Inferencia Estadística

Universidad de Málaga

Grado en Matemáticas

Junio de 2024

Contenidos

Muestra aleatoria simple

En otras asignaturas relacionadas con la Probabilidad y la Estadística, al estudiar las propiedades de un experimento aleatorio se suponían conocidos tanto el espacio de probabilidad $(\Omega, \mathcal{A}, P)$ como la variable aleatoria X que describe el experimento. En la práctica, se debería tener en cuenta que los datos conocidos suelen proceder de un conjunto de datos más grande; por ejemplo, si se quiere estudiar la estatura de la población española, entonces Ω no va a ser un conjunto de 47 millones de personas, sino que habrá que considerar solo una parte de la población. La principal preocupación de la Inferencia Estadística es inducir las propiedades una población estadística a partir de una porción de la misma.

1.1. Nociones básicas

Normalmente, el procedimiento para recoger datos de un experimento aleatorio consiste en realizar múltiples observaciones de una cierta variable de interés, lo que motiva la definición siguiente:

Definición 1. *Dada una variable aleatoria X, una* muestra aleatoria simple de X *es un vector aleatorio $(X_1, \ldots, X_n)$ constituido por variables aleatorias independientes y con la misma función de distribución que X.*

Las distribuciones de probabilidad con las que se trabajará habitualmente son aquellas que figuran en la Sección A.8 y la Sección A.9. En todos esos casos, la función de masa o densidad de X depende de uno o varios parámetros.

Definición 2. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X y supóngase que f_θ, la función de masa o densidad de X, depende de cierto parámetro θ perteneciente a una familia de parámetros $\Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$.*

(i) A Θ se le denomina espacio paramétrico.

(ii) El conjunto

$$\chi^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1, \ldots, x_n \in \chi\}$$

se denomina espacio muestral, *donde*

$$\chi_\theta = \{x \in \mathbb{R} : f_\theta(x) > 0\} \qquad \text{y} \qquad \chi = \bigcup_{\theta \in \Theta} \chi_\theta$$

Ejemplo 1. Se va a tratar de hallar la distribución de una muestra aleatoria simple $(X_1, \ldots, X_n)$ de $X \sim Bin(1, p)$ para cualquier $p \in (0,1)$. El espacio paramétrico sería $\Theta = (0,1) \subset \mathbb{R}$. Además, si $x \in \{0,1\}$,

$$P_p(X = x) = p^x (1-p)^{1-x},$$

luego

$$\chi_p = \{0,1\} = \chi,$$

mientras que el espacio muestral sería

$$\chi^n = \{(x_1, \ldots, x_n) : x_i \in \{0,1\} \text{ para todo } i = 1, \ldots, n\},$$

y la función de masa de la muestra vendría dada por

$$P_p(X_1 = x_1, \dots, X_n = x_n) = P_p(X_1 = x_1)\dots P_p(X_n = x_n) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

para todo $(x_1, \dots, x_n) \in \chi^n$. Cabe remarcar que en la primera igualdad se ha usado que las variables aleatorias son independientes.

Ejemplo 2. Se va a tratar de hallar la distribución de una muestra aleatoria simple $(X_1, \dots, X_n)$ de $X \sim U([a,b])$ para $a, b \in \mathbb{R}$, $a < b$. El espacio paramétrico ahora sería $\Theta = \{(a,b) \in \mathbb{R}^2 : a < b\}$. Además, si $x \in (a,b)$,

$$f_{a,b}(x) = \frac{1}{b-a},$$

luego

$$\chi_{a,b} = (a,b) \qquad \text{y} \qquad \chi = \bigcup_{(a,b)\in\Theta} (a,b) = \mathbb{R},$$

mientras que el espacio muestral sería

$$\chi^n = \mathbb{R}^n,$$

y la función de densidad de la muestra vendría dada por

$$f^n_{a,b}(x_1, \dots, x_n) = f_{a,b}(x_1)\dots f_{a,b}(x_n) = \begin{cases} \dfrac{1}{(b-a)^n} & \text{si } \min\limits_{i=1,\dots,n} x_i > a,\ \max\limits_{i=1,\dots,n} x_i < b \\ 0 & \text{en otro caso} \end{cases}$$

para todo $(x_1, \dots, x_n) \in \mathbb{R}^n$. Cabe remarcar que en la primera igualdad se ha usado que las variables aleatorias son independientes.

1.2. Estadísticos muestrales

Definición 3. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Un* estadístico muestral *o simplemente* estadístico *es una función $T : \mathbb{R}^n \to \mathbb{R}^m$ medible.*

Si T es un estadístico muestral y $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria, entonces $T(X_1, \dots, X_n)$ es otra variable aleatoria por ser T una función medible. Usualmente, en lugar de escribir los valores que toma T como una función de $(x_1, \dots, x_n) \in \mathbb{R}^n$, se escribe directamente como variable aleatoria. Esto es lo que se hará en las subsecciones siguientes, que reúnen los estadísticos muestrales más célebres.

Cabe advertir que la comprobación de que una función dada $T : \mathbb{R}^n \to \mathbb{R}^m$ es medible (y, por tanto, un estadístico muestral) escapa a los propósitos de esta asignatura. Lo habitual será trabajar con funciones inocentes e inofensivas, así que no existirán problemas de medibilidad.

1.2.1. Total muestral

Definición 4. *Dada una muestra aleatoria simple $(X_1, \dots, X_n)$ de una variable aleatoria X, se define el* total muestral *como*

$$T := \sum_{i=1}^n X_i$$

1.2.2. Media muestral

Definición 5. *Dada una muestra aleatoria simple $(X_1, \dots, X_n)$ de una variable aleatoria X, se define la* media muestral *como*

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

Proposición 1. *Sea X una variable aleatoria y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. Se verifican las siguientes propiedades:*

(i) *Si $E(X)=\mu<\infty$, entonces*
$$E(\overline{X})=\mu$$

(ii) *Si $E(X^2)<\infty$ y $\mathrm{Var}(X)=\sigma^2$, entonces*
$$\mathrm{Var}(\overline{X})=\frac{\sigma^2}{n}$$

(iii) *Si $E(\overline{X}^2)<\infty$, entonces*
$$E(\overline{X}^2)=\mathrm{Var}(\overline{X})+E(\overline{X})^2=\frac{\sigma^2}{n}+\mu$$

(iv) *La función generatriz de momentos de $\overline{X}$ viene dada por*
$$M_{\overline{X}}(t)=\Big(M_X\Big(\frac{t}{n}\Big)\Big)^n$$

(v) *Si $E(X)=\mu<\infty$, entonces $\overline{X}\xrightarrow{P}\mu$ y $\overline{X}\xrightarrow{\text{c.s.}}\mu$.*

(vi) *Si $E(X^2)<\infty$, entonces*
$$\frac{\overline{X}-\mu}{\sigma}\sqrt{n}\xrightarrow{\mathcal{L}}Z\sim N(0,1)$$

Demostración. Se va a utilizar constantemente que $E(X)=E(X_i)$ y $\mathrm{Var}(X)=\mathrm{Var}(X_i)$ para cualquier $i=1,\dots,n$ por poseer todas las variables aleatorias la misma distribución de probabilidad.

(i) En efecto,
$$E(\overline{X})=E\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)=\frac{1}{n}\sum_{i=1}^n E(X_i)=\frac{1}{n}\sum_{i=1}^n\mu=\frac{1}{n}n\mu=\mu$$

(ii) En efecto,
$$\mathrm{Var}(\overline{X})=\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)=\frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big)=\frac{1}{n^2}\sum_{i=1}^n\mathrm{Var}(X_i)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

En la tercera igualdad se ha usado que las variables aleatorias de la muestra aleatoria simple son independientes.

(iii) Esta se queda como ejercicio.

(iv) Nótese que todas las variables aleatorias de la muestra tienen la misma función generatriz de momentos por ser idénticamente distribuidas. Además,
$$M_{\overline{X}}(t)=E(e^{t\overline{X}})=E(e^{\frac{t}{n}\sum_{j=1}^n X_j})=E\Big(\prod_{j=1}^n e^{\frac{t}{n}X_j}\Big)=\prod_{j=1}^n E(e^{\frac{t}{n}X_j})=\prod_{j=1}^n M_{X_j}\Big(\frac{t}{n}\Big)=\Big(M_{X_i}\Big(\frac{t}{n}\Big)\Big)^n$$

para cualquier $i=1,\dots,n$. En la cuarta igualdad se ha usado que las variables aleatorias de la muestra aleatoria simple son independientes.

(v) Sea $\varepsilon>0$. Por la desigualdad de Chebyshev,
$$P(|\overline{X}-E(\overline{X})|>\varepsilon)\le\frac{\mathrm{Var}(\overline{X})}{\varepsilon^2},$$

o lo que es lo mismo, usando (i) y (ii),
$$P(|\overline{X}-\mu|>\varepsilon)\le\frac{\sigma^2}{n\varepsilon^2}$$

Por tanto, tomando límites cuando $n \to \infty$,

$$\lim_{n\to\infty} P(|\overline{X} - \mu| > \varepsilon) = 0,$$

luego $\overline{X} \xrightarrow{P} \mu$. La convergencia casi segura es trivial a partir del teorema de Kolmogorov.

(vi) Nótese que, si $S_n = X_1 + \ldots + X_n$, se tiene que

$$\overline{X} = \frac{S_n}{n}$$

Por tanto,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{\frac{S_n}{n} - \mu}{\frac{\sqrt{n}\sigma}{n}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$$

El teorema de Lindeberg-Lévy termina la demostración. □

1.2.3. Varianza muestral

Definición 6. *Dada una muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de una variable aleatoria* X*, se define la* varianza muestral *como*

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

Proposición 2. *Sea* X *una variable aleatoria y sea* $(X_1, \ldots, X_n)$ *una muestra aleatoria simple de* X*. Si* $E(X^2) < \infty$ *y* $\mathrm{Var}(X) = \sigma^2$*, entonces*

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

Demostración. Se tiene que

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

Por tanto,

$$\begin{aligned} E(S^2) &= E\Big(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2\Big) = \frac{1}{n}\sum_{i=1}^{n}E(X_i^2) - E(\overline{X}^2) = E(X^2) - E(\overline{X}^2) \\ &= \mathrm{Var}(X) + E(X)^2 - (\mathrm{Var}(\overline{X}) + E(\overline{X})^2) = \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2} = \Big(1 - \frac{1}{n}\Big)\sigma^2 = \frac{n-1}{n}\sigma^2 \end{aligned}$$

En la cuarta igualdad se ha despejado $E(X^2)$ de la igualdad $\mathrm{Var}(X) = E(X^2) - E(X)^2$, y lo mismo con $E(\overline{X}^2)$. □

1.2.4. Cuasivarianza muestral

Definición 7. *Dada una muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de una variable aleatoria* X*, se define la* cuasivarianza muestral *como*

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

Proposición 3. *Sea* X *una variable aleatoria y sea* $(X_1, \ldots, X_n)$ *una muestra aleatoria simple de* X*. Si* $E(X^2) < \infty$ *y* $\mathrm{Var}(X) = \sigma^2$*, entonces*

$$E(S_c^2) = \sigma^2$$

Demostración. Se tiene que

$$E(S_c^2) = E\Big(\frac{n}{n-1}S^2\Big) = \frac{n}{n-1}E(S^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2,$$

donde en la tercera igualdad se ha usado la proposición anterior. □

1.2.5. Momento muestral respecto al origen

Definición 8. *Dada una muestra aleatoria simple $(X_1,\dots,X_n)$ de una variable aleatoria X, se define el* momento muestral de orden k respecto al origen *como*

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

Proposición 4. *Sea X una variable aleatoria y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. Se verifican las siguientes propiedades:*

(i) Si $E(X^k)=\alpha_k<\infty$, entonces

$$E(m_k)=\alpha_k$$

(ii) Si $E(X^{2k})=\alpha_{2k}<\infty$, entonces

$$\mathrm{Var}(m_k)=\frac{\alpha_{2k}-\alpha_k^2}{n}$$

(iii) Si $E(X^{2k})=\alpha_{2k}<\infty$, entonces $m_k \xrightarrow{P} \alpha_k$ y $m_k \xrightarrow{\text{c.s.}} \alpha_k$.

(iv) Si $E(X^{2k})=\alpha_{2k}<\infty$, entonces

$$\frac{\sqrt{n}(m_k-\alpha_k)}{\sqrt{\alpha_{2k}-\alpha_k^2}} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$$

Demostración.

(i) En efecto,

$$E(m_k)=E\Big(\frac{1}{n}\sum_{i=1}^{n}X_i^k\Big)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^k)=\frac{1}{n}\sum_{i=1}^{n}\alpha_k=\frac{1}{n}n\alpha_k=\alpha_k$$

(ii) En efecto,

$$\begin{aligned}\mathrm{Var}(m_k)&=\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}X_i^k\Big)=\frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n}X_i^k\Big)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i^k)=\frac{1}{n^2}n\mathrm{Var}(X^k)=\frac{1}{n}\mathrm{Var}(X^k)\\&=\frac{E(X^{2k})-E(X^k)^2}{n}=\frac{\alpha_{2k}-\alpha_k^2}{n}\end{aligned}$$

En la tercera igualdad se ha usado que las variables X_i son independientes.

(iii) Sea $\varepsilon>0$. Por la desigualdad de Chebyshev,

$$P(|m_k-E(m_k)|>\varepsilon)\le\frac{\mathrm{Var}(m_k)}{\varepsilon^2}$$

Usando los dos apartados anteriores,

$$P(|m_k-\alpha_k|>\varepsilon)\le\frac{\alpha_{2k}-\alpha_k^2}{n\varepsilon^2}$$

En consecuencia, tomando límites,

$$\lim_{n\to\infty}P(|m_k-\alpha_k|>\varepsilon)=0,$$

lo que prueba que $m_k \xrightarrow{P} \alpha_k$. La convergencia casi segura es inmediata a partir del teorema de Kolmogorov.

(*iv*) Nótese que, si $S_n = X_1^k + \ldots + X_n^k$, se tiene que $S_n = nm_k$. Además,

$$\frac{S_n - nE(X^k)}{\sqrt{n\mathrm{Var}(X^k)}} = \frac{S_n - n\alpha_k}{\sqrt{n(\alpha_{2k} - \alpha_k^2)}} = \frac{n(m_k - \alpha_k)}{\sqrt{n}\sqrt{\alpha_{2k} - \alpha_k^2}} = \frac{\sqrt{n}(m_k - \alpha_k)}{\sqrt{\alpha_{2k} - \alpha_k^2}}$$

Como las variables aleatorias X_i^k son independientes y poseen la misma distribución, se puede aplicar el teorema de Lindeberg-Lévy para concluir la demostración. □

1.2.6. Momento muestral respecto a la media

Definición 9. *Dada una muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de una variable aleatoria* X*, se define el* momento muestral de orden k respecto a la media *como*

$$b_k := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

Proposición 5. *Sea* X *una variable aleatoria y sea* $(X_1, \ldots, X_n)$ *una muestra aleatoria simple de* X*. Si* $E(X^{2k}) = \alpha_{2k} < \infty$*, entonces*

$$b_k \xrightarrow{P} \mu_k$$

Demostración. Usando la fórmula del binomio de Newton,

$$b_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{k}\binom{k}{j}(-\overline{X})^{k-j}X_i^j = \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}\overline{X}^{k-j}\frac{1}{n}\sum_{i=1}^{n}X_i^j = \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}m_1^{k-j}m_j$$

Tomando límites en probabilidad en esta igualdad y usando que $m_j \xrightarrow{P} \alpha_j$ y $m_1^{k-j} \xrightarrow{P} \alpha_1^{k-j}$ para todo $j \in \{1, \ldots, k\}$, se tiene

$$b_k \xrightarrow{P} \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}\alpha_1^{k-j}\alpha_j = \sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}E(X)^{k-j}E(X^j) = \sum_{j=0}^{k}\binom{k}{j}(-E(X))^{k-j}E(X^j)$$

Ahora bien, usando de nuevo la fórmula del binomio y la linealidad de la esperanza,

$$\mu_k = E((X - E(X))^k) = E\Big(\sum_{j=0}^{k}\binom{k}{j}(-E(X))^{k-j}X^j\Big) = \sum_{j=0}^{k}\binom{k}{j}(-E(X))^{k-j}E(X^j)$$

Si se comparan las dos últimas expresiones, la demostración ha terminado. □

1.2.7. Máximo y mínimo

Definición 10. *Dada una muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de una variable aleatoria* X*, se definen el* máximo de $(X_1, \ldots, X_n)$ *y el* mínimo de $(X_1, \ldots, X_n)$ *como*

$$X_{(n)} := \max_{i=1,\ldots,n}\{X_i\} \qquad \text{y} \qquad X_{(1)} := \max_{i=1,\ldots,n}\{X_i\},$$

respectivamente.

Proposición 6. *Sea* X *una variable aleatoria y sea* $(X_1, \ldots, X_n)$ *una muestra aleatoria simple de* X*. Si* F *es la función de distribución de* X*, entonces la función de distribución de* $X_{(n)}$ *viene dada por*

$$F_{X_{(n)}}(x) = F(x)^n,$$

y la de $X_{(1)}$*, por*

$$F_{X_{(1)}}(x) = 1 - (1 - F(x))^n$$

Demostración. En efecto, si $x \in \mathbb{R}$,

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(\max_{i=1,\dots,n} \{X_i\} \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} F_{X_i}(x) = F(x)^n,$$

donde en la cuarta igualdad se ha usado que las variables X_i son independientes, y en la última, que todas tienen la misma función de distribución que X. Respecto al mínimo,

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) = 1 - P(\min_{i=1,\dots,n} \{X_i\} > x) = 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - \prod_{i=1}^{n} (1 - F_{X_i}(x)) = 1 - (1 - F(x))^n, \end{aligned}$$

donde, una vez más, se ha usado la independencia de las variables X_i y también que $F_{X_i} = F$ para todo $i = 1, \dots, n$. □

1.2.8. Estadístico ordenado

Definición 11. *Dada una muestra aleatoria simple $(X_1, \dots, X_n)$ de una variable aleatoria, el* estadístico ordenado de *$(X_1, \dots, X_n)$ es el vector aleatorio $(X_{(1)}, \dots, X_{(n)})$, donde $X_{(1)} = \min\{X_1, \dots, X_n\}$ y, para cada $i = 2, \dots, n$,*

$$X_{(i)} = \min(\{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(i-1)}\})$$

Para que la definición anterior no cause confusión, puede decirse de forma poco rigurosa que $X_{(r)}$ es la r-ésima componente “más pequeña” de la muestra. Así, según esta definición, $X_{(n)}$ es el máximo de $(X_1, \dots, X_n)$, tal y como se había definido antes.

