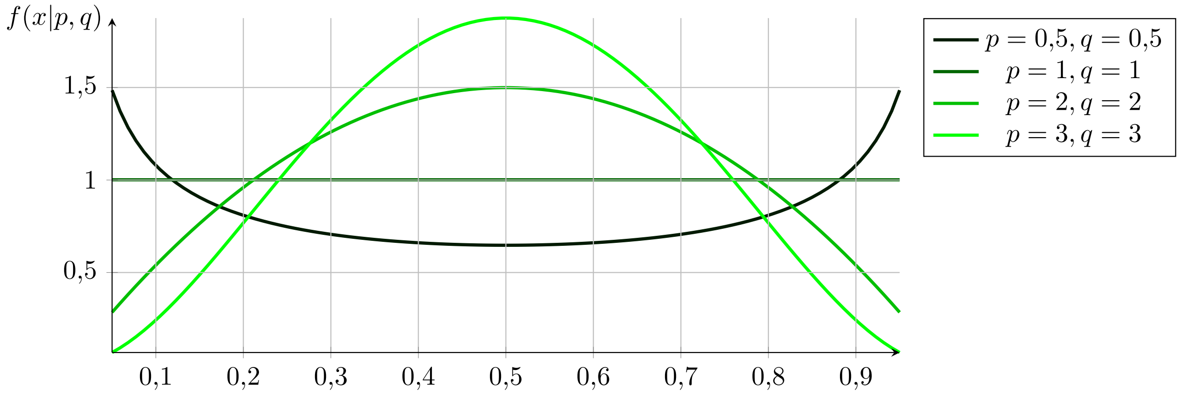


Figura 6: Densidad de la distribución beta con $p = q$.

**Proposición 19.** *La función característica de la distribución* $beta(x|p,q)$ *viene dada por*

$$\varphi_X(t) = 1 + \sum_{k=1}^{\infty} \frac{(it)^k}{k!} \frac{\beta(p+k,q)}{\beta(p,q)}.$$

*Demostración.* Desarrollamos la función característica utilizando el desarrollo de la exponencial

$$E[e^{itX}] = \frac{1}{\beta(p,q)} \int_0^1 e^{itx} x^{p-1}(1-x)^{q-1} dx = \frac{1}{\beta(p,q)} \int_0^1 \sum_{k=0}^{\infty} \frac{(itx)^k}{k!} x^{p-1}(1-x)^{q-1} dx$$

$$= \frac{1}{\beta(p,q)} \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \int_0^1 x^{p+k-1}(1-x)^{q-1} dx = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \frac{\beta(p+k,q)}{\beta(p,q)} = 1 + \sum_{k=1}^{\infty} \frac{(it)^k}{k!} \frac{\beta(p+k,q)}{\beta(p,q)}. \quad \square$$

**Corolario 20.** *Sea* $X \sim beta(x|p,q)$*. Entonces, para cada* $k \geq 1$ *se tiene que*

$$E[X^k] = \frac{\beta(p+k,q)}{\beta(p,q)}.$$

**Corolario 21.** *Sea* $X \sim beta(x|p,q)$*. Entonces,* $E[X] = \frac{p}{p+q}$ *y* $Var(X) = \frac{pq}{(p+q)^2(p+q+1)}$.

*Demostración.* Por el Corolario 20 y la Proposición 18 tenemos que

$$E[X] = \frac{\beta(p+1,q)}{\beta p, q} = \frac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q)} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} = \frac{\Gamma(p+1)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+1)} = \frac{p}{p+q}$$

y

$$E[X^2] = \frac{\beta(p+2,q)}{\beta p, q} = \frac{\Gamma(p+2)\Gamma(q)}{\Gamma(p+q+2)} \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} = \frac{\Gamma(p+2)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+2)} = \frac{p(p+1)}{(p+q)(p+q+1)}.$$

Por último, es directo calcular $Var(X) = E[X^2] - E[X]^2$. □

### 1.2.5. Distribución de Cauchy

**Definición 6.** Sea $\mu \in \mathbb{R}$ y $\sigma > 0$. Definimos la distribución $Cauchy(x|\mu, \sigma)$ como la distribución que tiene función de densidad

$$f(x|\mu,\sigma) = \frac{1}{\sigma\pi} \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^2} = \frac{\sigma}{\pi(\sigma^2 + (x-\mu)^2)}, \ x \in \mathbb{R}.$$

