



Inferencia Estadística

Apuntes

Andrés Herrera Poyatos
Universidad de Granada
andreshp9@gmail.com

Índice

1. Familias de distribuciones	2
1.1. Distribuciones discretas	2
1.2. Distribuciones continuas	2
1.2.1. Distribución uniforme	2
1.2.2. Distribución normal	2
1.2.3. Distribución gamma	2
1.2.4. Distribución beta	5
1.2.5. Distribución de Cauchy	5
1.2.6. Distribución de Laplace	5
1.2.7. Distribución T de Student	5
1.2.8. Distribución de Dirichlet	5
2. Estimación de parámetros	5
2.1. Método de los momentos	5
2.2. Estimadores insesgados	6
2.3. Estadísticos suficientes	6
2.4. Método de la máxima verosimilitud de Fisher	7

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 y, por tanto, también su varianza, como sigue

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

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

1.2.2. Distribución normal

1.2.3. Distribución gamma

La familia 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 1. Se define la función gamma como la aplicación $\Gamma : (0, \infty) \rightarrow (0, \infty)$ dada por

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

Proposición 1. 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^{α}/α 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

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

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. \square

Proposición 2 (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 obtenemos

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

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 2. Sean $\alpha, \beta > 0$. Definimos la distribución $\text{Gamma}(x|\alpha, \beta)$ como la distribución que tiene función de densidad

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

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

Proposición 3. La función de densidad de la distribución $\text{Gamma}(\alpha, \beta)$ verifica las siguientes propiedades:

- Si $0 < \alpha < 1$, entonces $f(x|\alpha, \beta)$ es decreciente y $f(x) \rightarrow \infty$ para $x \rightarrow 0$.
- 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

$$f'(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-2} e^{-\beta x} [(\alpha - 1) - \beta x];$$

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

La Figura 1 muestra la función de densidad de la distribución gamma para distintos valores de α . El parámetro β se denomina parámetro de escala debido a su influencia en la escala de la función de densidad. La Figura 2 muestra la función de densidad de la distribución gamma para distintos valores de β .

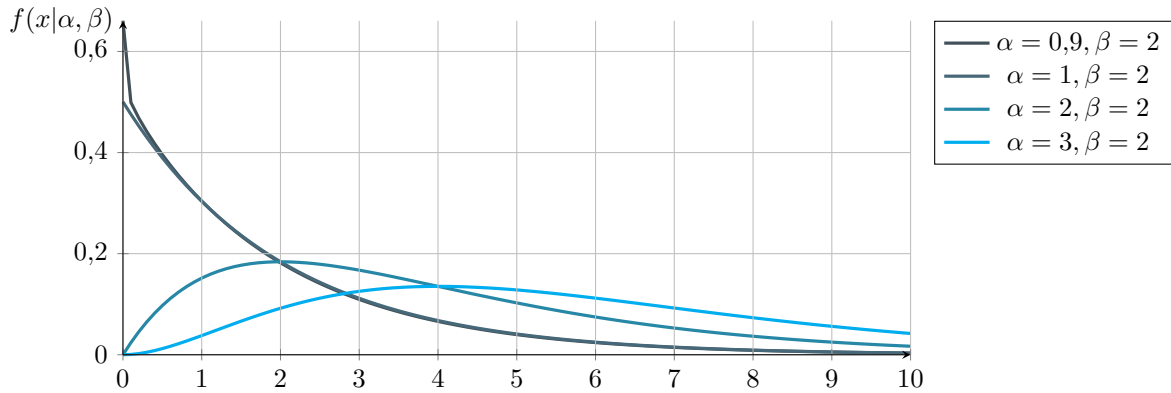


Figura 1: Densidad de la distribución gamma con distintos valores de α .

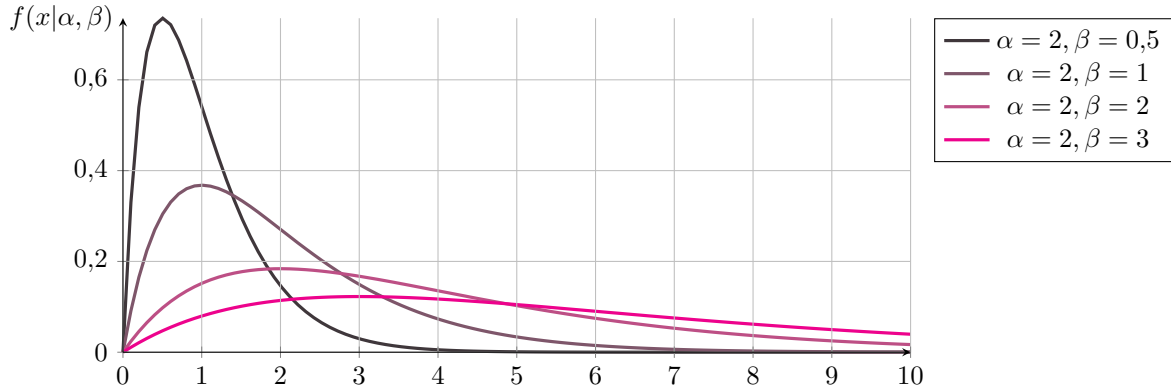


Figura 2: Densidad de la distribución gamma con distintos valores de β .