Proposición 7. *Sea X una variable aleatoria y sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de X. Si F es la función de distribución de X, entonces, para todo $r = 1, \dots, n$ y todo $x \in \mathbb{R}$,*

$$P(X_{(r)} \leq x) = \sum_{j=r}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}$$

Demostración. Sea $x \in \mathbb{R}$. Se va a sacrificar cierta rigurosidad matemática con objeto de facilitar la comprensión. El suceso $X_{(r)} \leq x$ es equivalente al suceso

al menos r de los X_i son menores o iguales que x

A su vez, este suceso es equivalente a

r de los X_i son menores o iguales que x y $n - r$ son mayores que x, o

$r + 1$ de los X_i son menores o iguales que x y $n - r - 1$ son mayores que x, o

...

n de los X_i son menores o iguales que x y 0 son mayores que x,

o lo que es lo mismo,

$$\bigcup_{j=r}^{n} (j \textit{ de los } X_i \textit{ son menores o iguales que } x \textit{ y } n - j \textit{ son mayores que } x)$$

Ahora bien, la probabilidad de que j de los X_i sean menores o iguales que x y $n - j$ sean mayores que x es precisamente

$$\binom{n}{j} P(X_i \leq x)^j P(X_i > x)^{n-j} = \binom{n}{j} F(x)^j (1 - F(x))^{n-j},$$

Por tanto,

$$P\Big(\bigcup_{j=r}^{n}(j \textit{ de los } X_i \textit{ son menores o iguales que } x \textit{ y } n-j \textit{ son mayores que } x)\Big) = \sum_{j=r}^{n}\binom{n}{j}F(x)^j(1-F(x))^{n-j},$$

donde se ha usado que, para $j = r,\dots,n$, los sucesos

$$j \textit{ de los } X_i \textit{ son menores o iguales que } x \textit{ y } n-j \textit{ son mayores que } x$$

son disjuntos. □

1.3. Función de distribución empírica

Definición 12. *Dada una muestra aleatoria simple* $(X_1,\dots,X_n)$ *de una variable aleatoria* X*, se define la* función de distribución empírica de la muestra *como*

$$F_n^*(x) := \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{(-\infty,x]}(X_i)$$

para cada $x \in \mathbb{R}$.

Aunque suele decirse que F_n^* es una función de $\mathbb{R}$ en $[0,1]$, en realidad es una función de $\mathbb{R}$ en el conjunto de variables aleatorias sobre $(\Omega,\mathcal{A},P)$, siendo éste el espacio de probabilidad sobre el cual X_i es variable aleatoria para cada $i \in \{1,\dots,n\}$. Así, para cada $x \in \mathbb{R}$, $F_n^*(x)$ es una variable aleatoria dada por

$$F_n^*(x)(\omega) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{(-\infty,x]}(X_i(\omega))$$

para cada $\omega \in \Omega$. Además,

$$F_n^*(x) = \begin{cases} 0 & \text{si } x < X_{(1)} \\ \frac{1}{n} & \text{si } X_{(1)} \le x < X_{(2)} \\ \frac{2}{n} & \text{si } X_{(2)} \le x < X_{(3)} \\ \vdots & \\ 1 & \text{si } x \ge X_{(n)} \end{cases}$$

Proposición 8. *Sea* X *una variable aleatoria, sea* $(X_1,\dots,X_n)$ *una muestra aleatoria simple de* X *y sea* F *la función de distribución de* X*. Para todo* $x \in \mathbb{R}$*, se verifican las siguientes propiedades:*

(i) $nF_n^*(x) \sim Bin(n,F(x))$.

(ii) $E(F_n^*(x)) = F(x)$.

(iii) $\mathrm{Var}(F_n^*(x)) = \dfrac{F(x)(1-F(x))}{n}$.

(iv) $F_n^*(x) \xrightarrow{P} F(x)$.

(v) $F_n^*(x) \xrightarrow{\text{c.s.}} F(x)$.

(vi) $\dfrac{\sqrt{n}(F_n^*(x)-F(x))}{\sqrt{F(x)(1-F(x))}} \xrightarrow{\mathcal{L}} Z \sim N(0,1)$.

Demostración.

(i) Nótese que

$$P(\mathcal{I}_{(-\infty,x]}(X_i)=1) = P(X_i \le x) = F(x), \qquad P(\mathcal{I}_{(-\infty,x]}(X_i)=0) = P(X_i > x) = 1-F(x)$$

En consecuencia, para todo $i \in \{1,\dots,n\}$,

$$\mathcal{I}_{(-\infty,x]}(X_i) \sim \mathrm{Be}(F(x)),$$

luego

$$\sum_{i=1}^{n} \mathcal{I}_{(-\infty,x]}(X_i) = nF_n^*(x) \sim \mathrm{Bin}(n, F(x))$$

(ii) Se tiene que

$$E(F_n^*(x)) = \frac{1}{n}E(nF_n^*(x)) = \frac{1}{n}nF(x) = F(x),$$

donde en la segunda igualdad se ha utilizado el apartado anterior, teniendo en cuenta que la esperanza de una distribución binomial de parámetros n y p es np.

(iii) Se tiene que

$$\begin{aligned}\mathrm{Var}(F_n^*(x)) = \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{(-\infty,x]}(X_i)\Big) &= \frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n}\mathcal{I}_{(-\infty,x]}(X_i)\Big) = \frac{1}{n^2}\mathrm{Var}(nF_n^*(x)) \\ &= \frac{nF(x)(1-F(x))}{n^2} = \frac{F(x)(1-F(x))}{n},\end{aligned}$$

donde en la segunda igualdad se ha usado la independencia de las variables X_i y en la cuarta se ha utilizado el apartado primero, teniendo en cuenta que la varianza de una distribución binomial de parámetros n y p es $np(1-p)$.

(iv) Sea $\varepsilon > 0$. La desigualdad de Chebyshev permite afirmar

$$P(|F_n^*(x) - E(F_n^*(x))| > \varepsilon) \leq \frac{\mathrm{Var}(F_n^*(x))}{\varepsilon^2},$$

o lo que es lo mismo, por los dos apartados anteriores,

$$P(|F_n^*(x) - F(x)| > \varepsilon) \leq \frac{F(x)(1-F(x))}{n\varepsilon^2},$$

Tomando límites,

$$\lim_{n\to\infty} P(|F_n^*(x) - F(x)| > \varepsilon) = 0,$$

lo que prueba que $F_n^*(x) \xrightarrow{P} F(x)$.

(v) Consecuencia directa del teorema de Kolmogorov.

(vi) Sea

$$S_n = \sum_{i=1}^{n}\mathcal{I}_{(-\infty,x]}(X_i) = nF_n^*(x)$$

Entonces

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \frac{nF_n^*(x) - nF(x)}{\sqrt{n^2\mathrm{Var}(F_n^*(x))}} = \frac{nF_n^*(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} = \frac{\sqrt{n}(F_n^*(x) - F(x))}{\sqrt{F(x)(1-F(x))}},$$

y el teorema de Lindeberg-Lévy concluye la demostración. □

Muestra aleatoria simple de la distribución normal

En este tema, como el título indica, se relacionará lo visto en el tema anterior con la distribución de probabilidad de más renombre, pues el muestreo aleatorio simple de la distribución normal da lugar a otras distribuciones de buena reputación.

2.1. Distribución normal

Definición 13. *Dados $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, se dice que X sigue una* distribución normal de parámetros μ y σ^2, *y se denota $X \sim N(\mu, \sigma^2)$, cuando para cada $x \in \mathbb{R}$*

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

Si $\mu = 0$ y $\sigma^2 = 1$, se habla de distribución normal estándar.

Proposición 9. *Si $X \sim N(\mu, \sigma^2)$, entonces*

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

Demostración. Sea $\varphi \colon \mathbb{R} \to \mathbb{R}$ la función definida por

$$\varphi(x) = \frac{x-\mu}{\sigma}$$

La función φ es de clase 1 y estrictamente monótona, así que, por la Proposición 35,

$$Z = \varphi \circ X = \frac{X-\mu}{\sigma}$$

es una variable aleatoria absolutamente continua, y su función de densidad es

$$f_Y(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|,$$

donde $\varphi^{-1} \colon \mathbb{R} \to \mathbb{R}$ está definida por $\varphi^{-1}(y) = \sigma y + \mu$. Por tanto,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\sigma y+\mu-\mu)^2}{2\sigma^2}} |\sigma| = \frac{|\sigma|}{\sqrt{2\pi}|\sigma|} e^{-\frac{\sigma^2 y^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

concluyéndose que $Z \sim N(0,1)$. □

Como consecuencia de la proposición anterior, el cálculo de probabilidades de distribuciones normales se reduce al cálculo de probabilidades de la distribución normal estándar, para el que se puede hacer uso de las famosas tablas de probabilidades.

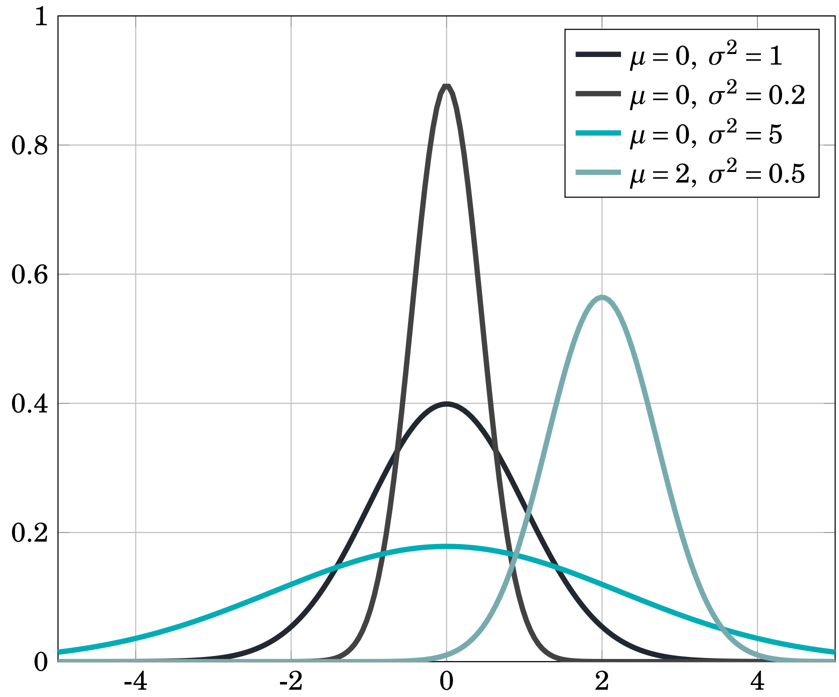

Figura 1. Gráficas de las funciones de densidad de algunas distribuciones normales.

Proposición 10. *Si $Z \sim N(0,1)$, entonces $M_Z(t) = e^{\frac{t^2}{2}}$ para todo $t \in \mathbb{R}$.*

Demostración. En efecto,

$$M_Z(t) = E(e^{tZ}) = \int_{\mathbb{R}} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz)}\,dz = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz+t^2)+\frac{1}{2}t^2}\,dz$$
$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} e^{\frac{t^2}{2}}\,dz = e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2}\,dz \overset{(*)}{=} e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = e^{\frac{t^2}{2}},$$

donde en $(*)$ se ha realizado el cambio de variable $z = t + x$, $dz = dx$, y en la igualdad siguiente se ha usado que la integral sobre $\mathbb{R}$ de la función de densidad de Z vale 1. □

Corolario 1. *Si $X \sim N(\mu,\sigma^2)$, entonces*

(i) $M_X(t) = e^{\frac{1}{2}t^2\sigma^2 + t\mu}$.

(ii) $E(X) = \mu$.

(iii) $\mathrm{Var}(X) = \sigma^2$.

Demostración. En efecto, usando que $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ y empleando la proposición anterior,

$$M_X(t) = M_{\sigma Z+\mu}(t) = E(e^{t(\sigma Z+\mu)}) = E(e^{t\sigma Z})E(e^{t\mu}) = M_Z(t\sigma)e^{t\mu} = e^{\frac{1}{2}t^2\sigma^2} e^{t\mu} = e^{\frac{1}{2}t^2\sigma^2+t\mu}$$

Por tanto, $E(X) = M_X'(0) = \mu$, y como $E(X^2) = M_X''(0) = \sigma^2 + \mu^2$, el apartado tercero está regalado. □

Proposición 11. *Se verifican las siguientes propiedades:*

(i) Si $X \sim N(\mu_X,\sigma_X^2)$ e $Y \sim N(\mu_Y,\sigma_Y^2)$ son variables aleatorias independientes, entonces

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

(ii) Si $X \sim N(\mu,\sigma^2)$, entonces, para todo $a \in \mathbb{R}$,

$$aX \sim N(a\mu, a^2\sigma^2)$$

Demostración. Ejercicio. □

Corolario 2. *Si* $(X_1,\ldots,X_n)$ *es una muestra aleatoria simple de* $X \sim N(\mu,\sigma^2)$*, entonces*

$$\overline{X} \sim N\Big(\mu, \frac{\sigma^2}{n}\Big)$$

Demostración. Por el primer apartado de la proposición anterior,

$$\sum_{i=1}^{n} X_i \sim N\Big(\sum_{i=1}^{n}\mu, \sum_{i=1}^{n}\sigma^2\Big) = N(n\mu, n\sigma^2),$$

luego

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\Big(\frac{1}{n}n\mu, \frac{1}{n^2}n\sigma^2\Big) = N\Big(\mu, \frac{\sigma^2}{n}\Big),$$

donde se ha aplicado el segundo apartado de la proposición anterior. □

2.2. Distribución gamma

La estrecha relación de la distribución gamma con la distribución normal quedará al descubierto en la sección siguiente. Por ahora, merece la pena mencionar las propiedades básicas de esta distribución.

Definición 14. *Dados* $\alpha > 0$ *y* $\beta > 0$*, se dice que* X *sigue una* distribución gamma de parámetros α y β*, y se denota* $X \sim Ga(\alpha,\beta)$*, cuando*

$$f(x) = \begin{cases} \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & si\ x > 0 \\ 0 & en\ otro\ caso, \end{cases}$$

donde

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x}\, dx$$

es la denominada función gamma.

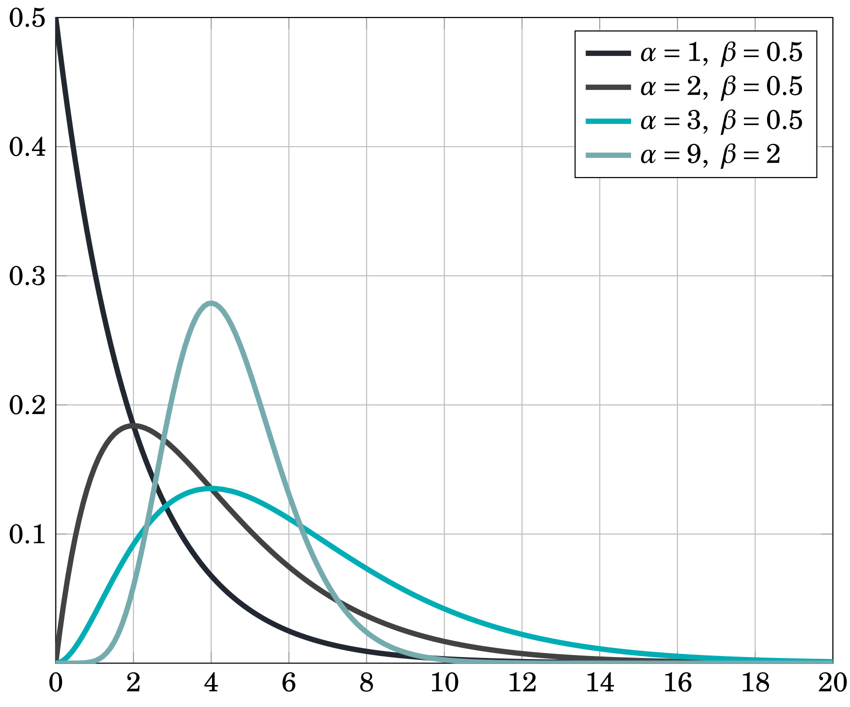

Figura 2. Gráficas de las funciones de densidad de algunas distribuciones gamma.

Proposición 12. *La función gamma verifica las siguientes propiedades:*

(i) $\Gamma(z+1) = z\Gamma(z)$ *para todo* $z > 0$.

(ii) $\Gamma(n+1) = n!$ *para todo* $n \in \mathbb{N}$.

(iii) $\Gamma\big(\frac{1}{2}\big) = \sqrt{\pi}$.

Demostración. Solo se va a demostrar el último apartado. Se tiene que

$$\Gamma\big(\frac{1}{2}\big) = \int_0^\infty x^{-\frac{1}{2}} e^{-x}\, dx = \int_0^\infty \frac{e^{-x}}{\sqrt{x}}\, dx \overset{(*)}{=} \int_0^\infty 2e^{-y^2}\, dy = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi},$$

donde en $(*)$ se ha realizado el cambio de variable $x = y^2$, $dx = 2y\,dy$. □

Proposición 13. *Si* $X \sim Ga(\alpha, \beta)$*, entonces*

(i) $E(X) = \dfrac{\alpha}{\beta}$.

(ii) $\mathrm{Var}(X) = \dfrac{\alpha}{\beta^2}$.

(iii) $M_X(t) = \big(1 - \dfrac{t}{\beta}\big)^{-\alpha}$ *siempre que* $t < \beta$.

(iv) $E(X^r) = \dfrac{\Gamma(\alpha + r)}{\Gamma(\alpha)\beta^r}$ *para todo* $r \in \mathbb{N}$.

Demostración. Ejercicio. □

Proposición 14. *Se verifican las siguientes propiedades:*

(i) Si $X_1, \ldots, X_n$ *son variables aleatorias independientes con* $X_i \sim Ga(\alpha_i, \beta)$ *para cada* $i \in \{1, \ldots, n\}$*, entonces*

$$\sum_{i=1}^n X_i \sim Ga\big(\sum_{i=1}^n \alpha_i, \beta\big)$$

(ii) Si $X \sim Ga(\alpha, \beta)$*, entonces* $cX \sim Ga(\alpha, \frac{\beta}{c})$ *para todo* $c > 0$.