La distribución está bien definida. En efecto, utilizando el cambio de variable $x = \sigma y - \mu$ obtenemos

$$\int_{-\infty}^{\infty} \frac{1}{1 + \left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\infty} \frac{\sigma}{1+y^2} dy = \sigma\pi.$$

**Proposición 22.** *La función característica de la distribución $Cauchy(x|\mu, \sigma)$ viene dada por*

$$\varphi_X(t) = e^{i\mu t - \sigma|t|}.$$

*Demostración.* En primer lugar, demostramos el resultado para $Cauchy(x|0, 1)$. Tenemos que

$$\varphi_X(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1+z^2} dz = e^{-|t|},$$

donde la última igualdad se explica en [2]. Ahora, si $X \sim Cauchy(x|\mu, \sigma)$, entonces $Y = (X - \mu)/\sigma \sim Cauchy(x|0, 1)$ y, por tanto, obtenmos

$$\varphi_X(t) = E[e^{itX}] = E[e^{it(\sigma Y + \mu)}] = e^{it\mu}\varphi_Y(\sigma t) = e^{i\mu t - \sigma|t|}.$$ □

Nótese que la función característica de la distribución de Cauchy no es diferenciable en 0. Consecuentemente, esta distribución no tiene momentos de orden mayor o igual que 1.

**Proposición 23.** *Sean $X$ e $Y$ dos variables aleatorias independientes con distribuciones $Cauchy(x|\mu_1, \sigma_1)$ y $Cauchy(x|\mu_2, \sigma_2)$ respectivamente. Entonces, $X + Y \sim Cauchy(x|\mu_1 + \mu_2, \sigma_1 + \sigma_2)$.*

*Demostración.* Nótese que $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{it(\mu_1+\mu_2) - |t|(\sigma_1+\sigma_2)}$. La prueba finaliza al darse cuenta de que ésta es la función característica de $Cauchy(x|\mu_1 + \mu_2, \sigma_1 + \sigma_2)$. □

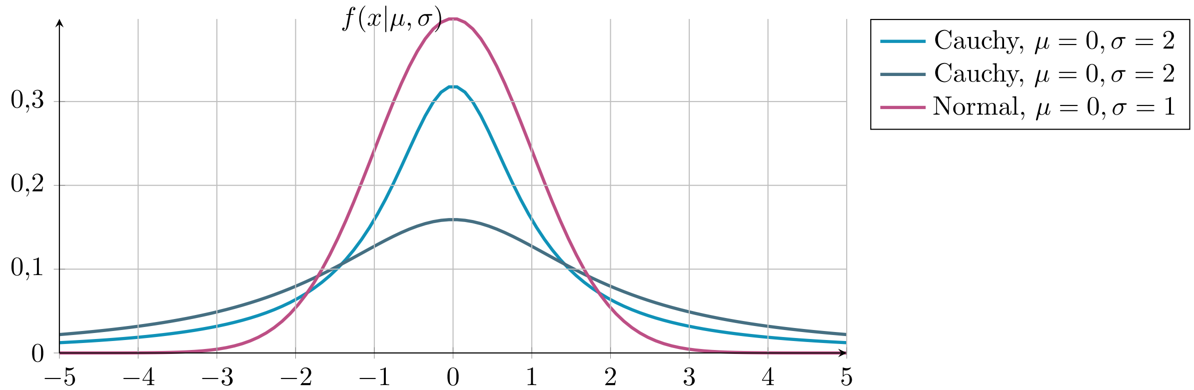


Figura 7: Densidad de la distribución de Cauchy comparada con la distribución normal.

### 1.2.6. Distribución de Laplace

**Definición 7.** Sea $\mu \in \mathbb{R}$ y $\sigma > 0$. Definimos la distribución de Laplace, y la denotamos $Laplace(x|\mu, \sigma)$ como la distribución que tiene función de densidad