Proposición 4. La función característica de la distribución $\text{Gamma}(\alpha, \beta)$ viene dada por $\varphi_X(t) = (\beta - it)^{-\alpha}$.

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

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

1.2.4. Distribución beta**1.2.5. Distribución de Cauchy****1.2.6. Distribución de Laplace****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 θ_0 a partir de una muestra x_1, \dots, x_n . Para ello buscamos una función T_n de manera que podamos decir $\theta_0 \approx T_n(x_1, \dots, x_n)$.

Definición 3. Un estimador puntual es una función medible $T_n(X_1, \dots, X_n)$ que toma valores en Ω , donde Ω 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 θ_0 se corresponde con la media de la distribución $f(X|\theta_0)$. En tal caso, parece claro que el mejor estimador para θ_0 será la media muestral $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Sin embargo, en general no sabemos qué estimador hay que utilizar. Buscamos técnicas que nos proporcionen estimadores que sean razonables. En ocasiones queremos estimar $g(\theta_0)$, donde g es determinada transformación de Ω en otro espacio más manejable.

Para comprobar cómo de bueno es un estimador podemos definir una función de pérdida $L(\theta, T)$ que indique la pérdida asociada a estimar un parámetro mediante un estimador T si su verdadero valor es θ . A partir de la función de pérdida definimos 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

$$R_T^L(\theta) = E_\theta[L(\theta, T)].$$

El estimador óptimo es el que minimiza uniformemente la función de riesgo, esto es, T es estimador óptimo si para cada estimador T' se tiene que

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

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

$$\begin{aligned} m_1 &= \mu_1(\theta_1, \dots, \theta_k), \\ m_2 &= \mu_2(\theta_1, \dots, \theta_k), \\ &\vdots \\ m_k &= \mu_k(\theta_1, \dots, \theta_k). \end{aligned} \tag{1}$$

EJEMPLO 5 (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 $\bar{X} = \theta$ y $m_2 = \theta^2 + \sigma^2$. La solución claramente es $\theta = \bar{X}$ y

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

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

EJEMPLO 6 (Distribución binomial):

2.2. Estimadores insesgados

Definición 4. 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 su sesgo es nulo.

Proposición 7. Sea X_1, \dots, X_n una muestra de alguna población con función de densidad $f(X|\theta_0)$. Definimos la varianza muestral como

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

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

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

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

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

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

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

2.3. Estadísticos suficientes

Intuitivamente, un estadístico es suficiente para el parámetro θ si utiliza toda la información contenida en la muestra aleatoria con respecto a θ .

Un criterio para ver si un estadístico es suficiente viene dado por el teorema de factorización de Neyman.

Teorema 8. Sea X_1, X_2, \dots, X_n una muestra de una distribución con una función de densidad de probabilidad $f(x; \theta)$. Se dice que $T(X_1, X_2, \dots, X_n)$ es un estadístico suficiente de θ si y solo si la función de verosimilitud puede factorizarse de la siguiente forma

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

2.4. Método de la máxima verosimilitud de Fisher

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

Definición 5. 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 θ del parámetro a estimar, la verosimilitud proporciona la credibilidad que se le da a θ para los datos x_1, \dots, x_n . Buscamos una aproximación $\hat{\theta}$ de θ_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 6. Para cada elemento $x = (x_1, \dots, x_n)$ del espacio muestral, definimos $\hat{\theta}(x) \in \Omega$ como un máximo 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.
- 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.

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 9 (Distribución normal): Consideremos una muestra X_1, \dots, 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, $\theta = \bar{x}$. Además, es rutinario comprobar que \bar{x} es el mínimo absoluto de h . Por tanto, \bar{x} es el máximo absoluto de $L(\theta|x)$. Tenemos pues $\hat{\theta}(x) = \bar{x}$.

A continuación pretendemos extender el método de la máxima verosimilitud para estimar $g(\theta)$, donde $g: \Omega \rightarrow \Omega'$ 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 7. 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 máximo 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 10 (Invarianza de Zehna). *Para cualquier aplicación sobreyectiva $g : \Omega \rightarrow \Omega'$ 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 θ , entonces $L^*(g(\hat{\theta}), x) = L(\hat{\theta}|x)$ por definición. Consecuentemente, $g(\hat{\theta})$ es estimador máximo verosímil de $g(\theta)$. \square

Referencias

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