Demostración. Ejercicio. □

Proposición 15. *Si* $Z \sim N(0,1)$*, entonces*

$$Z^2 \sim Ga\big(\frac{1}{2}, \frac{1}{2}\big)$$

Demostración. Tenemos que, para todo $z \geq 0$,

$$F_{Z^2}(z) = P(Z^2 \leq z) = P(|Z| \leq \sqrt{z}) = P(-\sqrt{z} \leq Z \leq \sqrt{z}) = \int_{-\sqrt{z}}^{\sqrt{z}} f_Z(t)\,dt = \int_{-\infty}^{\sqrt{z}} f_Z(t)\,dt - \int_{-\infty}^{-\sqrt{z}} f_Z(t)\,dt$$

$$= F_Z(\sqrt{z}) - (1 - \int_{-\sqrt{z}}^{\infty} f_Z(t)\,dt) \overset{(*)}{=} F_Z(\sqrt{z}) - (1 - \int_{-\infty}^{\sqrt{z}} f_Z(t)\,dt) = 2F_Z(\sqrt{z}) - 1,$$

donde en $(*)$ se ha usado que la función de densidad de Z es una función par. Por tanto,

$$f_{Z^2}(z) = F'_{Z^2}(z) = 2F'_Z(\sqrt{z})\frac{1}{2\sqrt{z}} = \frac{f_Z(\sqrt{z})}{\sqrt{z}} = \frac{1}{\sqrt{2\pi z}} e^{-\frac{z}{2}} = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} z^{\frac{1}{2}-1} e^{-\frac{1}{2}z},$$

lo que demuestra que $Z^2 \sim Ga(\frac{1}{2}, \frac{1}{2})$. □

2.3. Distribución χ^2 de Pearson

Definición 15. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X \sim N(0,1)$. La variable aleatoria*

$$V = \sum_{i=1}^{n} X_i^2$$

se dice que sigue una distribución χ^2 con n grados de libertad, *y se denota $V \sim \chi_n^2$.*

Proposición 16. *La función de densidad de $V \sim \chi_n^2$ es*

$$f_n(v) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} v^{\frac{n}{2}-1} e^{-\frac{v}{2}} & \text{si } v > 0 \\ 0 & \text{en otro caso} \end{cases}$$

En otras palabras, $V \sim \mathrm{Ga}(\frac{n}{2}, \frac{1}{2})$.

Demostración. No es asunto de esta asignatura. □

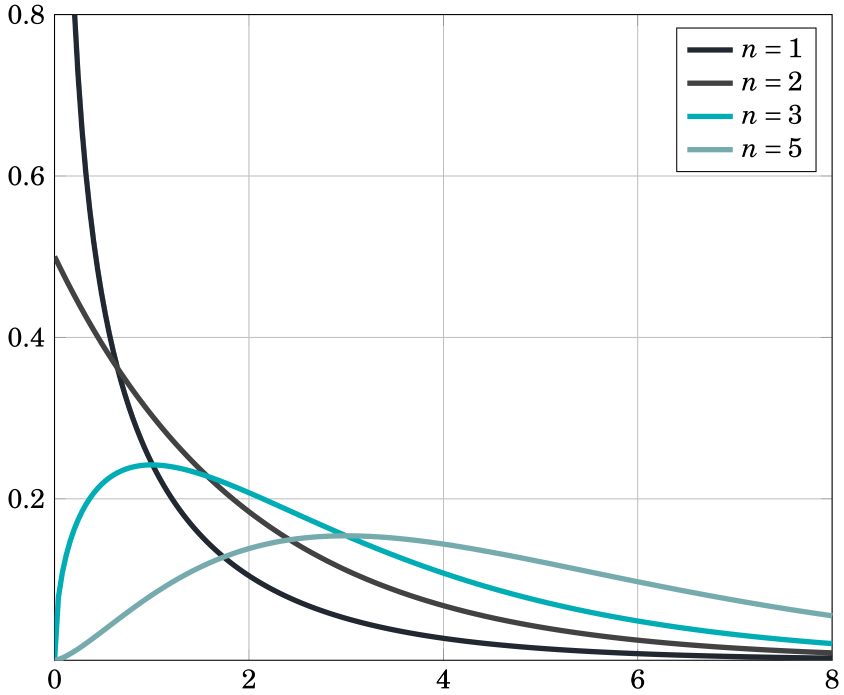

Figura 3. Gráficas de las funciones de densidad de algunas distribuciones χ_n^2.

Corolario 3. *Si $V \sim \chi_n^2$, entonces*

(i) $E(V) = n$.

(ii) $\mathrm{Var}(V) = 2n$.

(iii) $M_V(t) = (1-2t)^{-\frac{n}{2}}$ *siempre que* $t < \dfrac{1}{2}$.

(iv) $E(X^r) = \dfrac{2^r \Gamma(r + \frac{n}{2})}{\Gamma(\frac{n}{2})}$.

Demostración. No hay más que aplicar la proposición anterior y la Proposición 13. □

Corolario 4. *Si $V_i \sim \chi^2_{n_i}$ $(i = 1, \dots, k)$ son variables aleatorias independientes, entonces*

$$\sum_{i=1}^{k} V_i \sim \chi^2_{\sum_{i=1}^{k} n_i}$$

Demostración. No hay más que aplicar la proposición anterior y la Proposición 14. □

Corolario 5. *Si $X \sim N(0,1)$, entonces $X^2 \sim \chi^2_1$.*

Demostración. No hay más que aplicar la proposición anterior y la Proposición 15. □

Corolario 6. *Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple de $X \sim N(\mu, \sigma^2)$, entonces*

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$$

Demostración. Llamamos, para cada $i \in \{1, \dots, n\}$, $Y_i = \left(\frac{X_i - \mu}{\sigma}\right)^2$. Por el corolario anterior se tiene $Y_i \sim \chi^2_1$ para todo $i \in \{1, \dots, n\}$, luego

$$M_{\sum_{i=1}^{n} Y_i}(t) = \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n} (1 - 2t)^{-\frac{1}{2}} = (1 - 2t)^{-\frac{n}{2}}$$

para todo $t < \frac{1}{2}$, de donde se deduce inmediatamente la afirmación del enunciado. □

Corolario 7. *Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple de $X \sim N(\mu, \sigma^2)$, entonces*

$$\left(\frac{\overline{X} - \mu}{\sigma}\right)^2 n \sim \chi^2_1$$

Demostración. No hay más que aplicar el corolario anterior con $n = 1$ y el Corolario 2. □

2.4. Distribución *t* de Student

Definición 16. *Si $Z \sim N(0,1)$ y $V \sim \chi^2_n$ son independientes, la variable aleatoria*

$$T = \frac{Z}{\sqrt{\frac{V}{n}}}$$

se dice que sigue una distribución t con n grados de libertad, *y se denota $T \sim t_n$.*

Proposición 17. *Una variable aleatoria T que sigue una distribución t con n grados de libertad verifica las siguientes propiedades:*

(i) La función de densidad de T es

$$f_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

para todo $t \in \mathbb{R}$.

(ii) Si $n > 1$,

$$E(T) = 0$$

(iii) Si $n > 2$,

$$\mathrm{Var}(T) = \frac{n}{n-2}$$

Demostración. No es asunto de esta asignatura. □

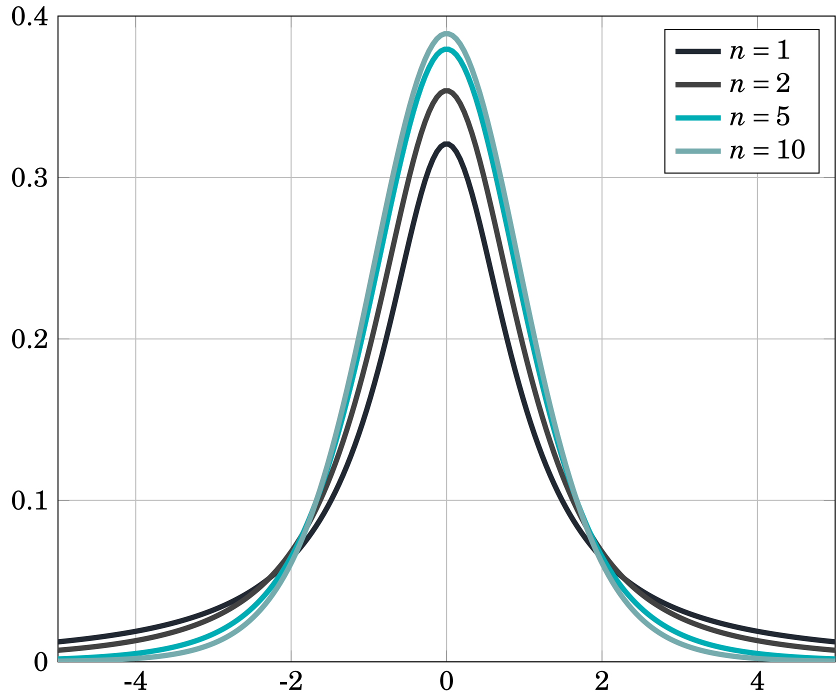

Figura 4. Gráficas de las funciones de densidad de algunas distribuciones t_n.

2.5. Distribución F de Snedecor

Definición 17. *Si $V \sim \chi^2_n$ y $W \sim \chi^2_m$ son independientes, la variable aleatoria*

$$F = \frac{\frac{V}{n}}{\frac{W}{m}}$$

se dice que sigue una distribución F con n y m grados de libertad, *y se denota $F \sim F_{n,m}$.*

Proposición 18. *Una variable aleatoria F que sigue una distribución F con n y m grados de libertad verifica las siguientes propiedades:*

(i) La función de densidad de F es

$$f_{n,m}(x) = \begin{cases} \dfrac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\Big(\dfrac{n}{m}\Big)^{\frac{n}{2}} x^{\frac{n}{2}-1}\Big(1+\dfrac{n}{m}x\Big)^{-\frac{n+m}{2}} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

(ii) Si $m > 2$,

$$E(F) = \frac{m}{m-2}$$

(iii) Si $m > 4$,

$$\mathrm{Var}(F) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$$

Demostración. No es asunto de esta asignatura. □

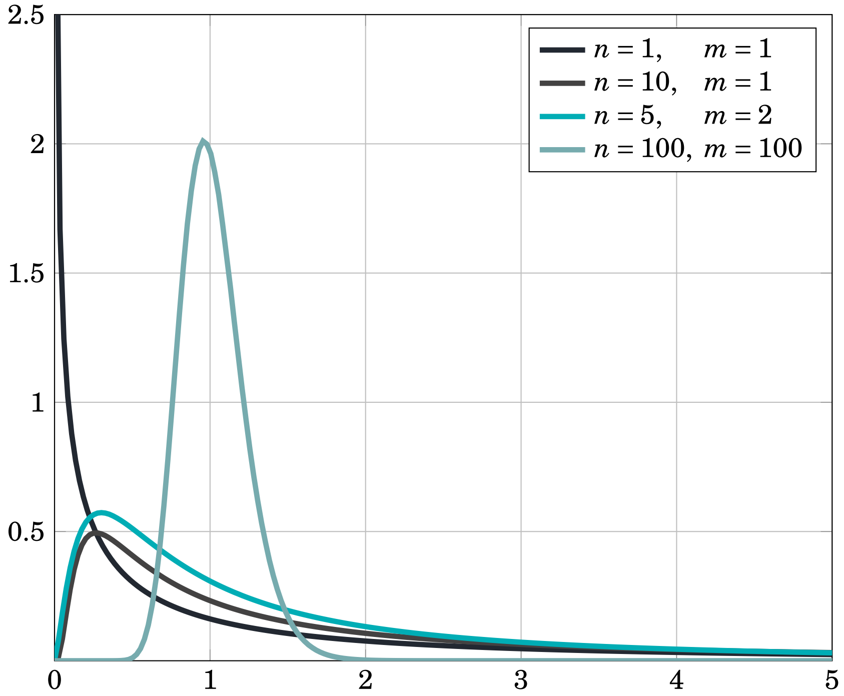

Figura 5. Gráficas de las funciones de densidad de algunas distribuciones $F_{n,m}$.

2.6. Propiedades de la media muestral y la cuasivarianza muestral

Proposición 19. *Considérese una muestra aleatoria simple* $(X_1,\ldots,X_n)$ *de* $X \sim N(\mu,\sigma^2)$. *Entonces* $\overline{X}$ *y* $(X_1-\overline{X},\ldots,X_n-\overline{X})$ *son independientes, es decir, se verifica*

$$M_{(\overline{X},X_1-\overline{X},\ldots,X_n-\overline{X})}(t,t_1,\ldots,t_n) = M_{\overline{X}}(t)M_{(X_1-\overline{X},\ldots,X_n-\overline{X})}(t_1,\ldots,t_n)$$

para todo $t \in \mathbb{R}$ *y todo* $(t_1,\ldots,t_n) \in \mathbb{R}^n$.

Demostración. Antes de empezar, dado $(t_1,\ldots,t_n)\in\mathbb{R}^n$, será conveniente denotar

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

Se tiene que

$$\begin{aligned} M_{(X_1-\overline{X},\ldots,X_n-\overline{X})}(t_1,\ldots,t_n) &= E\left(e^{\sum_{i=1}^n t_i(X_i-\overline{X})}\right) = E\left(e^{\sum_{i=1}^n t_iX_i-\overline{X}\sum_{i=1}^n t_i}\right) = E\left(e^{\sum_{i=1}^n t_iX_i-\bar{t}\sum_{i=1}^n X_i}\right) \\ &= E\left(e^{\sum_{i=1}^n X_i(t_i-\bar{t})}\right) = E\Big(\prod_{i=1}^n e^{X_i(t_i-\bar{t})}\Big) = \prod_{i=1}^n E\left(e^{X_i(t_i-\bar{t})}\right) = \prod_{i=1}^n M_{X_i}(t_i-\bar{t}) \end{aligned}$$

Además,

$$M_{\overline{X}}(t) = E(e^{t\overline{X}}) = E\left(e^{\sum_{i=1}^n \frac{t}{n}X_i}\right) = E\Big(\prod_{i=1}^n e^{\frac{t}{n}X_i}\Big) = \prod_{i=1}^n E(e^{\frac{t}{n}X_i}) = \prod_{i=1}^n M_{X_i}\Big(\frac{t}{n}\Big) = \prod_{i=1}^n e^{\frac{1}{2}\frac{t^2}{n^2}\sigma^2+\frac{t}{n}\mu} = e^{\frac{1}{2}\frac{t^2}{n}\sigma^2+t\mu}$$

Por otro lado,

$$\begin{aligned}
M_{(\overline{X},X_1-\overline{X},\dots,X_n-\overline{X})}(t,t_1,\dots,t_n) &= E\big(e^{t\overline{X}+\sum_{i=1}^n t_i(X_i-\overline{X})}\big) = E\big(e^{t\overline{X}+\sum_{i=1}^n t_iX_i-\sum_{i=1}^n t_i\overline{X})}\big) \\
&= E\big(e^{\sum_{i=1}^n t_iX_i-\overline{X}(\sum_{i=1}^n t_i-t)}\big) = E\big(e^{\sum_{i=1}^n X_i(t_i-\frac{1}{n}(\sum_{i=1}^n t_i-t))}\big) \\
&= E\big(\prod_{i=1}^n e^{X_i(t_i-\frac{1}{n}(\sum_{i=1}^n t_i-t))}\big) = \prod_{i=1}^n E\big(e^{X_i(t_i-\frac{1}{n}(\sum_{i=1}^n t_i-t))}\big) \\
&= \prod_{i=1}^n E\big(e^{X_i(t_i-\overline{t}+\frac{t}{n})}\big) = \prod_{i=1}^n E\big(e^{X_i(t_i-\overline{t})}e^{X_i\frac{t}{n}}\big) = \prod_{i=1}^n E(e^{X_i(t_i-\overline{t})})E(e^{X_i\frac{t}{n}}) \\
&= \prod_{i=1}^n M_{X_i}(t_i-\overline{t})M_{X_i}(\frac{t}{n}) = \prod_{i=1}^n M_{X_i}(t_i-\overline{t})e^{\frac{1}{2}\frac{t^2}{n^2}\sigma^2+\frac{t}{n}\mu} \\
&= e^{\frac{1}{2}\frac{t^2}{n}\sigma^2+t\mu}\prod_{i=1}^n M_{X_i}(t_i-\overline{t}),
\end{aligned}$$

así que se verifica la igualdad del enunciado. □

Proposición 20 (Lema de Fisher). *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X \sim N(\mu,\sigma^2)$. Entonces*

(i) $\overline{X}$ *y* S_c^2 *son variables aleatorias independientes.*

(ii) $\overline{X} \sim N\big(\mu,\dfrac{\sigma^2}{n}\big)$.

(iii) $\dfrac{(n-1)S_c^2}{\sigma^2} \sim \chi^2_{n-1}$.

(iv) $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$

Demostración.

(i) Considérese la función $g\colon \mathbb{R}^n \to \mathbb{R}$ dada por

$$g(t_1,\dots,t_n) = \frac{1}{n-1}\sum_{i=1}^n t_i^2$$

Como $\overline{X}$ e $Y = (X_1-\overline{X},\dots,X_n-\overline{X})$ son independientes y g es continua, entonces $\overline{X}$ y $g(Y) = S_c^2$ también son independientes.

(ii) Como $X_i \sim N(\mu,\sigma^2)$ para todo $i \in \{1,\dots,n\}$, entonces

$$\sum_{i=1}^n X_i \sim N\big(\sum_{i=1}^n \mu, \sum_{i=1}^n \sigma^2\big) = N(n\mu, n\sigma^2),$$

luego, empleando la Proposición 11,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim N\big(\mu,\frac{\sigma^2}{n}\big),$$

(iii) Nótese que

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

Se sabe, por el Corolario 6, que

$$V = \sum_{i=1}^n \big(\frac{X_i-\mu}{\sigma}\big)^2 \sim \chi^2_n$$

Pero

$$V=\sum_{i=1}^{n}\frac{(X_i-\overline{X}+\overline{X}-\mu)^2}{\sigma^2}=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}+\sum_{i=1}^{n}\frac{(\overline{X}-\mu)^2}{\sigma^2}+\cancel{\frac{2(\overline{X}-\mu)}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})}$$
$$=\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}+\Big(\frac{\overline{X}-\mu}{\sigma}\Big)^2 n=U+W$$

El Corolario 7 también permite afirmar que $W\sim\chi_1^2$. Por tanto, por el Corolario 3,

$$M_V(t)=(1-2t)^{-\frac{n}{2}}\qquad \text{y}\qquad M_W(t)=(1-2t)^{-\frac{1}{2}}$$

para todo $t<\frac{1}{2}$. Así, usando que U, V y W son independientes,

$$M_U(t)=M_{V-W}(t)=\frac{M_V(t)}{M_W(t)}=(1-2t)^{-\frac{n}{2}+\frac{1}{2}}=(1-2t)^{-\frac{n-1}{2}}$$

Se concluye que $U\sim\chi_{n-1}^2$.

(*iv*) Por (*ii*) y por (*iii*), respectivamente,

$$Z=\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim N(0,1)\qquad \text{y}\qquad V=\frac{(n-1)S_c^2}{\sigma^2}\sim\chi_{n-1}^2$$

Por tanto,

$$\frac{Z}{\sqrt{\frac{V}{n-1}}}=\frac{\sqrt{n-1}\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)S_c^2}{\sigma^2}}}=\frac{\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\frac{\sqrt{S_c^2}}{\sigma}}=\frac{(\overline{X}-\mu)\sqrt{n}}{\sqrt{S_c^2}}\sim t_{n-1},$$

teniéndose en cuenta que Z y V son independientes por serlo $\overline{X}$ y S_c^2, como se probó en el apartado primero. □

Corolario 8. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X\sim N(\mu_X,\sigma_X^2)$ y sea $(Y_1,\dots,Y_m)$ una muestra aleatoria simple de $Y\sim N(\mu_Y,\sigma_Y^2)$, siendo X e Y variables aleatorias independientes. Entonces*

(*i*) $\dfrac{S_{c,X}^2\sigma_Y^2}{S_{c,Y}^2\sigma_X^2}\sim F_{n-1,m-1}$

(*ii*) $\sqrt{\dfrac{n+m-2}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\,\dfrac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{(n-1)S_{c,X}^2}{\sigma_X^2}+\frac{(m-1)S_{c,Y}^2}{\sigma_Y^2}}}\sim t_{n+m-2}$

Demostración.