$$f(x|\mu, \sigma) = 2\sigma e^{-|x-\mu|/\sigma}, \ x \in \mathbb{R}.$$

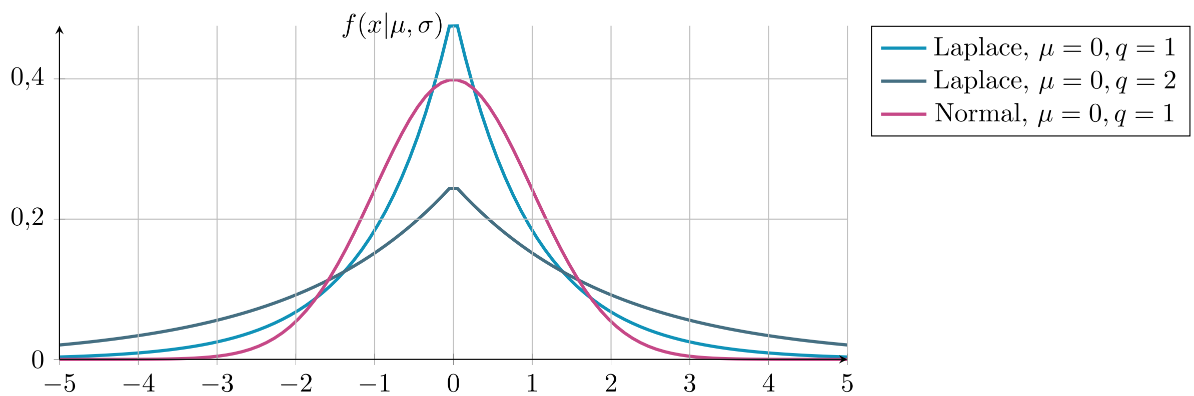


Figura 8: Densidad de la distribución beta con distintos valores de $p$.

### 1.2.7. Distribución T de Student

### 1.2.8. Distribución de Dirichlet

## 2. Estimación de parámetros

Supongamos que estamos estudiando un fenómeno aleatorio que sabemos que sigue una distribución $f(X|\theta_0)$, donde $\theta_0 \in \Omega$ es un parámetro que no es conocido. Nuestro objetivo es estimar el parámetro $\theta_0$ a partir de una muestra $x_1, \ldots, x_n$. Para ello buscamos una función $T_n$ de manera que podamos decir $\theta_0 \approx T_n(x_1, \ldots, x_n)$.

**Definición 8.** Un estimador puntuales una función medible $T_n(X_1, \dots, X_n)$ que toma valores $\Omega$, donde $\Omega$ es el dominio del parámetro a estimar. Una estimación es la evaluación obtenida por un estimador sobre una muestra $x_1, \dots, x_n$, esto es, $T_n(x_1, \dots, x_n)$.

En múltiples situaciones encontramos estimadores de calidad de forma natural. Por ejemplo, imaginemos que el parámetro $\theta_0$ se corresponde con la media de la distribución $f(X|\theta_0)$. En tal caso, parece claro que el mejor estimador para $\theta_0$ será la media muestral $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$. Sin embargo, en general no sabemos qué estimador hay que utilizar. Buscamos técnicas que nos proporcionen estimadores que sean razonables. En ocasiones querremos estimar $g(\theta_0)$, donde $g$ es determinada transformación de $\Omega$ en otro espacio más manejable.

## 2.1. Método de los momentos

El método de los momentos es, probablemente, el método más antiguo para estimar parámetros. Fue propuesto por Pearson al finales del siglo XIX. En muchos casos los resultados de este método son mejorables. Sin embargo, siempre es un último recurso en el caso de que no podamos aplicar otros métodos.

Sea $X_1, \dots, X_n$ una muestra de un fenómeno con función de distribución $f(X|\theta)$ con $\theta \in \Omega$. Suponemos que $\theta = (\theta_1, \dots, \theta_k)$ con $\theta_i \in \mathbb{R}$. Definimos los momentos de la muestra como $m_j = \frac{1}{n}\sum_{i=1}^n X_i^j$. En media se debería cumplir que $m_j = E_\theta X^j$ para todo $j$ tal que $E_\theta X^j$ existe. Nótese que $E_\theta X^j = \mu_j(\theta_1, \dots, \theta_k)$ es una función que depende de $\theta_1, \theta_2, \dots, \theta_k$. El método de los momentos propone como estimador a una solución del sistema de ecuaciones

$$\begin{aligned} m_1 &= \mu_1(\theta_1, \dots, \theta_k), \\ m_2 &= \mu_1(\theta_1, \dots, \theta_k), \\ &\vdots \\ m_k &= \mu_1(\theta_1, \dots, \theta_k). \end{aligned} \tag{1}$$

Ejemplo 24 (Distribución normal): Supongamos que $X_1, \dots, X_n$ son muestras de una distribución normal $N(\theta, \sigma^2)$. En el contexto anterior, los parámetros a estimar son $\theta_1 = \theta, \theta_2 = \sigma^2$. En este caso el sistema (1) viene dado por las ecuaciones $\overline{X} = \theta$ y $m_2 = \theta^2 + \sigma^2$. La solución claramente es $\theta = \overline{X}$ y

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

En este caso, los estimadores obtenidos coinciden con nuestra intuición. Este método es más útil cuando no disponemos de un estimador intuitivo.