(*i*) Se tiene que

$$\frac{S_{c,X}^2\sigma_Y^2}{S_{c,Y}^2\sigma_X^2}=\frac{\frac{(n-1)S_{c,X}^2}{\sigma_X^2}}{\frac{(m-1)S_{c,Y}^2}{\sigma_Y^2}}\,\frac{m-1}{n-1}=\frac{\frac{V}{n-1}}{\frac{W}{m-1}}\sim F_{n-1,m-1}$$

Se ha usado que, por el lema de Fisher, $V\sim\chi_{n-1}^2$ y $W\sim\chi_{m-1}^2$.

(*ii*) Por el lema de Fisher, se sabe que

$$\overline{X}\sim N\big(\mu_X,\frac{\sigma_X^2}{n}\big)\qquad \text{y}\qquad \overline{Y}\sim N\big(\mu_Y,\frac{\sigma_Y^2}{m}\big)$$

Por tanto,

$$\overline{X}-\overline{Y}\sim N\big(\mu_X-\mu_Y,\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\big)$$

En consecuencia,

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

Además, también por el lema de Fisher,

$$\frac{(n-1)S_{c,X}^2}{\sigma_X^2} \sim \chi_{n-1}^2 \qquad \text{y} \qquad \frac{(m-1)S_{c,Y}^2}{\sigma_Y^2} \sim \chi_{m-1}^2$$

Así,

$$V = \frac{(n-1)S_{c,X}^2}{\sigma_X^2} + \frac{(m-1)S_{c,Y}^2}{\sigma_Y^2} \sim \chi_{n-1+m-1}^2 = \chi_{n+m-2}^2$$

Total, que se tiene

$$\frac{Z}{\sqrt{\frac{V}{n+m-2}}} \sim t_{n+m-2}$$

Si se compara esta variable aleatoria con la del enunciado, no queda nada que probar. □

Estimación puntual paramétrica

Considérese una variable aleatoria X cuya función de distribución depende de un cierto parámetro θ desconocido, del que solo se sabe que pertene a un cierto espacio paramétrico Θ. El objetivo de este tema consistirá en estimar este parámetro θ. Para ello, serán de cierta utilidad los estadísticos muestrales, pues el hecho de conocer, por ejemplo, la media o la varianza de una muestra aleatoria simple de X permitirá obtener cierta información del parámetro θ.

En gran parte del tema será necesario referirse a la función de masa o densidad de X dependiendo de si es discreta o absolutamente continua. Para facilitar la escritura y no tener que distinguir entre función de masa y de densidad constantemente, será frecuente denotar por f_θ a ambas.

3.1. Función de verosimilitud

Definición 18. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde θ pertenece a un espacio paramétrico Θ, y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. La* función de verosimilitud de $(X_1,\dots,X_n)$ *es la función $L\colon \chi^n \times \Theta \to \mathbb{R}$ dada por*

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

Si g_θ es la función de masa o densidad del vector aleatorio $(X_1,\dots,X_n)$, usando la independencia de estas n variables aleatorias puede probarse que

$$g_\theta(x_1,\dots,x_n) = \prod_{i=1}^{n} f_\theta(x_i), \quad (x_1,\dots,x_n) \in \mathbb{R}^n$$

Si a esto se le suma la costumbre de denotar de la misma manera a la función de masa o densidad de $(X_1,\dots,X_n)$ y a la de X, queda justificada la expresión

$$L(x_1,\dots,x_n;\theta) = f_\theta(x_1,\dots,x_n), \qquad (x_1,\dots,x_n;\theta) \in \chi^n \times \Theta$$

Ejemplo 3. Defínase la variable aleatoria

$$X \equiv \begin{cases} 1 & \text{si un alumno duerme menos de 6 horas} \\ 0 & \text{en otro caso} \end{cases}$$

Si p es la probabilidad de que un alumno duerma menos de 6 horas, se tiene $X \sim Ber(p)$. Sea

$$(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10}) = (0,0,0,0,0,1,0,0,0,0) \in \chi^{10}$$

Entonces

$$L(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10};p) = \prod_{i=1}^{10} P_p(X = x_i) = p(1-p)^9$$

3.2. Estadísticos suficientes

Definición 19. *Sea X una variable aleatoria con función de masa o densidad f_θ, sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X y sea $T\colon \mathbb{R}^n \to \mathbb{R}$ un estadístico muestral. Se dice que T es un* estadístico suficiente para θ *si para cada $(x_1,\dots,x_n) \in \chi^n$, el cociente*

$$\frac{f_\theta(x_1,\dots,x_n)}{g_\theta(T(x_1,\dots,x_n))}$$

no depende de θ, donde g_θ es la función de masa o densidad de la variable aleatoria $T(X_1,\dots,X_n)$.

Nótese que, en el caso discreto,

$$\begin{aligned}
\frac{p_\theta(x_1,\dots,x_n)}{q_\theta(T(x_1,\dots,x_n))} &= \frac{P_\theta(X_1 = x_1,\dots,X_n = x_n)}{P_\theta(T(X_1,\dots,X_n) = T(x_1,\dots,x_n))} \\
&= \frac{P_\theta(X_1 = x_1,\dots,X_n = x_n, T(X_1,\dots,X_n) = T(x_1,\dots,x_n))}{P_\theta(T(X_1,\dots,X_n) = T(x_1,\dots,x_n))} \\
&= P_\theta(X_1 = x_1,\dots,X_n = x_n \,\big|\, T(X_1,\dots,X_n) = T(x_1,\dots,x_n))
\end{aligned}$$

Es por esto que será habitual emplear la notación

$$\frac{p_\theta(x_1,\dots,x_n)}{q_\theta(T(x_1,\dots,x_n))} \equiv P(X_1 = x_1,\dots,X_n = x_n \,\big|\, T = t)$$

Ejemplo 4. Sea $X \sim Ber(p)$ y sea (X_1,X_2,X_3) una muestra aleatoria simple de X. Considérense los estadísticos muestrales

$$T_1(X_1,X_2,X_3) = \sum_{i=1}^{3} X_i, \qquad T_2(X_1,X_2,X_3) = \text{máx}\{X_1,X_2,X_3\}$$

Recogemos en una tabla los valores que pueden tomar $(X_1,X_2,X_3), T_1$ y T_2 con probabilidad positiva.

(X_1,X_2,X_3)	T_1	T_2
$(0,0,0)$	0	0
$(1,0,0)$	1	1
$(0,1,0)$	1	1
$(0,0,1)$	1	1
$(1,1,0)$	2	1
$(0,1,1)$	2	1
$(1,0,1)$	2	1
$(1,1,1)$	3	1

Veamos si T_1 es un estadístico suficiente. Se tiene que

$$\begin{aligned}
&P(X_1 = 0, X_2 = 0, X_3 = 0 \,|\, T_1 = 0) = 1, && P(X_1 = 1, X_2 = 1, X_3 = 0 \,|\, T_1 = 2) = \frac{1}{3}, \\
&P(X_1 = 1, X_2 = 0, X_3 = 0 \,|\, T_1 = 1) = \frac{1}{3}, && P(X_1 = 0, X_2 = 1, X_3 = 1 \,|\, T_1 = 2) = \frac{1}{3}, \\
&P(X_1 = 1, X_2 = 1, X_3 = 0 \,|\, T_1 = 1) = \frac{1}{3}, && P(X_1 = 1, X_2 = 0, X_3 = 1 \,|\, T_1 = 2) = \frac{1}{3}, \\
&P(X_1 = 1, X_2 = 0, X_3 = 1 \,|\, T_1 = 1) = \frac{1}{3}, && P(X_1 = 1, X_2 = 1, X_3 = 1 \,|\, T_1 = 3) = 1,
\end{aligned}$$

y como esto no depende de p, entonces T_1 es un estadístico suficiente. Sin embargo, T_2 no lo es, ya que

$$P(X_1 = 1, X_2 = 0, X_3 = 0 \,|\, T_2 = 1) = \frac{p(1-p)^2}{1-(1-p)^3}$$

Ejemplo 5. Sean $X \sim P(\lambda)$ y $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X. Se considera el estadístico muestral
$$T = \sum_{i=1}^{n} X_i$$
Veamos si T es un estadístico suficiente para λ. Si $(x_1,\ldots,x_n) \in (\mathbb{N}\cup\{0\})^n$,
$$P(X_1 = x_1,\ldots,X_n = x_n \,\big|\, T = t) = \frac{P(X_1 = x_1,\ldots,X_n = x_n, T = t)}{P(T = t)}$$
Se tiene que
$$P(X_1 = x_1,\ldots,X_n = x_n, T = t) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} = e^{-n\lambda}\frac{\lambda^t}{\prod_{i=1}^n x_i!}$$
Por otro lado, usando que $T \sim P(n\lambda)$ por ser X_i variables aleatorias independientes con $X_i \sim P(\lambda)$,
$$P(T = t) = e^{-n\lambda}\frac{(n\lambda)^t}{t!}$$
En consecuencia,
$$P(X_1 = x_1,\ldots,X_n = x_n \,\big|\, T = t) = \frac{t!}{n^t \prod_{i=1}^n x_i!}$$
Como esto no depende de λ, se tiene que T es un estadístico suficiente para λ.

Teorema 1 (Teorema de factorización de Fisher-Neyman). *Un estadístico muestral T es suficiente para $\theta \in \Theta$ si y solo si existen funciones $g\colon T(\chi^n)\times\Theta \to \mathbb{R}$, $h\colon \chi^n \to \mathbb{R}$ no negativas tales que*
$$L(x_1,\ldots,x_n;\theta) = g(T(x_1,\ldots,x_n);\theta)h(x_1,\ldots,x_n)$$
para todo $(x_1,\ldots,x_n) \in \chi^n$.

Demostración. Solo se va a probar para el caso discreto. Sea T un estadístico suficiente para θ. Entonces
$$L(x_1,\ldots,x_n;\theta) = P_\theta(X_1 = x_1,\ldots,X_n = x_n) = P_\theta(X_1 = x_1,\ldots,X_n = x_n \,\big|\, T = t)P_\theta(T = t),$$
donde $t = T(x_1,\ldots,x_n)$. Basta considerar
$$h(x_1,\ldots,x_n) = P_\theta(X_1 = x_1,\ldots,X_n = x_n \,\big|\, T = t), \qquad g(t,\theta) = P(T = t),$$
que son funciones no negativas (nótese que h está bien definida porque T es suficiente para θ). Para el recíproco, supóngase que
$$L(x_1,\ldots,x_n;\theta) = g(t;\theta)h(x_1,\ldots,x_n)$$
con h y g no negativas. Dado $t \in T(\chi^n)$, consideremos $A_t = \big\{(y_1,\ldots,y_n) \in \chi^n : T(y_1,\ldots,y_n) = t\big\}$. Entonces se tiene
$$P_\theta(T = t) = P_\theta((X_1,\ldots,X_n) \in A_t) = \sum_{(y_1,\ldots,y_n)\in A_t} p_\theta(y_1,\ldots,y_n) = \sum_{(y_1,\ldots,y_n)\in A_t} g(T(y_1,\ldots,y_n);\theta)h(y_1,\ldots,y_n)$$
Por otra parte,
$$\begin{aligned} P_\theta(X_1 = x_1,\ldots,X_n = x_n, T = t) &= P_\theta(X_1 = x_1,\ldots,X_n = x_n, T(X_1,\ldots,X_n) = T(x_1,\ldots,x_n)) \\ &= P_\theta(X_1 = x_1,\ldots,X_n = x_n) \\ &= L(x_1,\ldots,x_n;\theta) \\ &= g(t;\theta)h(x_1,\ldots,x_n) \end{aligned}$$

En consecuencia,

$$
\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n \,\big|\, T = t) &= \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P_\theta(T = t)} \\
&= \frac{g(t;\theta) h(x_1, \ldots, x_n)}{g(t;\theta) \sum_{(y_1,\ldots,y_n) \in A_t} h(y_1, \ldots, y_n)} \\
&= \frac{h(x_1, \ldots, x_n)}{\sum_{(y_1,\ldots,y_n) \in A_t} h(y_1, \ldots, y_n)}
\end{aligned}
$$

Como esta expresión no depende de θ, se concluye que T es suficiente para θ. □

Ejemplo 6. Sea $X \sim N(\mu, \sigma^2)$ (con σ^2 un valor conocido) y sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Considérese el estadístico muestral

$$
T = \sum_{i=1}^{n} X_i
$$

Entonces, para cada $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$
\begin{aligned}
L(x_1, \ldots, x_n; \mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2} \\
&= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2} e^{-\frac{1}{2\sigma^2} (n\mu^2 - 2\mu \sum_{i=1}^{n} x_i)} \\
&= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2} e^{-\frac{1}{2\sigma^2} (n\mu^2 - 2\mu t)} \\
&= h(x_1, \ldots, x_n) g(t; \mu),
\end{aligned}
$$

donde

$$
h(x_1, \ldots, x_n) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2} \qquad \text{y} \qquad g(t; \mu) = e^{-\frac{1}{2\sigma^2} (n\mu^2 - 2\mu t)}
$$

son funciones no negativas. Por tanto, T es un estadístico suficiente para μ.

Ejemplo 7. Sean $X \sim P(\lambda)$ y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Considérese el estadístico muestral

$$
T = \sum_{i=1}^{n} X_i
$$

Entonces, para cada $(x_1, \ldots, x_n) \in (\mathbb{N} \cup \{0\})^n$,

$$
L(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} = e^{-n\lambda} \frac{\lambda^t}{\prod_{i=1}^{n} x_i!} = h(x_1, \ldots, x_n) g(t, \lambda),
$$

donde

$$
h(x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^{n} x_i!} \qquad \text{y} \qquad g(t; \lambda) = e^{-n\lambda} \lambda^t
$$

son funciones no negativas. Por tanto, T es un estadístico suficiente para λ.

3.3. Estadísticos conjuntamente suficientes

Hasta ahora se ha trabajado con estadísticos que toman valores en $\mathbb{R}$ y con variables aleatorias cuya función de masa o densidad depende de un parámetro unidimensional $\theta \in \mathbb{R}$. El objetivo de esta sección es generalizar los estadísticos suficientes al caso multidimensional. Esto resultará útil en la práctica cuando, por ejemplo, se trabaje con distribuciones normales cuya media y varianza sean desconocidas.

Definición 20. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta \in \mathbb{R}^m$. Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X y sea $T\colon \mathbb{R}^n \to \mathbb{R}^k$ un estadístico muestral. Se dice que T es un* estadístico conjuntamente suficiente para *θ si para cada $(x_1,\dots,x_n) \in \chi^n$, el cociente*

$$\frac{f_\theta(x_1,\dots,x_n)}{g_\theta(T(x_1,\dots,x_n))}$$

es independiente de θ, donde g_θ es la función de masa o densidad del vector aleatorio k-dimensional $T(X_1,\dots,X_n)$.

Teorema 2. *Un estadístico $T\colon \mathbb{R}^n \to \mathbb{R}^k$ es conjuntamente suficiente para $\theta \in \Theta \subset \mathbb{R}^m$ si y solo si existen funciones $g\colon T(\chi^n)\times\Theta \to \mathbb{R}$, $h\colon \chi^n \to \mathbb{R}$ no negativas tales que para todo $(x_1,\dots,x_n)\in\chi^n$ se verifica*

$$L(x_1,\dots,x_n;\theta_1,\dots,\theta_m) = g(t_1,\dots,t_k;\theta_1,\dots,\theta_m)h(x_1,\dots,x_n),$$

donde $t_i = T_i(x_1,\dots,x_n)$ para cada $i \in \{1,\dots,k\}$.

Demostración. Análoga a la del teorema anterior. □

Ejemplo 8. Sea $X \sim N(\mu,\sigma^2)$, sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X y considérese el estadístico muestral (T_1,T_2), donde

$$T_1 = \sum_{i=1}^n X_i, \qquad T_2 = \sum_{i=1}^n X_i^2$$

Veamos que (T_1,T_2) es conjuntamente suficiente para (μ,σ^2). En efecto, si $(x_1,\dots,x_n)\in\mathbb{R}^n$,

$$\begin{aligned}
L(x_1,\dots,x_n;\mu,\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n}\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 + n\mu^2 - 2\mu\sum_{i=1}^n x_i\right)} \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n}\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\left(t_2 + n\mu^2 - 2\mu t_1\right)} \\
&= h(x_1,\dots,x_n)g(t_1,t_2;\mu,\sigma^2),
\end{aligned}$$

donde

$$h(x_1,\dots,x_n) = \frac{1}{\left(\sqrt{2\pi}\right)^n}, \qquad g(t_1,t_2;\mu,\sigma^2) = \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\left(t_2+n\mu^2-2\mu t_1\right)},$$

son funciones no negativas.

Proposición 21. *Sean X una variable aleatoria y $(X_1,\dots,X_n)$ una muestra aleatoria simple de X.*

(i) Si $T_i = X_i$, $i\in\{1,\dots,n\}$, entonces el estadístico $(T_1,\dots,T_n)$ es conjuntamente suficiente para θ.

(ii) Si $T_i = X_{(i)}$, $i\in\{1,\dots,n\}$, entonces el estadístico $(T_1,\dots,T_n)$ es conjuntamente suficiente para θ.

Demostración. Ejercicio. □

3.4. Familia exponencial uniparamétrica

Definición 21. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta \in \Theta \subset \mathbb{R}$. Se dice que f_θ* pertenece a una familia exponencial uniparamétrica *si existen funciones $a,c\colon \Theta \to \mathbb{R}$, $b,d\colon \chi \to \mathbb{R}$, con a y b son no negativas, tales que para cada $x\in\chi$ se verifica*

$$f_\theta(x) = a(\theta)b(x)e^{c(\theta)d(x)}$$

Proposición 22. *Sea X una variable aleatoria cuya función de masa o densidad f_θ pertenece a una familia exponencial uniparamétrica. Si $(X_1,\ldots,X_n)$ es una muestra aleatoria simple de X, entonces*

$$T=\sum_{i=1}^{n} d(X_i)$$

es un estadístico muestral suficiente para θ.