Ejemplo 25 (Distribución binomial):

## 2.2. Método de la máxima verosimilitud de Fisher

El método de la máxima verosimilitud es la técnica más utilizada para obtener estimadores.

**Definición 9.** Sea $x_1, \dots, x_n$ una muestra de un fenómeno con función de distribución $f(X|\theta_0)$, donde $\theta_0 \in \Omega$. Se define la función de verosimilitud para cada $\theta \in \Omega$ como $L(\theta|x) = \prod_{i=1}^n f(x_i|\theta)$.

Para cada posible valor $\theta$ del parámetro a estimar, la verosimilitud proporciona la credibilidad que se le da a $\theta$ para los datos $x_1, \dots, x_n$. Buscamos una aproximación $\hat{\theta}$ de $\theta_0$ en base a la muestra obtenida.

Parece lógico que si asumimos que los datos son correctos, entonces una buena aproximación será aquella en la que los datos sean coherentes, esto es, la probabilidad de que se den datos similares a la muestra observada debe ser lo más alta posible.

**Definición 10.** Para cada elemento $x = (x_1, \ldots, x_n)$ del espacio muestral, definimos $\hat{\theta}(x) \in \Omega$ como un máximo global de $L(\theta|x)$. El estimador máximo verosímil (EMV) de una muestra $X$ se define como $\hat{\theta}(X)$.

El estimador máximo verosímil presenta principalmente dos problemas.

- Cálculo del estimador. Para calcular $\hat{\theta}(X)$ es necesario maximizar una función. Muchas veces esto es complejo incluso para funciones de densidad comunes. Es más, puede suceder que la verosimilitud presente múltiples máximos globales y, por tanto, el estimador máximo verosímil no está bien definido. Necesitaremos condiciones sobre la distribución que nos permitan asegurar la buena definición del estimador máximo verosímil.

- Sensibilidad numérica. El valor $\hat{\theta}(x)$ puede cambiar considerablemente para pequeñas variaciones de $x$. Nos preguntamos qué condiciones debe verificar la función de distribución para evitar este comportamiento.

Para adentrarnos en el estudio de estos problemas necesitaremos teoría general de estimadores. Antes de desarrollarla realizaremos varios ejemplos de cálculo de estimadores máximo verosímiles.

**Comentario 1.** Los máximos globales de la función $L(\theta|x)$ se corresponden con los máximos globales de la función $\log L(\theta|x) = \sum_{i=1}^n \log f(x_i|\theta)$. En múltiples ocasiones es más sencillo maximizar esta última expresión.

Ejemplo 26 (Distribución normal): Consideremos una muestra $X_1, \ldots, X_n$ de un fenómeno con distribución $N(\theta, 1)$. En primer lugar, calculamos la función de verosimilitud

$$L(\theta|x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i-\theta)^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^n (x_i-\theta)^2/2}.$$

En virtud del Comentario 1 maximizamos la función $-\sum_{i=1}^n (x_i - \theta)^2/2 - n/2 \log(2\pi)$. Maximizar esta función equivale a minimizar $h(\theta) = \sum_{i=1}^n (x_i - \theta)^2$. Derivando, obtenemos que $h'(\theta) = 0$ si, y solo si, $\theta = \overline{x}$. Además, es rutinario comprobar que $\overline{x}$ es el mínimo absoluto de $h$. Por tanto, $\overline{x}$ es el máximo absoluto de $L(\theta|x)$. Tenemos pues $\hat{\theta}(x) = \overline{x}$.

A continuación pretendemos extender el método de la máxima verosimilitud para estimar $g(\theta)$, donde $g : \Omega \to \Omega'$ sobreyectiva. Si la aplicación $g$ fuese inyectiva, entonces podemos definir de norma natural la verosimilitud de $\eta \in \Omega'$ como $L^*(\eta|x) = L(g^{-1}(\eta)|x)$. Claramente, el valor que maximiza $L^*(\eta|x)$, que denotaremos $\hat{g}(x)$, es $g(\hat{\theta}(x))$. Sin embargo, los casos que presentan relevancia práctica son aquellos en los que $g$ no es inyectiva ya que de esta forma conseguimos reducir la dimensionalidad del espacio de parámetros. Necesitamos extender la definición de verosimilitud para abordar esta problemática.

**Definición 11.** En el contexto anterior, definimos la verosimilitud inducida por $g$ como

$$L^*(\eta|x) = \sup\{L(\theta|x) : \theta \in g^{-1}(\eta)\}.$$

El valor $\hat{g}(x)$ que maximiza $L^*(\eta|x)$ se denomina estimador maximo verosímil de $g(\theta)$.

La definición anterior es artificial en el sentido de que se realiza con el fin de poder mantener la propiedad de invarianza del estimador máximo verosímil, que se recoge en el siguiente teorema.

**Teorema 27** (Invarianza de Zehna)**.** *Para cualquier aplicación sobreyectiva $g : \Omega \to \Omega'$ se tiene que $\hat{g}(X) = g(\hat{\theta}(X))$.*
*Demostración.* En primer lugar, por la definición de la verosimilitud inducida se tiene que

$$\sup_{\eta \in \Omega'} L^*(\eta|x) = \sup_{\eta \in \Omega'} \sup\{L(\theta|x) : \theta \in g^{-1}(\eta)\} = \sup_{\theta \in \Omega} L(\theta|x)$$

Por tanto, la verosimilitud tiene un máximo global si, y solo si, lo tiene la verosimilitud inducida, en cuyo caso la credibilidad de ambos coincide. Si existe un EMV de $\theta$, entonces $L^*(g(\hat{\theta}), x) = L(\hat{\theta}|x)$ por definición. Consecuentemente, $g(\hat{\theta})$ es estimador máximo verosímil de $g(\theta)$. □

## 2.3. Teoría general de estimadores

En esta sección introduciremos conceptos y definiciones relacionados con estimadores arbitrarios. El objetivo de esta teoría es dotarnos de herramientas que nos permitan abordar el estudio práctico de estimadores concretos, como el estimador máximo verosímil.

### 2.3.1. Score. Hipótesis de regularidad. Función de información de Fischer

**Definición 12.** Sea $\Omega \subset \mathbb{R}^m$ un abierto y sea $\{f(X|\theta) : \theta \in \Omega\}$ una familia de funciones de densidad. Sea $X = (X_1, \ldots, X_n)$ una muestra que sigue una distribución con función de densidad $f(X|\theta_0)$. Si la función de verosimilitud para los valores $x = (x_1, \ldots, x_n)$ es diferenciable en $\theta \in \Omega$, entonces definimos el score de $\theta$ como el gradiente de la función verosimilitud en $\theta$ y lo denotamos $S(x; \theta)$.

Intuitivamente el score indica la sensibilidad de la verosimilitud en un punto. Nos centraremos en el estudio del score cuando $\Omega$ es un abierto de $\mathbb{R}$. En tal caso

$$S(x; \theta) = \frac{\partial}{\partial \theta} \log f(x|\theta) = \frac{\frac{\partial}{\partial \theta} f(x|\theta)}{f(x|\theta)}.$$

Supongamos que el score de $\theta$ existe para cualesquiera valores de la muestra $x_1, \ldots, x_n$. En tal caso es natural considerar la función $E_{X|\theta}[S(X; \theta)]$, que depende solamente de $\theta$. Si $\theta$ fuese el parámetro a estimar, entonces $E_{X|\theta}[S(X; \theta)]$ mide la sensibilidad media de la verosimilitud en $\theta$.

**Lema 28.** *Sea $\Omega \subset \mathbb{R}$ un abierto y sea $\{f(X|\theta) : \theta \in \Omega\}$ una familia de funciones de densidad que verifican las siguientes condiciones de regularidad:*

*a) Para cualesquiera valores $x = (x_1, \ldots, x_n)$ la función $L(x; \theta)$ es diferenciable para todo $\theta \in \Omega$.*

*b) Se verifica*
$$\frac{\partial}{\partial \theta} \int_X f(x|\theta) dx = \int_X \frac{\partial}{\partial \theta} f(x|\theta) dx.$$