Demostración. Se tiene que

$$L(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n} f_\theta(x_i)=a(\theta)^n e^{c(\theta)\sum_{i=1}^{n} d(x_i)}\prod_{i=1}^{n} b(x_i)=a(\theta)^n e^{c(\theta)t}\prod_{i=1}^{n} b(x_i)=g(t;\theta)h(x_1,\ldots,x_n),$$

donde

$$h(x_1,\ldots,x_n)=\prod_{i=1}^{n} b(x_i), \qquad g(t;\theta)=a(\theta)^n e^{c(\theta)t}$$

son funciones no negativas. □

Ejemplo 9. Sea $X\sim P(\lambda)$, sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X y sea

$$T=\sum_{i=1}^{n} X_i$$

Dado $x\in\mathbb{N}\cup\{0\}$, se tiene que

$$P_\lambda(X=x)=e^{-\lambda}\frac{\lambda^x}{x!}=e^{-\lambda}\frac{e^{x\log(\lambda)}}{x!}=a(\lambda)b(x)e^{c(\lambda)d(x)},$$

donde

$$a(\lambda)=e^{-\lambda}, \qquad b(x)=\frac{1}{x!}, \qquad c(\lambda)=\log(\lambda), \qquad d(x)=x,$$

siendo $a,c\colon(0,\infty)\to\mathbb{R}$, $b,d\colon\mathbb{N}\cup\{0\}\to\mathbb{R}$ con a y b no negativas. Por la proposición anterior, T es un estadístico suficiente para θ.

3.5. Familia exponencial k-paramétrica

Definición 22. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta\in\Theta\subset\mathbb{R}^k$. Se dice que f_θ* pertenece a una familia exponencial k-paramétrica *si existen funciones $a,c_i\colon\Theta\to\mathbb{R}$, $b,d_i\colon\chi\to\mathbb{R}$, con a y b no negativas e $i\in\{1,\ldots,k\}$, tales que para cada $x\in\chi$ se verifica*

$$f_\theta(x)=a(\theta_1,\ldots,\theta_k)b(x)e^{\sum_{i=1}^{k} c_i(\theta_1,\ldots,\theta_k)d_i(x)}$$

Proposición 23. *Sea X una variable aleatoria cuya función de masa o densidad f_θ pertenece a una familia exponencial k-paramétrica. Si $(X_1,\ldots,X_n)$ es una muestra aleatoria simple de X, entonces*

$$(T_1,\ldots,T_k)=\left(\sum_{j=1}^{n} d_1(X_j),\ldots,\sum_{j=1}^{n} d_k(X_j)\right)$$

es un estadístico muestral conjuntamente suficiente para $\theta=(\theta_1,\ldots,\theta_k)$.

Demostración. Se tiene que

$$\begin{aligned}
L(x_1,\dots,x_n;\theta_1,\dots,\theta_k) &= \prod_{j=1}^{n} a(\theta_1,\dots,\theta_k)b(x_j)e^{\sum_{i=1}^{k} c_i(\theta_1,\dots,\theta_k)d_i(x_j)} \\
&= a(\theta_1,\dots,\theta_k)^n e^{\sum_{j=1}^{n}\sum_{i=1}^{k} c_i(\theta_1,\dots,\theta_k)d_i(x_j)} \prod_{j=1}^{n} b(x_j) \\
&= a(\theta_1,\dots,\theta_k)^n e^{\sum_{i=1}^{k} c_i(\theta_1,\dots,\theta_k)\sum_{j=1}^{n} d_i(x_j)} \prod_{j=1}^{n} b(x_j) \\
&= a(\theta_1,\dots,\theta_k)^n e^{\sum_{i=1}^{k} c_i(\theta_1,\dots,\theta_k)t_i} \prod_{j=1}^{n} b(x_j) \\
&= g(t_1,\dots,t_k;\theta_1,\dots,\theta_k)h(x_1,\dots,x_n),
\end{aligned}$$

donde

$$h(x_1,\dots,x_n) = \prod_{j=1}^{n} b(x_j), \qquad g(t_1,\dots,t_k;\theta_1,\dots,\theta_k) = a(\theta_1,\dots,\theta_k)^n e^{\sum_{i=1}^{k} c_i(\theta_1,\dots,\theta_k)t_i},$$

son funciones no negativas. □

Ejemplo 10. Sea $X \sim N(\mu,\sigma^2)$, sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X y considérese el estadístico

$$(T_1,T_2) = \left(\sum_{j=1}^{n} X_j^2, \sum_{j=1}^{n} X_j\right)$$

Se tiene que

$$f_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x^2+\mu^2-2x\mu)} = a(\mu,\sigma^2)b(x)e^{c_1(\mu,\sigma^2)d_1(x)+c_2(\mu,\sigma^2)d_2(x)},$$

donde

$$a(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\mu^2}{2\sigma^2}}, \qquad b(x)=1, \qquad c_1(x) = -\frac{1}{2\sigma^2}, \qquad c_2(x) = \frac{\mu}{\sigma^2}, \qquad d_1(x) = x^2, \qquad d_2(x) = x$$

En consecuencia, por la proposición anterior, (T_1,T_2) es un estadístico conjuntamente suficiente para (μ,σ^2).

3.6. Estadísticos suficientes minimales

Definición 23. *Un estadístico $T\colon \mathbb{R}^n \to \mathbb{R}$ suficiente para θ es un* estadístico suficiente minimal *si para cualquier estadístico muestral $T^*\colon \mathbb{R}^n \to \mathbb{R}$ suficiente para θ se verifica la siguiente propiedad: dados $(x_1,\dots,x_n),(y_1,\dots,y_n) \in \mathbb{R}^n$,*

$$T^*(x_1,\dots,x_n) = T^*(y_1,\dots,y_n) \implies T(x_1,\dots,x_n) = T(y_1,\dots,y_n)$$

Teorema 3. *Sea X una variable aleatoria con función de masa o densidad f_θ y sea $T\colon \mathbb{R}^n \to \mathbb{R}$ una función que verifica la siguiente propiedad: dados $(x_1,\dots,x_n),(y_1,\dots,y_n) \in \chi^n$,*

$$\frac{f_\theta(x_1,\dots,x_n)}{f_\theta(y_1,\dots,y_n)} \text{ no depende de } \theta \iff T(x_1,\dots,x_n) = T(y_1,\dots,y_n)$$

Entonces T es un estadístico suficiente minimal para θ.

Demostración. De nuevo, se probará solo en el caso discreto. Veamos primero que T es un estadístico suficiente para θ. Para cada $t \in T(\chi^n)$, se considera el conjunto

$$A_t = \{(y_1,\dots,y_n) \in \chi^n : T(y_1,\dots,y_n) = t\}$$

Entonces, si $(x_1,\dots,x_n)\in\chi^n$ y $t=T(x_1,\dots,x_n)$,

$$\frac{p_\theta(x_1,\dots,x_n)}{q_\theta(T(x_1,\dots,x_n))}=\frac{p_\theta(x_1,\dots,x_n)}{\sum_{(y_1,\dots,y_n)\in A_t}p_\theta(y_1,\dots,y_n)}=\frac{1}{\sum_{(y_1,\dots,y_n)\in A_t}\frac{p_\theta(y_1,\dots,y_n)}{p_\theta(x_1,\dots,x_n)}}$$

Como para todo $(y_1,\dots,y_n)\in A_t$ se tiene que $T(x_1,\dots,x_n)=T(y_1,\dots,y_n)$, entonces, por hipótesis, los cocientes

$$\frac{p_\theta(y_1,\dots,y_n)}{p_\theta(x_1,\dots,x_n)}$$

no dependen de θ, luego T es suficiente para θ. Veamos ahora que T es minimal. Considérese un estadístico T^* suficiente para θ. Entonces existen funciones g y h no negativas tales que para todo $(x_1,\dots,x_n)\in\chi^n$ se tiene

$$L(x_1,\dots,x_n;\theta)=g(T^*(x_1,\dots,x_n);\theta)h(x_1,\dots,x_n)$$

Por tanto, si $(x_1,\dots,x_n),(y_1,\dots,y_n)\in\chi^n$ son tales que $T^*(x_1,\dots,x_n)=T^*(y_1,\dots,y_n)$, entonces

$$\frac{p_\theta(x_1,\dots,x_n)}{p_\theta(y_1,\dots,y_n)}=\frac{L(x_1,\dots,x_n;\theta)}{L(y_1,\dots,y_n;\theta)}=\frac{g(T^*(x_1,\dots,x_n),\theta)h(x_1,\dots,x_n)}{g(T^*(y_1,\dots,y_n),\theta)h(y_1,\dots,y_n)}=\frac{h(x_1,\dots,x_n)}{h(y_1,\dots,y_n)}$$

Esto no depende de θ, luego $T(x_1,\dots,x_n)=T(y_1,\dots,y_n)$ y se concluye que T es minimal. □

Ejemplo 11. Sea $X\sim P(\lambda)$ y sean $(x_1,\dots,x_n),(y_1,\dots,y_n)\in\chi^n$. Entonces

$$\frac{p_\lambda(x_1,\dots,x_n)}{p_\lambda(y_1,\dots,y_n)}=\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\frac{\prod_{i=1}^n y_i!}{e^{-n\lambda}\lambda^{\sum_{i=1}^n y_i}}=\frac{\lambda^{\sum_{i=1}^n x_i}\prod_{i=1}^n y_i!}{\lambda^{\sum_{i=1}^n y_i}\prod_{i=1}^n x_i!}$$

Este cociente no depende de λ si y solo si $\sum_{i=1}^n x_i=\sum_{i=1}^n y_i$, luego, por el teorema anterior, $T=\sum_{i=1}^n X_i$ es un estadístico suficiente minimal para λ.

3.7. Información de Fisher

Definición 24. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde θ pertenece a un espacio paramétrico Θ. La* información de Fisher de X *es la función $\mathcal{I}_X:\Theta\to\mathbb{R}$ dada por*

$$\mathcal{I}_X(\theta)=E\left(\left(\frac{\partial\log(f(X,\theta))}{\partial\theta}\right)^2\right),$$

siempre que dicha expresión tenga sentido.

Evidentemente, para que la información de Fisher esté bien definida será necesario que $\chi\times\Theta$ sea un abierto de $\mathbb{R}^2$, que la función $(x,\theta)\mapsto\log(f(x,\theta))$, $(x,\theta)\in\chi\times\Theta$ tenga derivadas parciales con respecto a θ, que $\frac{\partial\log(f(X,\theta))}{\partial\theta}$ sea una variable aleatoria... y ese tipo de cosas. Ni que decir tiene que en esta asignatura todas estas rigurosiadades son completamente irrelevantes.

Proposición 24. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta\in\Theta$, y sea $\mathcal{I}_X$ la información de fischer de X. Entonces*

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

siempre que se disfrute de la regularidad necesaria.

Demostración. Solo se va a probar el caso absolutamente continuo. En primer lugar, si $(x,\theta)\in\chi\times\Theta$, por la regla de la cadena,

$$\frac{\partial\log(f(x,\theta))}{\partial\theta}=\frac{\frac{\partial f(x,\theta)}{\partial\theta}}{f(x,\theta)}$$

Derivando de nuevo,

$$\frac{\partial^2 \log(f(x,\theta))}{\partial\theta^2} = \frac{\frac{\partial^2 f(x,\theta)}{\partial\theta^2} f(x,\theta) - \left(\frac{\partial f(x,\theta)}{\partial\theta}\right)^2}{f(x,\theta)^2}$$

Así, por la linealidad de la esperanza,

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

Pero

$$E\left(\frac{\frac{\partial^2 f(X,\theta)}{\partial\theta^2}}{f(X,\theta)}\right) = \int_{\chi} \frac{\frac{\partial^2 f(x,\theta)}{\partial\theta^2}}{f(x,\theta)} f(x,\theta)\,dx = \int_{\chi} \frac{\partial^2 f(x,\theta)}{\partial\theta^2}\,dx = \frac{\partial^2}{\partial\theta^2}\int_{\chi} f(x,\theta)\,dx = \frac{\partial^2}{\partial\theta^2} 1 = 0,$$

deduciéndose inmediatamente la igualdad del enunciado. □

Proposición 25. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta \in \Theta$. Sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X. Entonces*

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = n\mathcal{I}_X(\theta)$$

Demostración. En efecto,

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = E\left(\left(\frac{\partial \log(f(X_1,\ldots,X_n,\theta))}{\partial\theta}\right)^2\right) = -E\left(\frac{\partial^2 \log(f(X_1,\ldots,X_n,\theta))}{\partial\theta^2}\right)$$

Pero

$$f(X_1,\ldots,X_n,\theta) = \prod_{i=1}^{n} f(X_i,\theta),$$

luego

$$\log(f(X_1,\ldots,X_n,\theta)) = \sum_{i=1}^{n} \log(f(X_i,\theta)),$$

y por tanto, derivando,

$$\frac{\partial \log(f(X_1,\ldots,X_n,\theta))}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \log(f(X_i,\theta))}{\partial\theta}$$

Derivando de nuevo,

$$\frac{\partial^2 \log(f(X_1,\ldots,X_n,\theta))}{\partial\theta^2} = \sum_{i=1}^{n} \frac{\partial^2 \log(f(X_i,\theta))}{\partial\theta^2}$$

En consecuencia, por la linealidad de la esperanza,

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = E\left(\sum_{i=1}^{n} \frac{\partial^2 \log(f(X_i,\theta))}{\partial\theta^2}\right) = \sum_{i=1}^{n} E\left(\frac{\partial^2 \log(f(X_i,\theta))}{\partial\theta^2}\right) = \sum_{i=1}^{n} \mathcal{I}_{X_i}(\theta) = \sum_{i=1}^{n} \mathcal{I}_X(\theta) = n\mathcal{I}_X(\theta),$$

donde se ha usado que todas las X_i poseen la misma distribución que X. □

Ejemplo 12. Sea $X \sim Exp(\lambda)$. Entonces, para $x > 0$,

$$f(x,\lambda) = \lambda e^{-\lambda x},$$

luego

$$\log(f(x,\lambda)) = \log(\lambda) - \lambda x,$$

y por tanto,

$$\frac{\partial \log(f(x,\lambda))}{\partial\lambda} = \frac{1}{\lambda} - x$$

Derivando otra vez,

$$\frac{\partial^2 \log(f(x,\lambda))}{\partial\lambda^2} = -\frac{1}{\lambda^2}$$

Consecuentemente,

$$\mathcal{I}_X(\lambda) = -E\left(-\frac{1}{\lambda^2}\right) = E\left(\frac{1}{\lambda^2}\right) = \int_0^\infty \frac{1}{\lambda^2} f(x,\lambda)\,dx = \frac{1}{\lambda^2}$$

Ejemplo 13. Sea $X \sim P(\lambda)$. Entonces, para $x \in \mathbb{N}\cup\{0\}$,

$$f(x,\lambda) = e^{-\lambda}\frac{\lambda^x}{x!},$$

luego

$$\log(f(x,\lambda)) = -\lambda + x\log(\lambda) - \log(x!),$$

y por tanto,

$$\frac{\partial \log(f(x,\lambda))}{\partial\lambda} = -1 + \frac{x}{\lambda}$$

Derivando otra vez,

$$\frac{\partial^2 \log(f(x,\lambda))}{\partial\lambda^2} = -\frac{x}{\lambda^2}$$

Consecuentemente,

$$\mathcal{I}_X(\lambda) = -E\left(-\frac{X}{\lambda^2}\right) = E\left(\frac{X}{\lambda^2}\right) = \frac{1}{\lambda^2}E(X) = \frac{1}{\lambda}$$

Ejemplo 14. Sea $X \sim P(\lambda)$ y sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X. Sea

$$T = \sum_{i=1}^n X_i,$$

que ya se sabe que es suficiente para λ. Tratemos de hallar $\mathcal{I}_T(\lambda)$. Por la propiedad reproductiva de la distribución de Poisson, $T \sim P(n\lambda)$, así que la función de masa de T es, para $t \in \mathbb{N}\cup\{0\}$,

$$f(t,\lambda) = e^{-n\lambda}\frac{(n\lambda)^t}{t!}$$

Razonando como en el ejemplo anterior, se obtiene

$$\mathcal{I}_T(\lambda) = \frac{1}{\lambda^2}E(T) = \frac{n}{\lambda}$$

Obsérvese que, por la proposición anterior,

$$\mathcal{I}_T(\lambda) = \mathcal{I}_{(X_1,\ldots,X_n)}(\lambda)$$

3.8. Estimadores

Definición 25. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde $\theta \in \Theta \subset \mathbb{R}^k$.*

(i) Un estadístico $T: \mathbb{R}^n \to \mathbb{R}^k$ se dice que es un estimador de θ *si $T(\mathbb{R}^n) \subset \Theta$.*

(ii) Dado un estimador T de θ, se define el error cuadrático medio de T *como*

$$ECM_T := E\left((T-\theta)^2\right),$$

donde se está denotando por T a la variable aleatoria $T(X_1,\ldots,X_n)$.

(iii) Dado un estimador T de θ, se define el sesgo de T *como*

$$b_T(\theta) := E(T) - \theta$$

(iv) Un estimador T de θ se dice que es insesgado *si su sesgo es nulo, esto es, si*

$$E(T) = \theta$$

(v) Dados dos estimadores T_1, T_2 de θ, se dice que T_1 es preferible frente a T_2 *si*

$$ECM_{T_1} \leq ECM_{T_2}$$

Proposición 26. *Si T es un estimador de θ, entonces*

$$ECM_T = \operatorname{Var}(T) + b_T(\theta)^2$$

Demostración. Efectivamente,

$$\begin{aligned}
ECM_T &= E\left((T-\theta)^2\right) \\
&= E\left((T - E(T) + E(T) - \theta)^2\right) \\
&= E\left((T - E(T)^2\right) + E\left((E(T)-\theta)^2\right) - 2E(T-E(T))E(E(T)-\theta) \\
&= E\left((T - E(T)^2\right) + E\left((E(T)-\theta)^2\right) \\
&= E\left((T - E(T)^2\right) + (E(T)-\theta)^2 \\
&= \operatorname{Var}(T) + b_T(\theta)^2,
\end{aligned}$$

donde en la tercera igualdad se ha usado que $T - E(T)$ y $E(T) - \theta$ son independientes (pues esta última variable aleatoria es una constante). □

Corolario 9. *Dados dos estimadores T_1, T_2 de θ insesgados, se tiene que T_1 es preferible frente a T_2 si y solo si* $\operatorname{Var}(T_1) \leq \operatorname{Var}(T_2)$*.*

Demostración. Trivial a partir de la proposición anterior. □

Ejemplo 15. Sea X una variable aleatoria con función de masa o densidad f_μ, donde $\mu = E(X)$. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X y sea

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

Se tiene que

$$E(T) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$

luego T es un estimador insesgado de μ. Hallemos su error cuadrático medio:

$$ECM_T = \operatorname{Var}(T) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{\sigma^2}{n},$$

donde $\sigma^2 = \operatorname{Var}(X)$.

Ejemplo 16. En las mismas circunstancias del ejemplo anterior, considérese el estimador

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

Si se recuerda la Proposición 2, llamando $\sigma^2 = \operatorname{Var}(X)$, se tiene

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

Por tanto, S^2 no es insesgado para σ^2.