*Entonces, $E_{X|\theta}[S(X; \theta)] = 0$.*
*Demostración.* Tenemos que

$$E_{X|\theta}[S(X; \theta)] = \int_X \frac{\frac{\partial}{\partial \theta} f(x|\theta)}{f(x|\theta)} f(x|\theta) dx = \int_X \frac{\partial}{\partial \theta} f(x|\theta) dx = \frac{\partial}{\partial \theta} \int_X f(x; \theta) dx = \frac{\partial}{\partial \theta} 1 = 0. \qquad \square$$

En el resultado anterior aparecen por primera vez hipótesis de regularidad sobre las distribuciones a estudiar. Nótese que en la práctica normalmente vamos a trabajar con distribuciones que verifiquen estas hipótesi. La hipótesis b) se verifica si la función la derivada de la función de densidad con respecto a $\theta$ es continua e integramos en un intervalo [3]. Consecuentemente, todas las distribuciones continuas estudiadas, exceptuando la distribución de Laplace, cumplen estas hipótesi de regularidad (su función de densidad es de clase infinito con respecto a $\theta$).

En lo que sigue habitualmente exigiremos unas hipótesis de regularidad más fuertes, asumiendo que $\Omega \subset \mathbb{R}^m$. En tal caso, se indicará que el resultado es verdad bajo hipótesis de regularidad. Las hipótesis que consideraremos son las siguientes:

i) Para cualquier muestra $x = (x_1, \ldots, x_n)$, la verosimilitud $f(x|\theta)$ es dos veces derivable en $\Omega$.

ii) $\frac{\partial^i}{\partial \theta^i} \int_X f(x|\theta)dx = \int_X \frac{\partial^i}{\partial \theta^i} f(x|\theta)dx$ para $i = 1, 2$.

Se deja como ejercicio razonar que todas las distribuciones continuas estudiadas, exceptuando la distribuciónd de Laplace, verifican estas hipótesis de regularidad.

**Definición 13.** Sea $\Omega \subset \mathbb{R}$ un abierto y sea $\{f(X|\theta) : \theta \in \Omega\}$ una familia de funciones de densidad para la cual siempre existe el score. Dado $\theta \in \Omega$, definimos la función de información de Fischer en $\theta$ como el segundo momento de la variable aleatoria $S(X;\theta)$, donde $X$ sigue una distribución con función de densidad $f(X|\theta)$. Se denota $\mathcal{I}(\theta) := E_{X|\theta}[S(X;\theta)^2] \geq 0$.

**Lema 29.** *Bajo hipótesis de regularidad tenemos que* $\mathcal{I}(\theta) = Var_{X|\theta}(S(X;\theta))$

$$\mathcal{I}(\theta) = E_{X|\theta}\left[-\frac{\partial^2}{\partial \theta^2} \log f(X;\theta)\right].$$

*Demostración.* En primer lugar, podemos escribir

$$\frac{\partial^2}{\partial \theta^2} \log f(X|\theta) = \frac{\frac{\partial^2}{\partial \theta^2} f(X|\theta)}{f(X|\theta)} - \left(\frac{\frac{\partial}{\partial \theta} f(X|\theta)}{f(X|\theta)}\right)^2 = \frac{\frac{\partial^2}{\partial \theta^2} f(X|\theta)}{f(X|\theta)} - \left(\frac{\partial}{\partial \theta} \log f(X|\theta)\right)^2.$$

La demostración finaliza al tomar esperanzas en la expresión anterior y darse cuenta de que

$$E_{X|\theta}\left[\frac{\frac{\partial^2}{\partial \theta^2} f(X|\theta)}{f(X|\theta)}\right] = \int_X \frac{\partial^2}{\partial \theta^2} f(x|\theta)\, dx = \frac{\partial^2}{\partial \theta^2} \int_X f(x|\theta)\, dx = \frac{\partial^2}{\partial \theta^2}\, 1 = 0. \qquad \square$$

Ejemplo 30: Calculamos la función de información de Fisher para la distribución normal.

### 2.3.2. Estimadores insesgados

Para comprobar cómo de bueno es un estimador $T$ podemos definir una función de pérdida $L(\theta, T(X))$ que indique la pérdida asociada a estimar un parámetro mediante un estimador $T$ si su verdadero valor es $\theta$. A partir de la función de pérdida deinfimos la función de riesgo, que asocia a cada posible valor del parámetro la pérdida media asociada al estimador. La función de riesgo viene dada por

$$R_T^L(\theta) = E_{X|\theta}[L(\theta, T(X))].$$

Un estimador $T$ que "minimice uniformemente" la función de riesgo hará mejores estimaciones en media. Con minimizar uniformemente queremos decir que para cada estimador $T'$ se tiene que

$$R_T^L(\theta) \leq R_{T'}^L(\theta) \ \forall \ \theta \in \Omega.$$

En la siguiente definición introducimos un tipo particular de estimadores que minimizan determinada función de riesgo.