Ejemplo 17. En las mismas circunstancias del ejemplo anterior, considérese el estimador

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

Si se recuerda la Proposición 3, llamando $\sigma^2 = \mathrm{Var}(X)$, se tiene

$$E(S_c^2) = \sigma^2$$

Por tanto, S_c^2 no es insesgado para μ, pero sí es insesgado para σ^2.

Definición 26. *Un estimador T de θ es un* estimador insesgado uniformemente de mínima varianza *para θ si $E(T) = \theta$ y $\mathrm{Var}(T) \leq \mathrm{Var}(T^*)$ para cualquier estimador insesgado T^* de θ con $E(T^*) = \theta$.*

Proposición 27 (Desigualdad de Cramér-Rao). *Dado un estimador T de θ, se tiene*

$$\mathrm{Var}(T) \geq \frac{(1 + b_T'(\theta))^2}{n\mathcal{I}_X(\theta)}$$

bajo ciertas condiciones de regularidad que no van a detallarse. En particular, si T es insesgado, entonces

$$\mathrm{Var}(T) \geq \frac{1}{n\mathcal{I}_X(\theta)}$$

Demostración. Hay que creérselo. □

Definición 27. *Un estimador insesgado T de θ se dice que es* eficiente *si alcanza la igualdad en la desigualdad de Cramér-Rao, esto es, si*

$$\mathrm{Var}(T) = \frac{1}{n\mathcal{I}_X(\theta)}$$

Corolario 10. *Todo estimador eficiente es insesgado uniformemente de mínima varianza.*

Demostración. Trivial a partir de la desigualdad de Cramér-Rao. □

Ejemplo 18. Sea $X \sim P(\lambda)$ y sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Veamos si $T = \overline{X}$ (que ya se sabe que es insesgado para λ) es insesgado uniformemente de mínima varianza para λ. Se tiene

$$E(T) = \lambda, \qquad \mathrm{Var}(T) = \frac{\lambda}{n}, \qquad \mathcal{I}_X(\lambda) = \frac{1}{\lambda},$$

luego

$$\mathrm{Var}(T) = \frac{1}{n\mathcal{I}_X(\lambda)}$$

Como T es un estimador eficiente, entonces es insesgado uniformemente de mínima varianza para λ.

Definición 28. *Se dice que una sucesión $\{T_n\}_{n\in\mathbb{N}}$ de estimadores de θ es* consistente *si*

$$\lim_{n\to\infty} ECM_{T_n} = 0$$

Ejemplo 19. Sea X una variable aleatoria con función de masa o función de densidad f_μ y $\mu = E(X)$. Sea $\sigma^2 = \mathrm{Var}(X)$, sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias independientes con la misma distribución que X y sea

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

una sucesión de estimadores de μ. Entonces

$$E(T_n) = \mu \xrightarrow{n\to\infty} \mu, \qquad \mathrm{Var}(T_n) = \frac{\sigma^2}{n} \xrightarrow{n\to\infty} 0,$$

de donde se deduce inmediatamente que $\{T_n\}_{n\in\mathbb{N}}$ es consistente.

3.9. Métodos de estimación del parámetro θ

En las subsecciones siguientes se exponen algunos métodos conocidos para estimar el parámetro θ del que depende la distribución de una variable aleatoria X.

3.9.1. Método de máxima verosimilitud

Definición 29. *Sea X una variable aleatoria con función de masa o densidad f_θ, donde θ pertenece a un espacio paramétrico Θ, y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. Si para algún $x \in \chi^n$ se tiene que $\hat{\theta}$ es un máximo global de la función $\theta \mapsto L(x,\theta)$, $\theta \in \Theta$, se dirá que $\hat{\theta}$ es un* estimador maximo-verosímil del parámetro θ.

Una condición necesaria pero no suficiente para que $\hat{\theta}$ sea un estimador maximo-verosímil de θ es que resuelva la ecuación

$$\frac{\partial L(x,\theta)}{\partial \theta} = 0,$$

o lo que es lo mismo,

$$\frac{\partial \log(L(x,\theta))}{\partial \theta} = 0,$$

donde se ha usado que

$$\frac{\partial \log(L(x,\theta))}{\partial \theta} = \frac{1}{L(x,\theta)} \frac{\partial L(x,\theta)}{\partial \theta} \qquad \text{y} \qquad L(x,\theta) > 0$$

En la práctica, la ecuación segunda suele ser más fácil de resolver que la primera. Una solución $\hat{\theta}$ de dicha ecuación es un candidato a máximo local de la función $\theta \mapsto L(x,\theta)$, $\theta \in \Theta$, pero podría no serlo. Ahora bien, si se verifica la condición

$$\left.\frac{\partial^2 \log(L(x,\theta))}{\partial^2 \theta}\right|_{\theta=\hat{\theta}} < 0,$$

entonces $\hat{\theta}$ sí que es un máximo local y, por tanto, basta comprobar que es un máximo global para que pueda afirmarse que $\hat{\theta}$ es un estimador maximo-verosímil de θ. Quedan al margen sutilezas del estilo: θ debe pertenecer al interior de Θ, las derivadas parciales con respecto a θ deben existir, la ecuación a resolver debe tener soluciones...

Ejemplo 20. Si $X \sim B(1,p)$, hallemos el estimador maximo-verosímil del parámetro p. Se tiene

$$L(x,p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, \quad (x,p) \in \{0,1\}^n \times (0,1)$$

Por tanto,

$$\log(L(x,p)) = \log(p)\sum_{i=1}^n x_i + \log(1-p)\left(n - \sum_{i=1}^n x_i\right)$$

En consecuencia,

$$\frac{\partial \log(L(x,p))}{\partial p} = 0 \iff \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p} = 0 \iff \sum_{i=1}^n x_i - p\cancel{\sum_{i=1}^n} x_i - np + p\cancel{\sum_{i=1}^n} x_i = 0 \iff p = \frac{\sum_{i=1}^n x_i}{n}$$

Por otro lado,

$$\frac{\partial^2 \log(L(x,p))}{\partial p^2} = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}$$

Sea

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

Entonces $\hat{p}$ es el estimador maximo-verosímil del parámetro p, ya que

$$\left.\frac{\partial^2 \log(L(x,p))}{\partial p^2}\right|_{p=\hat{p}} = -\frac{n\hat{p}}{\hat{p}^2} - \frac{n(1-\hat{p})}{(1-\hat{p})^2} = -\frac{n}{\hat{p}} - \frac{n}{1-\hat{p}} < 0$$

Ejemplo 21. Si $X \sim N(\mu,\sigma^2)$, hallemos el estimador maximo-verosímil del parámetro μ. Se tiene

$$L(x,\mu) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}, \quad (x,\mu) \in \mathbb{R}^n \times (0,\infty)$$

Por tanto,

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

Derivando e igualando a cero,

$$\frac{\partial \log(L(x,\mu))}{\partial \mu} = 0 \iff \frac{\sum_{i=1}^n (x_i-\mu)}{\sigma^2} = 0 \iff \sum_{i=1}^n x_i - n\mu = 0 \iff \mu = \frac{\sum_{i=1}^n x_i}{n}$$

Llamando $\hat{\mu}$ a la única solución de la ecuación, se tiene

$$\left.\frac{\partial^2 \log(L(x,\mu))}{\partial \mu^2}\right|_{\mu=\hat{\mu}} = -\frac{n}{\sigma^2} < 0,$$

así que $\hat{\mu}$ es el estimador maximo-verosímil del parámetro μ.

Definición 30. *Sea X una variable aleatoria cuya función de masa o densidad depende de un parámetro θ perteneciente a un espacio paramétrico Θ, sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X y sea $g\colon \Theta \to \mathbb{R}$ una función medible.*

(i) La función de verosimilitud inducida por *g es la función $M\colon \chi^n \times g(\Theta) \to \mathbb{R}$ dada por*

$$M(x,\lambda) = \sup_{\theta \in g^{-1}(\lambda)} L(x,\theta)$$

(ii) Si para algún $x \in \chi^n$ se tiene que $\hat{\lambda}$ es un máximo global de la función $\lambda \mapsto M(x,\lambda), \lambda \in g(\Theta)$, se dirá que $\hat{\lambda}$ es un estimador maximo-verosímil del parámetro *$\lambda = g(\theta)$.*

Teorema 4 (Teorema de invarianza de Zehna). *Si $g\colon \Theta \to \mathbb{R}$ es una función medible y $\hat{\theta}$ es un estimador maximo-verosímil de θ, entonces $g(\hat{\theta})$ es un estimador maximo-verosímil de $g(\theta)$.*

Demostración. Sea $\hat{x} \in \chi^n$ tal que $\hat{\theta}$ es máximo global de la función $\theta \mapsto L(\hat{x},\theta)$, y sea $\hat{\lambda} = g(\hat{\theta})$. Para todo $\lambda \in g(\Theta)$ se tiene

$$M(\hat{x},\lambda) = \sup_{\theta \in g^{-1}(\lambda)} L(\hat{x},\theta) \le \sup_{\theta\in\Theta} L(\hat{x},\theta) = L(\hat{x},\hat{\theta})$$

Por otra parte,

$$M(\hat{x},\hat{\lambda}) = \sup_{\theta \in g^{-1}(\hat{\lambda})} L(\hat{x},\theta) = L(\hat{x},\hat{\theta}),$$

donde se ha usado que $\hat{\theta} \in g^{-1}(\hat{\lambda})$. Por tanto, para todo $\lambda \in g(\Theta)$ se tiene

$$M(\hat{x},\lambda) \le M(\hat{x},\hat{\lambda}),$$

luego $\hat{\lambda}$ es un máximo global de la función $\theta \mapsto M(\hat{x},\lambda), \lambda \in g(\Theta)$, y concluimos que es un estimador de máxima verosimilitud del parámetro $\lambda = g(\theta)$. □

3.9.2. Método de los momentos

Sea X una variable aleatoria cuya función de masa o densidad depende de k parámetros $\theta_1,\dots,\theta_k$, y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. El método de los momentos consiste en estimar $(\theta_1,\dots,\theta_k)$ mediante la solución del sistema

$$\begin{cases} \alpha_1(\theta_1,\dots,\theta_k) = m_1 \\ \alpha_2(\theta_1,\dots,\theta_k) = m_2 \\ \qquad\vdots \\ \alpha_k(\theta_1,\dots,\theta_k) = m_k \end{cases}$$

Se recuerda que α_j es el *momento de orden j respecto al origen de X*,

$$\alpha_j = E(X^j),$$

y que m_j es el *momento muestral de orden j respecto al origen de* $(X_1,\dots,X_n)$,

$$m_j = \frac{1}{n}\sum_{i=1}^n X_i^j$$

Ejemplo 22. Sea $X \sim U([0,\theta])$ y estimemos θ mediante el método de los momentos. Se tiene que

$$\alpha_1 = E(X) = \int_0^\theta \frac{x}{\theta}\,dx = \frac{\theta}{2}, \qquad m_1 = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X},$$

así que la estimación obtenida por el método de los momentos es $\hat\theta = 2\overline{X}$.

Ejemplo 23. Sea $X \sim N(\mu,\sigma^2)$ y estimemos (μ,σ^2) mediante el método de los momentos. Se tiene que

$$\alpha_1 = E(X) = \mu, \qquad \alpha_2 = E(X^2) = \mathrm{Var}(X) + E(X)^2 = \sigma^2 + \mu^2, \qquad m_1 = \overline{X}, \qquad m_2 = \frac{1}{n}\sum_{i=1}^n X_i^2 = \overline{X^2}$$

así que hay que resolver el sistema

$$\begin{cases} \mu = \overline{X} \\ \sigma^2 + \mu^2 = \overline{X^2} \end{cases}$$

La solución del sistema es $(\hat\mu,\hat\sigma^2)$, donde

$$\hat\mu = \overline{X}, \qquad \hat\sigma^2 = \overline{X^2} - \overline{X}^2 = S^2$$

Ejemplo 24. Sea $X \sim U([a,a+b])$ y estimemos a y b mediante el método de los momentos. Se tiene que

$$\begin{aligned} \alpha_1 &= \int_a^{a+b} \frac{x}{b}\,dx = \frac{(a+b)^2}{2b} - \frac{a^2}{2b} = \frac{b}{2} + a \\ \alpha_2 &= \int_a^{a+b} \frac{x^2}{b}\,dx = \frac{(a+b)^3}{3b} - \frac{a^3}{3b} = \frac{3a^2b + 3ab^2 + b^3}{3b} = a^2 + ab + \frac{b^2}{3} \end{aligned}$$

El sistema a resolver sería

$$\begin{cases} \frac{b}{2} + a = \overline{X} \\ a^2 + ab + \frac{b^2}{3} = \overline{X^2} \end{cases}$$

De la primera ecuación se obtiene

$$a = \overline{X} - \frac{b}{2}$$

Sustituyendo en la segunda,

$$\begin{aligned}
\left(\overline{X}-\frac{b}{2}\right)^2+\left(\overline{X}-\frac{b}{2}\right)b+\frac{b^2}{3}=\overline{X^2} &\iff \overline{X}^2+\frac{b^2}{4}-\cancel{b\overline{X}}+\cancel{b\overline{X}}-\frac{b^2}{2}+\frac{b^2}{3}=\overline{X^2}\\
&\iff \frac{b^2}{12}=S^2\\
&\iff b=\pm 2\sqrt{3S^2}
\end{aligned}$$

El sistema posee dos soluciones: $(\hat{a}_1,\hat{b}_1)$ y $(\hat{a}_2,\hat{b}_2)$, donde

$$\hat{a}_1=\overline{X}-\sqrt{3S^2},\qquad \hat{a}_2=\overline{X}+\sqrt{3S^2},\qquad \hat{b}_1=2\sqrt{3S^2},\qquad \hat{b}_2=-2\sqrt{3S^2}$$

Intervalos de confianza

Al igual que en el tema anterior, el objetivo es estimar el parámetro θ desconocido del que depende la distribución de una variable aleatoria dada. En el tema anterior esta estimación se hacía dando valores concretos de aproximaciones de θ, mientras que en este tema se tratarán de dar intervalos en los que puede encontrarse θ. Otro aspecto en el que este tema se desmarcará del anterior es que nos centraremos única y exclusivamente en la aproximación de los parámetros μ y σ^2 de distribuciones normales.

4.1. Nociones básicas

Definición 31. *Sea X una variable aleatoria cuya distribución depende de un parámetro $\theta \in \mathbb{R}$ y sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X. Un* intervalo de estimación de θ *es un vector aleatorio $(L(X_1,\dots,X_n),U(X_1,\dots,X_n))$, donde $L,U : \chi^n \to \mathbb{R}$ son tales que*

$$L(x_1,\dots,x_n) \leq U(x_1,\dots,x_n)$$

para todo $(x_1,\dots,x_n) \in \chi^n$.

En analogía a la notación de los intervalos, un intervalo de estimación suele denotarse con corchetes en los extremos en lugar de paréntesis. Por otra parte, será habitual dar las funciones L y U directamente como variables aleatorias, es decir, como funciones de $(X_1,\dots,X_n)$ en lugar de $(x_1,\dots,x_n)$.

Definición 32. *Dado $\alpha \in (0,1)$, se dice que un intervalo de estimación $[L(X_1,\dots,X_n),U(X_1,\dots,X_n)]$ es un* intervalo de confianza para θ de nivel de confianza $1-\alpha$ *si*

$$P_\theta(\theta \in [L(X_1,\dots,X_n),U(X_1,\dots,X_n)]) = P_\theta(L(X_1,\dots,X_n) \leq \theta \leq U(X_1,\dots,X_n)) \geq 1-\alpha$$

Habitualmente se denota

$$IC(\theta,1-\alpha) = [L(X_1,\dots,X_n),U(X_1,\dots,X_n)]$$

Siempre que se dé un intervalo de confianza, es de extrema importancia comprobar que las funciones L y U no dependen de θ.

4.2. Intervalos de confianza para una distribución normal

En las subsecciones siguientes se expondrán intervalos de confianza para los parámetros μ y σ^2 de la distribución normal, dependiendo de cuáles de dichos parámetros son conocidos. Si, por ejemplo, μ es desconocida, entonces el intervalo de confianza que se dé para σ^2 se debe poder calcular sin utilizar μ.

Si $p \in (0,1)$ y X es una variable aleatoria, se va a denotar por X_p a cualquier número real que verifique

$$F(X_p) = P(X \leq X_p) = p$$

4.2.1. Parámetro μ de la distribución normal conociendo σ^2

Proposición 28. *Sea X una variable aleatoria con $X \sim N(\mu,\sigma^2)$, sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X y denotemos*

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

Entonces

$$IC(\mu,1-\alpha) = \left[\overline{X} - \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right]$$

Demostración. Se tiene que

$$\begin{aligned}
P\left(\overline{X} - \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2} \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right) &= P\left(-\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2} \le \overline{X}-\mu \le \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right)\\
&= P\left(-Z_{1-\alpha/2} \le Z \le Z_{1-\alpha/2}\right)\\
&= P(Z \le Z_{1-\alpha/2}) - P(Z \le Z_{\alpha/2})\\
&= 1-\alpha,
\end{aligned}$$

donde se ha usado que $Z_{\alpha/2} = -Z_{1-\alpha/2}$ porque $Z \sim N(0,1)$ y la distribución normal es simétrica con respecto al origen. □

4.2.2. Parámetro μ de la distribución normal desconociendo σ^2

Proposición 29. *Sea X una variable aleatoria con $X \sim N(\mu,\sigma^2)$, sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X y denotemos*

$$T = \frac{(\overline{X}-\mu)\sqrt{n}}{\sqrt{S_c^2}}$$

Entonces

$$IC(\mu,1-\alpha) = \left[\overline{X} - \frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}, \overline{X} + \frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\right]$$

y además este intervalo de confianza no depende de σ^2.

Demostración. Se tiene que

$$\begin{aligned}
P\left(\overline{X} - \frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2} \le \mu \le \overline{X} + \frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\right) &= P\left(-\frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2} \le \overline{X}-\mu \le \frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\right)\\
&= P\left(-T_{1-\alpha/2} \le T \le T_{1-\alpha/2}\right)\\
&= P(T \le T_{1-\alpha/2}) - P(T \le T_{\alpha/2})\\
&= 1-\alpha,
\end{aligned}$$

donde se ha usado que, por el lema de Fisher, $T \sim t_{n-1}$ y por tanto $T_{\alpha/2} = -T_{1-\alpha/2}$ por la simetría respecto al origen de la distribución T de Student. □

4.2.3. Parámetro σ^2 de la distribución normal conociendo μ

Proposición 30. *Sea X una variable aleatoria con $X \sim N(\mu,\sigma^2)$, sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X y denotemos*

$$Q = \sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$$

Entonces

$$IC(\sigma^2, 1-\alpha) = \left[\frac{\sigma^2 Q}{Q_{1-\alpha/2}}, \frac{\sigma^2 Q}{Q_{\alpha/2}}\right]$$

Demostración. Se tiene que

$$P\left(\frac{\sigma^2 Q}{Q_{1-\alpha/2}} \leq \sigma^2 \leq \frac{\sigma^2 Q}{Q_{\alpha/2}}\right) = P\left(\frac{1}{Q_{1-\alpha/2}} \leq \frac{1}{Q} \leq \frac{1}{Q_{\alpha/2}}\right) = P\left(Q_{\alpha/2} \leq Q \leq Q_{1-\alpha/2}\right) = 1-\alpha$$

Nótese que el intervalo de confianza dado tiene sentido porque, al ser $\alpha \in (0,1)$, se tiene $\frac{\alpha}{2} \leq 1 - \frac{\alpha}{2}$, y como las funciones de distribución son crecientes, entonces $Q_{\alpha/2} \leq Q_{1-\alpha/2}$ y $\frac{1}{Q_{1-\alpha/2}} \leq \frac{1}{Q_{\alpha/2}}$. Además, la distribución de variable aleatoria $\sigma^2 Q = \sum_{i=1}^n (X_i - \mu)^2$ no depende de σ^2, solo de μ, luego el intervalo de confianza tampoco depende de σ^2. □

La variable Q de la proposición anterior verifica $Q \sim \chi_n^2$ por el Corolario 6, así que, en la práctica, $Q_{1-\alpha/2}$ y $Q_{\alpha/2}$ se calculan fácilmente con ayuda de la tabla de probabilidades de la distribución χ_n^2.

4.2.4. Parámetro σ^2 de la distribución normal desconociendo μ

Proposición 31. *Sea X una variable aleatoria con $X \sim N(\mu, \sigma^2)$, sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de X y denotemos*

$$Q = \frac{(n-1)S_c^2}{\sigma^2}$$

Entonces

$$IC(\sigma^2, 1-\alpha) = \left[\frac{\sigma^2 Q}{Q_{1-\alpha/2}}, \frac{\sigma^2 Q}{Q_{\alpha/2}}\right]$$

y además este intervalo de confianza no depende de μ.

Demostración. Se tiene que

$$P\left(\frac{\sigma^2 Q}{Q_{1-\alpha/2}} \leq \sigma^2 \leq \frac{\sigma^2 Q}{Q_{\alpha/2}}\right) = P\left(\frac{1}{Q_{1-\alpha/2}} \leq \frac{1}{Q} \leq \frac{1}{Q_{\alpha/2}}\right) = P\left(Q_{\alpha/2} \leq Q \leq Q_{1-\alpha/2}\right) = 1-\alpha$$

Razonando como en la proposición anterior, el intervalo de confianza dado tiene sentido, y además no depende ni de σ^2 ni de μ. □

La variable Q de la proposición anterior verifica $Q \sim \chi_{n-1}^2$ por el lema de Fisher, así que, en la práctica, $Q_{1-\alpha/2}$ y $Q_{\alpha/2}$ se calculan fácilmente con ayuda de la tabla de probabilidades de la distribución χ_{n-1}^2.

4.3. Intervalos de confianza para dos distribuciones normales

Dadas dos distribuciones normales independientes, el objetivo es dar intervalos de confianza para la diferencia de sus medias y el cociente de sus varianzas.

4.3.1. Diferencia de medias de distribuciones normales

Proposición 32. *Sean X e Y variables aleatorias independientes con $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, y sean $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_n)$ muestras aleatorias simples de X e Y, respectivamente. Denotemos*

$$T = \frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sqrt{S_c^2}}$$

Entonces

$$IC(\mu_X - \mu_Y, 1-\alpha) = \left[\overline{X} - \overline{Y} - \frac{\sqrt{S_c^2}}{\sqrt{n}} T_{1-\alpha/2}, \ \overline{X} - \overline{Y} + \frac{\sqrt{S_c^2}}{\sqrt{n}} T_{1-\alpha/2}\right]$$

Demostración. Razonando como siempre, se verifica

$$P\left(\overline{X}-\overline{Y}-\frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\leq \mu_X-\mu_Y\leq \overline{X}-\overline{Y}+\frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\right)=1-\alpha,$$

aplicando el lema de Fisher a la muestra aleatoria simple $(X_1-Y_1,\dots,X_n-Y_n)$ de $X-Y$ para obtener $T\sim t_{n-1}$ y por tanto $T_{\alpha/2}=-T_{1-\alpha/2}$. □

Si las muestras aleatorias simples de las que se dispone no son del mismo tamaño, habrá que trabajar un poco más.

Proposición 33. *Sean X e Y variables aleatorias independientes con $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$, y sean $(X_1,\dots,X_n)$, $(Y_1,\dots,Y_m)$ muestras aleatorias simples de X e Y, respectivamente. Supongamos que $\sigma_X^2=\sigma_Y^2=\sigma^2$ y denotemos*

$$T=\frac{\dfrac{\overline{X}-\overline{Y}-\mu_X+\mu_Y}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{\dfrac{\frac{(n-1)S_{c,X}^2}{\sigma^2}+\frac{(m-1)S_{c,Y}^2}{\sigma^2}}{n+m-2}}},\qquad S=\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_{c,X}^2+(m-1)S_{c,Y}^2}{n+m-2}}$$

Entonces

$$IC(\mu_X-\mu_Y,1-\alpha)=\left[\overline{X}-\overline{Y}-T_{1-\alpha/2}S,\ \overline{X}-\overline{Y}+T_{1-\alpha/2}S\right]$$

Demostración. Razonando como siempre, se verifica

$$P\left(\overline{X}-\overline{Y}-T_{1-\alpha/2}S\leq \mu_X-\mu_Y\overline{X}-\overline{Y}+T_{1-\alpha/2}S\right)=1-\alpha,$$

aplicando el Corolario 8 para obtener $T\sim t_{n+m-2}$ y por tanto $T_{\alpha/2}=-T_{1-\alpha/2}$. □

Ejemplo 25. Sean $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$ variables aleatorias independientes de las que se conocen las muestras aleatorias simples

$$\begin{aligned}(X_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8)&=(86,87,56,93,84,93,75,79)\\(Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7,Y_8)&=(80,79,58,91,77,82,74,66)\end{aligned}$$

Entonces

$$(X_1-Y_1,X_2-Y_2,X_3-Y_3,X_4-Y_4,X_5-Y_5,X_6-Y_6,X_7-Y_7,X_8-Y_8)=(6,8,-2,2,7,11,1,13)$$

Se trata de dar un intervalo de confianza para $\mu_X-\mu_Y$ de nivel de confianza 0.99. Tal y como se ha razonado antes, este intervalo es de la forma

$$IC(\mu_X-\mu_Y,1-\alpha)=\left[\overline{X}-\overline{Y}-\frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2},\ \overline{X}+\overline{Y}-\frac{\sqrt{S_c^2}}{\sqrt{n}}T_{1-\alpha/2}\right]$$

Hallemos los datos necesarios:

(*i*) $\alpha=0.01$.

(*ii*) $\overline{X}=\frac{1}{8}(86+87+56+93+84+93+75+79)=81.625.$

(*iii*) $\overline{Y}=\frac{1}{8}(80+79+58+91+77+82+74+66)=75.875.$

(*iv*) $\overline{X}-\overline{Y}=5.75.$

(*v*) $\sqrt{S_c^2} = \sqrt{\frac{1}{7}\left(0.25^2 + 2.25^2 + 7.75^2 + 3.75^2 + 1.25^2 + 5.25^2 + 4.75^2 + 7.25^2\right)} \approx 5.11999.$

(*vi*) $T_{0.995} \approx 3.499$ (véase la tabla de la distribución t_{n-1} con $n = 8$).

(*vii*) $\frac{\sqrt{S_c^2}}{\sqrt{8}} T_{1-\alpha/2} \approx 6.33387.$

En consecuencia,

$$IC(\mu_X - \mu_Y, 0.99) = [-0.58387, 12.08387]$$

4.3.2. Cociente de varianzas de distribuciones normales

Proposición 34. *Sean X e Y variables aleatorias independientes con $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, y sean $(X_1, \ldots, X_n)$, $(Y_1, \ldots, Y_m)$ muestras aleatorias simples de X e Y, respectivamente. Denotemos*

$$F = \frac{S_{c,X}^2 \sigma_Y^2}{S_{c,Y}^2 \sigma_X^2}$$

Entonces

$$IC\left(\frac{\sigma_X^2}{\sigma_Y^2}, 1-\alpha\right) = \left[\frac{S_{c,X}^2}{S_{c,Y}^2}\frac{1}{F_{1-\alpha/2}}, \frac{S_{c,X}^2}{S_{c,Y}^2}\frac{1}{F_{\alpha/2}}\right]$$

Demostración. Razónese como en las últimas cinco proposiciones. □

Nótese que, por el Corolario 8, la variable aleatoria de la proposición anterior verifica $F \sim F_{n-1,m-1}$, así que, en la práctica, $F_{1-\alpha/2}$ y $F_{\alpha/2}$ se pueden encontrar fácilmente en las tablas de probabilidades de la distribución F de Snedecor.

TEMA 5

Contraste de hipótesis

5.1. Nociones básicas

Definición 33. *Dada una variable aleatoria cuya distribución depende de un parámetro $\theta \in \Theta \subset \mathbb{R}^k$, una* hipótesis sobre θ *no es más que una expresión del estilo*

$$H: \theta \in A$$

para cualquier $A \subset \mathbb{R}^n$.

Ejemplo 26. Si $X \sim N(\mu, \sigma^2)$, una hipótesis sobre μ sería $H: \mu = 27$.

Definición 34. *Considérese una variable aleatoria cuya distribución depende de un parámetro $\theta \in \Theta$, y sea $\Theta_0 \subset \Theta$. Dos hipótesis del estilo*

$$H_0: \theta \in \Theta_0, \qquad H_1: \theta \notin \Theta_0,$$

se denominan, respectivamente, hipótesis nula *e* hipótesis alternativa.

(i) Dado $\theta_0 \in \Theta$, se denomina contraste de hipótesis bilateral *a cualquier par de hipótesis de la forma*

$$H_0: \theta = \theta_0, \qquad H_1: \theta \neq \theta_0$$

(ii) Dado $\theta_0 \in \Theta$, se denomina contraste de hipótesis unilateral *a cualquier par de hipótesis de la forma*

$$H_0: \theta \geq \theta_0, \qquad H_1: \theta < \theta_0,$$

o de la forma

$$H_0: \theta \leq \theta_0, \qquad H_1: \theta > \theta_0$$

Si se tiene una hipótesis nula H_0 y una hipótesis alternativa H_1, se pueden tomar dos decisiones: considerar H_0 como verdadera y H_1 como falsa, o lo contrario, considerar H_0 como falsa y H_1 como verdadera. En otras palabras, se puede *aceptar* H_0 y *rechazar* H_1, o al revés, *rechazar* H_0 y *aceptar* H_1.

Definición 35. *Un criterio que establece cuándo se debe aceptar H_0 y cuándo se debe aceptar H_1 se denomina* test de hipótesis.

Ejemplo 27. Dada una variable aleatoria X y una muestra aleatoria simple $(X_1, \ldots, X_n)$ de X, un test de hipótesis sería

Aceptamos H_0 si $\overline{X} > 3$ y rechazamos H_0 si $\overline{X} \leq 3$

para cualquier par de hipótesis complementarias H_0 y H_1.

Definición 36. *Dada una variable aleatoria X y dado $S \subset \mathbb{R}$, si se tiene un test de hipótesis del estilo*

Aceptamos H_0 si $X \notin S$ y rechazamos H_0 si $X \in S$,

se dice que S es la región crítica, *mientras que S^c se conoce como* región de aceptación.

Definición 37. *Dado un test de hipótesis con región crítica S, se define el* nivel de significación del test *como*

$$\alpha := \sup_{\theta \in \Theta_0} P_\theta(X \in S)$$

En el caso particular del contraste de hipótesis bilateral, como $\Theta_0 = \{\theta_0\}$, entonces

$$\alpha = P_{\theta_0}(X \in S)$$

En este caso, es habitual escribir

$$\alpha \equiv P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta})$$

5.2. Contraste de hipótesis para una distribución normal

En toda esta sección, X será una variable aleatoria con $X \sim N(\mu, \sigma^2)$ y $(X_1, \dots, X_n)$ será una muestra aleatoria simple de X. Se van a exponer los tests de hipótesis más utilizados para hipótesis relacionadas con cada parámetro de la distribución normal, dependiendo de si se conocen o no dichos parámetros. Por ejemplo, si no se conoce μ, habrá que emplear un test de hipótesis que no dependa de μ. Cabe destacar que las demostraciones de los casos unilaterales van a ser todas omitidas.

5.2.1. Contraste bilateral sobre μ conociendo σ^2

Proposición 35. *Considérense las hipótesis*

$$H_0: \mu = \mu_0, \qquad H_1: \mu \neq \mu_0$$

Dado $\alpha \in (0,1)$, un test de hipótesis con nivel de significación α es

$$\text{Aceptamos } H_0 \text{ si } |Z_0| \le Z_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } |Z_0| > Z_{1-\alpha/2},$$

donde

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad \text{y} \quad Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Demostración. En efecto, si suponemos que H_0 es cierta, entonces $Z = Z_0 \sim N(0,1)$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P(-Z_{1-\alpha/2} \le Z_0 \le Z_{1-\alpha/2}) = P(Z_{\alpha/2} \le Z_0 \le Z_{1-\alpha/2}) = 1 - \alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1 - (1 - \alpha) = \alpha$$

□

Ejemplo 28. Sea $X \sim N(\mu, 49)$, sea $\alpha = 0.01$ y sean

$$H_0: \mu = 20, \qquad H_1: \mu \neq 20$$

Considérese la muestra aleatoria simple

$$(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}) = (14, 17, 15, 16, 19, 27, 16, 12, 17, 24)$$

Entonces

$$\overline{X} = 17.7, \qquad Z = \frac{17.7 - 20}{\frac{7}{\sqrt{10}}} \approx -1.04, \qquad Z_{0.995} \approx 2.575$$

Como $|Z| \le Z_{0.995}$, el test de hipótesis anterior dice que se debe aceptar H_0.

5.2.2. Contraste bilateral sobre μ desconociendo σ^2

Proposición 36. *Considérense las hipótesis*

$$H_0: \mu = \mu_0, \qquad H_1: \mu \neq \mu_0$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *que no depende de* σ^2 *es*

$$\text{Aceptamos } H_0 \text{ si } |T_0| \leq T_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } |T_0| > T_{1-\alpha/2},$$

donde

$$T_0 = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \mu)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

Demostración. En efecto, si suponemos que H_0 es cierta, entonces $T = T_0 \sim t_{n-1}$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P(-T_{1-\alpha/2} \leq T_0 \leq T_{1-\alpha/2}) = P(T_{\alpha/2} \leq T_0 \leq T_{1-\alpha/2}) = 1 - \alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1 - (1 - \alpha) = \alpha$$

□

Ejemplo 29. Sea $X \sim N(\mu, \sigma^2)$, sea $\alpha = 0.01$ y sean

$$H_0: \mu = 3, \qquad H_1: \mu \neq 3$$

Sea $(X_1, \ldots, X_{18})$ una muestra aleatoria simple de X con $\overline{X} = 3.911$ y $S_c^2 = 0.763$. Entonces

$$T = \frac{3.911 - 3}{\frac{\sqrt{0.763}}{\sqrt{18}}} \approx 18.247, \qquad T_{0.995} \approx 2.898$$

Como $|T| > T_{0.995}$, el test de hipótesis anterior dice que se debe rechazar H_0.

5.2.3. Contraste unilateral sobre μ conociendo σ^2

Proposición 37. *Considérense las hipótesis*

$$H_0: \mu \leq \mu_0, \qquad H_1: \mu > \mu_0$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *es*

$$\text{Aceptamos } H_0 \text{ si } Z_0 \leq Z_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } Z_0 > Z_{1-\alpha},$$

donde

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad \text{y} \qquad Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Proposición 38. *Considérense las hipótesis*

$$H_0: \mu \geq \mu_0, \qquad H_1: \mu < \mu_0$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *es*

$$\text{Aceptamos } H_0 \text{ si } Z_0 \geq -Z_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } Z_0 < -Z_{1-\alpha},$$

donde

$$Z_0 = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \qquad \text{y} \qquad Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

5.2.4. Contraste unilateral sobre μ desconociendo σ^2

Proposición 39. *Considérense las hipótesis*

$$H_0: \mu \leq \mu_0, \qquad H_1: \mu > \mu_0$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *que no depende de* σ^2 *es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \leq T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 > T_{1-\alpha},$$

donde

$$T_0 = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \mu)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

Proposición 40. *Considérense las hipótesis*

$$H_0: \mu \geq \mu_0, \qquad H_1: \mu < \mu_0$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *que no depende de* σ^2 *es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \geq -T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 < -T_{1-\alpha},$$

donde

$$T_0 = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \mu)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

5.2.5. Contraste bilateral sobre σ^2 conociendo μ

Proposición 41. *Considérense las hipótesis*

$$H_0: \sigma^2 = \sigma_0^2, \qquad H_1: \sigma^2 \neq \sigma_0^2$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *es*

$$\text{Aceptamos } H_0 \text{ si } Q_{\alpha/2} \leq Q_0 \leq Q_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } Q_0 < Q_{\alpha/2} \text{ o } Q_0 > Q_{1-\alpha/2},$$

donde

$$Q_0 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma_0^2} \qquad \text{y} \qquad Q = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Demostración. En efecto, si suponemos que H_0 es cierta, entonces $Q = Q_0$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P(Q_{\alpha/2} \leq Q_0 \leq Q_{1-\alpha/2}) = 1 - \alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1 - (1 - \alpha) = \alpha$$

□

Ejemplo 30. Sea $X \sim N(15, \sigma^2)$, sea $\alpha = 0.1$ y sean $H_0: \sigma^2 = 4$, $H_1: \sigma^2 \neq 4$. Considérese la muestra aleatoria simple

$$(X_1, X_2, X_3, X_4) = (12, 16, 9, 17)$$

Entonces

$$Q = \frac{(12-15)^2}{4} + \frac{(16-15)^2}{4} + \frac{(9-15)^2}{4} + \frac{(17-15)^2}{4} = 12.5, \qquad Q_{0.95} \approx 9.4877, \qquad Q_{0.05} \approx 0.7107$$

Como $Q > Q_{0.95}$, el test de hipótesis anterior dice que se debe rechazar H_0. La región de aceptación de dicho test de hipótesis sería $S^c \approx [0.7107, 9.4877]$.