**Definición 14.** Se denomina sesgo de un estimador $T$ de $g(\theta)$ a la diferencia entre la esperanza del estimador y el verdadero valor del parámetro a estimar. Diremos que un estimador es insesgado si para cualquier posible valor del parámetro a estimar su sesgo es nulo.

Nótese que el sesgo de un estimador es la función de riesgo asociada a la pérdida $L(\theta, T(X)) = \theta - T(X)$. Un estimador insesgado verifica $0 = \theta - E_{X|\theta}T(X)$ y, por tanto, minimiza uniformemente la función de riesgo. Aunque esta propiedad puede parecer a priori interesante, puede suceder que en la práctica el estimador insesgado no se comporte como es deseado debido a que la varianza $Var_\theta(T(X))$ sea muy alta.

El siguiente resultado nos muestra un ejemplo de estimador insesgado.

**Proposición 31.** *Sea $X_1, \ldots, X_n$ una muestra de alguna población con función de densidad $f(X|\theta_0)$. Definimos la varianza muestral como*

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

*Entonces, $S^2$ es un estimador insesgado de la varianza de la distribución.*

*Demostración.* Nótese que $\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$. Consecuentemente tenemos

$$E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \sum_{i=1}^{n} E\left[X_i^2\right] - nE[\overline{X}^2] = n(E\left[X_i^2\right] - E[\overline{X}^2]).$$

Utilizando que $Var(\overline{X}) = Var(X_i)/n$ y $E[\overline{X}] = E[X_i]$ obtenemos

$$E[X_i^2] - E[\overline{X}^2] = Var(X_i) + E[X_i]^2 - Var(\overline{X}) - E[\overline{X}]^2 = \frac{n-1}{n}Var(X_i).$$

Por tanto, $E[S^2] = Var(X_i)$ como se quería. □

### 2.3.3. Estadísticos suficientes

Sea $\{f(x|\theta) : \theta \in \Omega\}$ una familia de distribuciones. Sea $X = (X_1, \ldots, X_n)$ una muestra de la distribución $f(x|\theta_0)$. Nuestro objetivo es deducir el parámetro $\theta_0$ a partir de la muestra. El concepto de estadístico suficiente nos permitirá separar la información contenida en $X$ en dos partes. Una parte contiene toda la información útil sobre $\theta_0$ mientras que la otra parte no dependerá del parámetro $\theta$ que se escoja. Consecuentemente, podemos ignorar esta última parte.

Intuitivamente, un estadístico $T$ es suficiente para la familia de distribuciones considerada si $T(X)$ nos permite estimar $\theta_0$ tan bien como lo permite toda la muestra $X$. Procedemos a dar la definición matemática.

**Definición 15.** Un estadístico $T(X_1, X_2, \ldots, X_n)$ es suficiente para cada $t$ la distribución condicional de $X_1, X_2, \ldots, X_n$ respecto de $T(X) = t$ y $\theta$ no depende de $\theta$.

El teorema de factorización de Neyman nos proporciona un criterio práctio para ver si un estadístico es suficiente.

**Teorema 32.** *Sea $\{f(x|\theta) : \theta \in \Omega\}$ una familia de distribuciones. Sea $X = (X_1, \ldots, X_n)$ una muestra de la distribución $f(x|\theta)$. Sea $T(X_1, X_2, \ldots, X_n)$ un estadístico. Entonces, $T$ es suficiente si y solo si la función de verosimilitud puede factorizarse de la siguiente forma*

$$L(x_1, x_2, \ldots, x_n) = h(t; \theta)g(x_1, x_2, \ldots, x_n),$$

*donde* $t = T(x_1, \ldots, x_n)$ *y* $g(x_1, x_2, \ldots, x_n)$ *no depende de* $\theta$.

Si encontramos un estadístico suficiente, entonces podemos inferir el parámetro $\theta_0$ utilizando solamente la función $h(t; \theta)$. Interesa pues que el codominio del estadístico suficiente sea lo más simple posible. Los estadísticos suficientes son especialmente interesantes al aplicar el método de la máxima verosimilitud.

Ejemplo 33 (Distribución normal, media desconocida): Sea $X = (X_1, \ldots, X_n)$ una muestra de $N(x|\mu, \sigma^2)$ donde solamente $\sigma^2$ es conocido ($\theta = \mu$). Es fácil verificar que

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2 = (n-1)S^2 + n(\overline{x} - \mu)^2.$$

A partir de la igualdad anterior obtenemos

$$f(x|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\sum_{i=1}^{n}(x_i - \mu)^2/(2\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-((n-1)S^2 + n(\overline{x} - \mu)^2)/(2\sigma^2)).$$

Definimos

$$g(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-(n-1)S^2/(2\sigma^2)) \text{ y } h(t|\mu) = \exp(n(t - \mu)^2)/(2\sigma^2)).$$

Tenemos que $f(x|\theta) = h(\overline{x}|\theta)g(x_1, x_2, \ldots, x_n)$ y, por tanto, $T(x) = \overline{x}$ es suficiente.

Ejemplo 34 (Distribución normal, ambos parámetros son desconocidos): Sea $X = (X_1, \ldots, X_n)$ una muestra de $N(x|\mu, \sigma^2)$ donde $\mu$ y $\sigma^2$ son desconocidos ($\theta = (\mu, \sigma^2)$). Tenemos que

$$f(x|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\sum_{i=1}^{n}(x_i - \mu)^2/(2\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2)/(2\sigma^2)).$$

Por tanto, el estadístico $T(X_1, \ldots, X_n) = (\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i)$ es suficiente (tomamos $g(x_1, x_2, \ldots, x_n) = 1$). También podemos desarrollar $f(x|\theta)$ como sigue

$$f(x|\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-(\sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2)/(2\sigma^2)).$$

Consecuentemente, el estadístico $T(X_1, \ldots, X_n) = (\overline{x}, S^2)$ también es suficiente. Nótese que el estadístico $T(X_1, \ldots, X_n) = (X_1, \ldots, X_n)$ es trivialmente suficiente, pero no aporta ninguna información.

## 2.4. Estudio teórico del estimador máximo verosímil

En este punto nos preguntamos cuándo está bien definido el estimador máximo verosímil. En tal caso nos interesa saber si el método de la máxima verosimilitud obtendrá una mejor aproximación de $\theta_0$ cuanto mayor sea la muestra. Para desarrollar esta cuestión de forma teórica introducimos la siguiente definición.

**Definición 16.** Consideremos una familia de densidades $\{f(x|\theta) : \theta \in \Omega\}$. Un estimador $T_n$ es consistente si para cualquier $\theta_0 \in \Omega$ se verifica que para toda sucesión $X_n$ de variables aleatorias independientes e idénticamente distribuidas con función de distribución $f(x|\theta_0)$ la sucesión $T_n(X_1, \ldots, X_n)$ converge en probabilidad a $g(\theta_0)$, donde $g$ es una función fijada de antemano.

**Teorema 35.** *Bajo las condiciones anteriores se verifican las siguientes afirmaciones:*

a) *Existe* $n_0$ *tal que para cada* $n \geq n_0$ *la ecuación en probabilidad* $0 = \sum_{i=0}^{n} \frac{\partial}{\partial \theta} \log f(X_i|\theta)$ *tiene solución única. A esta solución se le llama* $\hat{\theta}_n(X_1, \ldots, X_n)$*. En dicho punto se maximiza la verosimilitud.*

b) $\hat{\theta}(X_1, \ldots, X_n)$ *converge en* $P_{\theta_0}$ *a* $\theta_0$*. De hecho, se puede probar que la convergencia es casi segura.*

*Demostración.* $0 > -\mathcal{I}(\theta) = \mathrm{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(X;\theta) \middle| \theta_0\right] = \frac{\partial}{\partial \theta} \mathrm{E}\left[\frac{\partial}{\partial \theta} \log f(X;\theta) \middle| \theta_0\right]$ es decreciente en un entorno de $\theta_0$. □

# Referencias


[1] Proof Wiki, Euler's Reflection Formula, https://proofwiki.org/wiki/Euler%27s_Reflection_Formula.

[2] Wikipedia, Residue theorem, https://en.wikipedia.org/wiki/Residue_theorem#Example.

[3] Wikipedia, Leibniz integral rule, https://en.wikipedia.org/wiki/Leibniz_integral_rule.