5.2.6. Contraste bilateral sobre σ^2 desconociendo μ

Proposición 42. *Considérense las hipótesis*

$$H_0 : \sigma^2 = \sigma_0^2, \qquad H_1 : \sigma^2 \neq \sigma_0^2$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *que no depende de* μ *es*

$$\text{Aceptamos } H_0 \text{ si } Q_{\alpha/2} \leq Q_0 \leq Q_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } Q_0 < Q_{\alpha/2} \text{ o } Q_0 > Q_{1-\alpha/2},$$

donde

$$Q_0 = \frac{(n-1)S_c^2}{\sigma_0^2} \qquad \text{y} \qquad Q = \frac{(n-1)S_c^2}{\sigma^2} \sim \chi^2_{n-1}$$

Demostración. En efecto, si suponemos que H_0 es cierta, entonces $Q = Q_0$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P\left(Q_{\alpha/2} \leq Q_0 \leq Q_{1-\alpha/2}\right) = 1-\alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1-(1-\alpha) = \alpha$$

□

Ejemplo 31. Sea $X \sim N(\mu,\sigma^2)$, sea $\alpha = 0.2$, sean $H_0 : \sigma^2 = 4$, $H_1 : \sigma^2 \neq 4$, y consideremos la muestra aleatoria simple

$$(X_1, X_2, X_3, X_4, X_5, X_6, X_7) = (7.1, 5.3, 4.7, 8, 9.9, 3.4, 3.6)$$

Entonces

$$Q = \frac{(n-1)S_c^2}{\sigma_0^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{4} \approx 8.72, \qquad Q_{0.1} \approx 2.204, \qquad Q_{0.9} \approx 10.645$$

Como $Q_{0.1} \leq Q \leq Q_{0.9}$, el test de hipótesis anterior dice que se debe aceptar H_0. La región de aceptación sería $[2.204, 10.645]$.

5.2.7. Contraste unilateral sobre σ^2 conociendo μ

Proposición 43. *Considérense las hipótesis*

$$H_0 : \sigma^2 \leq \sigma_0^2, \qquad H_1 : \sigma^2 > \sigma_0^2$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *es*

$$\text{Aceptamos } H_0 \text{ si } Q_0 \leq Q_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } Q_0 > Q_{1-\alpha},$$

donde

$$Q_0 = \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma_0^2} \qquad \text{y} \qquad Q = \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

Proposición 44. *Considérense las hipótesis*

$$H_0 : \sigma^2 \geq \sigma_0^2, \qquad H_1 : \sigma^2 < \sigma_0^2$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *es*

$$\text{Aceptamos } H_0 \text{ si } Q_0 \geq Q_{\alpha} \text{ y rechazamos } H_0 \text{ si } Q_0 < Q_{\alpha},$$

donde

$$Q_0 = \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma_0^2} \qquad \text{y} \qquad Q = \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2_n$$

5.3. Contraste de hipótesis para dos distribuciones normales

Esta sección será parecida a la anterior pero, como el título de la sección sugiere, ahora tenemos dos variables aleatorias independientes X e Y con $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, y dos muestras aleatorias simples $(X_1, \ldots, X_n)$, $(Y_1, \ldots, Y_m)$ de X e Y, respectivamente. Se dirá que las muestras son *apareadas* si $n = m$. Por último, se supondrá en todos los casos que las varianzas son desconocidas.

5.3.1. Contraste bilateral sobre μ_X y μ_Y para muestras apareadas

Proposición 45. *Considérense las hipótesis*

$$H_0 \colon \mu_X = \mu_Y, \qquad H_1 \colon \mu_X \neq \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras apareadas es*

$$\text{Aceptamos } H_0 \text{ si } |T_0| \leq T_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } |T_0| > T_{1-\alpha/2},$$

donde

$$T_0 = \frac{(\overline{X} - \overline{Y})\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

Demostración. Por ser las muestras apareadas, estamos en un caso particular de la Proposición 37. □

5.3.2. Contraste unilateral sobre μ_X y μ_Y para muestras apareadas

Proposición 46. *Considérense las hipótesis*

$$H_0 \colon \mu_X \leq \mu_Y, \qquad H_1 \colon \mu_X > \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras apareadas es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \leq T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 > T_{1-\alpha},$$

donde

$$T_0 = \frac{(\overline{X} - \overline{Y})\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

Demostración. Por ser las muestras apareadas, estamos en un caso particular de la Proposición 40. □

Proposición 47. *Considérense las hipótesis*

$$H_0 \colon \mu_X \geq \mu_Y, \qquad H_1 \colon \mu_X < \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras apareadas es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \geq -T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 < -T_{1-\alpha},$$

donde

$$T_0 = \frac{(\overline{X} - \overline{Y})\sqrt{n}}{\sqrt{S_c^2}} \qquad \text{y} \qquad T = \frac{(\overline{X} - \overline{Y} - \mu_X + \mu_Y)\sqrt{n}}{\sqrt{S_c^2}} \sim t_{n-1}$$

Demostración. Por ser las muestras apareadas, estamos en un caso particular de la Proposición 41. □

5.3.3. Contraste bilateral sobre μ_X y μ_Y para muestras no apareadas

Proposición 48. *Supóngase que* $\sigma_X^2 = \sigma_Y^2$ *y considérense las hipótesis*

$$H_0\colon \mu_X = \mu_Y, \qquad H_1\colon \mu_X \neq \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras no apareadas es*

$$\text{Aceptamos } H_0 \text{ si } |T_0| \le T_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } |T_0| > T_{1-\alpha/2},$$

donde

$$T_0 = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}, \qquad T = \frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \qquad \text{y} \qquad S = \sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}}$$

Demostración. En primer lugar, nótese que $T \sim t_{n+m-2}$ por el Corolario 8, teniendo en cuenta que $\sigma_X^2 = \sigma_Y^2$. Si suponemos que H_0 es cierta, entonces $T = T_0 \sim t_{n+m-2}$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P(-T_{1-\alpha/2} \le T_0 \le T_{1-\alpha/2}) = P(T_{\alpha/2} \le T_0 \le T_{1-\alpha/2}) = 1 - \alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1 - (1 - \alpha) = \alpha$$

□

5.3.4. Contraste unilateral sobre μ_X y μ_Y para muestras no apareadas

Proposición 49. *Supóngase que* $\sigma_X^2 = \sigma_Y^2$ *y considérense las hipótesis*

$$H_0\colon \mu_X \le \mu_Y, \qquad H_1\colon \mu_X > \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras no apareadas es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \le T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 > T_{1-\alpha},$$

donde

$$T_0 = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}, \qquad T = \frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \qquad \text{y} \qquad S = \sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}}$$

Proposición 50. *Supóngase que* $\sigma_X^2 = \sigma_Y^2$ *y considérense las hipótesis*

$$H_0\colon \mu_X \ge \mu_Y, \qquad H_1\colon \mu_X < \mu_Y$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras no apareadas es*

$$\text{Aceptamos } H_0 \text{ si } T_0 \ge -T_{1-\alpha} \text{ y rechazamos } H_0 \text{ si } T_0 < -T_{1-\alpha},$$

donde

$$T_0 = \frac{\overline{X} - \overline{Y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}, \qquad T = \frac{\overline{X} - \overline{Y} - \mu_X + \mu_Y}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \qquad \text{y} \qquad S = \sqrt{\frac{(n-1)S_{c,X}^2 + (m-1)S_{c,Y}^2}{n+m-2}}$$

5.3.5. Contraste bilateral sobre σ_X^2 y σ_Y^2 para muestras no apareadas

Proposición 51. *Considérense las hipótesis*

$$H_0: \sigma_X^2 = \sigma_Y^2, \qquad H_1: \sigma_X^2 \neq \sigma_Y^2$$

Dado $\alpha \in (0,1)$*, un test de hipótesis con nivel de significación* α *en el caso de muestras no apareadas es*

$$\text{Aceptamos } H_0 \text{ si } F_{\alpha/2} \leq F_0 \leq F_{1-\alpha/2} \text{ y rechazamos } H_0 \text{ si } F_0 < F_{\alpha/2} \text{ o } F_0 > F_{1-\alpha/2},$$

donde

$$F_0 = \frac{S_{c,X}^2}{S_{c,Y}^2} \qquad \text{y} \qquad F = \frac{S_{c,X}^2 \sigma_Y^2}{S_{c,Y}^2 \sigma_X^2} \sim F_{n-1,m-1}$$

Demostración. En primer lugar, nótese que $F \sim F_{n-1,m-1}$ por el Corolario 8. Si suponemos que H_0 es cierta, entonces $F = F_0$, luego

$$P(\text{Aceptar } H_0 \,|\, H_0 \text{ es cierta}) = P(F_{\alpha/2} \leq F_0 \leq F_{1-\alpha/2}) = 1 - \alpha,$$

luego

$$P(\text{Rechazar } H_0 \,|\, H_0 \text{ es cierta}) = 1 - (1 - \alpha) = \alpha$$

□

Quedan al margen el resto de contrastes para σ_X^2 y σ_Y^2 (contrastes unilaterales y casos de muestras apareadas), pues se razonan de forma totalmente análoga a los que ya se han visto.

APÉNDICE A

Recuerdos del pasado

Como el título sugiere, este apéndice no es más que un recordatorio de la asignatura *Probabilidad*. En consecuencia, las demostraciones de todos los resultados van a ser omitidas.

A.1. Espacio de probabilidad y σ-álgebra de Borel

Definición 38. *Sea Ω un conjunto cualquiera.*

(i) *Se dice que un subconjunto $\mathcal{A} \subset \mathcal{P}(\Omega)$ es una σ-*álgebra sobre Ω *cuando*

(i.1) $\Omega \in \mathcal{A}$.

(i.2) *Para todo $A \in \mathcal{A}$ se tiene que $A^c \in \mathcal{A}$.*

(i.3) *Para toda familia numerable $\{A_i\}_{i\in I}$ de elementos de $\mathcal{A}$ se tiene que $\bigcup_{i\in I} A_i \in \mathcal{A}$.*

(ii) *Dada una σ-álgebra $\mathcal{A}$ sobre Ω, la dupla $(\Omega, \mathcal{A})$ se denomina* espacio medible.

(iii) *Si $\mathcal{A}$ es una σ-álgebra sobre Ω, una* medida de probabilidad sobre $\mathcal{A}$ *no es más que una función $P : \mathcal{A} \to [0,1]$ verificando*

(ii.1) $P(\Omega) = 1$.

(ii.2) *Si $\{A_n\}_{n\in\mathbb{N}}$ es una familia de elementos de $\mathcal{A}$ disjuntos dos a dos, entonces*

$$P\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} P(A_n)$$

(iv) *Dado un espacio medible $(\Omega, \mathcal{A})$ y una medida de probabilidad P sobre $\mathcal{A}$, la terna $(\Omega, \mathcal{A}, P)$ se denomina* espacio de probabilidad.

Definición 39. *Sea Ω un conjunto cualquiera.*

(i) *Dada una familia de subconjuntos $E \subset \mathcal{P}(\Omega)$, a la menor σ-álgebra sobre Ω que contiene a E se le denomina σ-*álgebra generada por *E, y se denota por $\sigma(E)$.*

(ii) *Sea τ una topología sobre Ω. La σ-*álgebra de Borel sobre Ω*, denotada por $\mathcal{B}(\Omega)$ o $\mathcal{B}_\Omega$, es la que genera τ, es decir, $\mathcal{B}(\Omega) := \sigma(\tau)$.*

Se suele denotar por $\mathcal{B}^n$ a la σ-álgebra de Borel sobre $\mathbb{R}^n$ junto con la topología usual. Gracias a la proposición que sigue, en $\mathbb{R}^n$, el estudio de los borelianos normalmente se reducirá al estudio de los intervalos.

Proposición 52. *La σ-álgebra de Borel $\mathcal{B}^n$ está generada por cada una de las siguientes familias de intervalos:*

(i) $\mathcal{E}_1 = \{(a,b) = (a_1,b_1) \times \ldots \times (a_n,b_n) : a,b \in \mathbb{R}^n\}$.

(ii) $\mathcal{E}_2 = \{[a,b) = [a_1,b_1) \times \ldots \times [a_n,b_n) : a,b \in \mathbb{R}^n\}$.

(iii) $\mathcal{E}_3 = \{(a,b] = (a_1,b_1] \times \ldots \times (a_n,b_n] : a,b \in \mathbb{R}^n\}$.

(iv) $\mathcal{E}_4 = \{[a,b] = [a_1,b_1] \times \ldots \times [a_n,b_n] \colon a,b \in \mathbb{R}^n\}$.

(v) $\mathcal{E}_5 = \{(a,+\infty) = (a_1,+\infty) \times \ldots \times (a_n,+\infty) \colon a \in \mathbb{R}^n\}$.

(vi) $\mathcal{E}_6 = \{[a,+\infty) = [a_1,+\infty) \times \ldots \times [a_n,+\infty) \colon a \in \mathbb{R}^n\}$.

(vii) $\mathcal{E}_7 = \{(-\infty,b) = (-\infty,b_1) \times \ldots \times (-\infty,b_n) \colon b \in \mathbb{R}^n\}$.

(viii) $\mathcal{E}_8 = \{(-\infty,b] = (-\infty,b_1] \times \ldots \times (-\infty,b_n] \colon b \in \mathbb{R}^n\}$.

A.2. Variables aleatorias

Definición 40. *Sean* $(X,\mathcal{M}_X)$ *e* $(Y,\mathcal{M}_Y)$ *dos espacios medibles. Una función* $f\colon X \to Y$ *se dice que es* medible con respecto a $\mathcal{M}_X$ y $\mathcal{M}_Y$ *si* $f^{-1}(B) \in \mathcal{M}_X$ *para todo* $B \in \mathcal{M}_Y$. *Si* (X,τ_X) *y* (Y,τ_Y) *son subespacios topológicos de* $\mathbb{R}^n$ *y* $\mathbb{R}^m$ *(respectivamente) junto con la topología usual, se dirá que* f *es* medible*, a secas, si es medible con respecto a* $\mathcal{B}_X$ *y* $\mathcal{B}_Y$.

Definición 41. *Sea* $(\Omega,\mathcal{A},P)$ *un espacio de probabilidad. Una* variable aleatoria sobre $(\Omega,\mathcal{A},P)$ *es una función* $X\colon \Omega \to \mathbb{R}$ *medible con respecto a* $\mathcal{A}$ *y* $\mathcal{B}$*, es decir, tal que* $X^{-1}(B) \in \mathcal{A}$ *para cada* $B \in \mathcal{B}$.

Definición 42. *Sea* X *una variable aleatoria en un espacio de probabilidad* $(\Omega,\mathcal{A},P)$.

(i) La función $P_X\colon \mathcal{B} \to [0,1]$ *dada por* $P_X(B) = P(X^{-1}(B))$ *se denomina* ley de probabilidad de X *o* distribución de probabilidad de X*, y suele denotarse* $P(X^{-1}(B)) \equiv P(X \in B)$.

(ii) La función $F_X\colon \mathbb{R} \to [0,1]$ *dada por* $F_X(x) = P_X((-\infty,x])$ *se llama* función de distribución de X.

(iii) La función $p\colon \mathbb{R} \to [0,1]$ *dada por* $p(x) = P_X(\{x\}) \equiv P(X = x)$ *se denomina* función de masa de X.

(iv) El conjunto $D_X := \{x \in \mathbb{R}\colon p(x) > 0\}$ *se denomina* rango de X.

Proposición 53. *Sea* F_X *la función de distribución de una variable aleatoria* X*. Entonces*

(i) F_X *es creciente.*

(ii) F_X *es continua por la derecha.*

(iii) $\lim\limits_{x\to-\infty} F_X(x) = 0$.

(iv) $\lim\limits_{x\to+\infty} F_X(x) = 1$.

(v) $P_X((-\infty,a)) = \lim\limits_{x\to a^-} F_X(x) \equiv F_X(a^-)$ *para todo* $a \in \mathbb{R}$.

(vi) $P_X((a,+\infty)) = 1 - F_X(a)$ *para todo* $a \in \mathbb{R}$.

(vii) $P_X([a,+\infty)) = 1 - F_X(a^-)$ *para todo* $a \in \mathbb{R}$.

(viii) $p(a) = F_X(a) - F_X(a^-)$ *para todo* $a \in \mathbb{R}$.

(ix) F_X *es continua en* a *si y solo si* $p(a) = 0$.

Proposición 54. *Sea* $\varphi\colon \mathbb{R} \to \mathbb{R}$ *una función medible y sea* $X\colon \Omega \to \mathbb{R}$ *una variable aleatoria sobre* $(\Omega,\mathcal{A},P)$*. Entonces* $\varphi \circ X$ *también es una variable aleatoria sobre* $(\Omega,\mathcal{A},P)$.

Definición 43. *Dadas* n *variables aleatorias* $X_1,\ldots,X_n$ *en un espacio de probabilidad* $(\Omega,\mathcal{A},P)$*, se dice que las variables aleatorias son* independientes *si para todos* $B_1,\ldots,B_n \in \mathcal{B}$ *se verifica*

$$P(X_1 \in B_1,\ldots,X_n \in B_n) = P(X_1 \in B_1)\ldots P(X_n \in B_n)$$

Proposición 55. *Si* X *e* Y *son variables aleatorias independientes sobre* $(\Omega,\mathcal{A},P)$ *y* $f,g\colon \mathbb{R} \to \mathbb{R}$ *son medibles, entonces* $f(X)$ *y* $g(Y)$ *son variables aleatorias independientes sobre* $(\Omega,\mathcal{A},P)$.

A.3. Vectores aleatorios

Definición 44. *Dado un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$*, un* vector aleatorio de dimensión *n es una n-tupla* $(X_1, \ldots, X_n)$*, donde X_i es una variable aleatoria sobre* $(\Omega, \mathcal{A}, P)$ *para cada* $i \in \{1, \ldots, n\}$.

Definición 45. *Sea* $X = (X_1, \ldots, X_n)$ *un vector aleatorio en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$*. A la función* $P_X : \mathcal{B}^n \to [0,1]$ *dada por* $P_X(B) = P(X^{-1}(B))$ *se le llama* ley de probabilidad de X *o* distribución de probabilidad de X*, y suele denotarse* $P(X^{-1}(B)) \equiv P(X \in B)$.

Definición 46. *Dados n vectores aleatorios $X_1, \ldots, X_n$ en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$*, se dice que los vectores aleatorios son* independientes *si para todos* $B_1, \ldots, B_n \in \mathcal{B}^n$ *se verifica*

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \ldots P(X_n \in B_n)$$

Proposición 56. *Si* $X = (X_1, \ldots, X_n)$ *e* $Y = (Y_1, \ldots, Y_m)$ *son vectores aleatorios independientes en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$ *y* $f : \mathbb{R}^n \to \mathbb{R}^l$, $g : \mathbb{R}^m \to \mathbb{R}^k$ *son continuas, entonces* $f(X)$ *y* $g(Y)$ *son vectores aleatorios independientes.*

Gran parte de la teoría relacionada con vectores aleatorios se desarrolla de forma totalmente análoga al caso unidimensional. Es por ello que de aquí en adelante se focalizará en el estudio de las variables aleatorias.

A.4. Clasificación de variables aleatorias

Definición 47. *Sea X una variable aleatoria en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$.

(i) Se dice que X es discreta *si existe un boreliano D_X finito o infinito numerable tal que* $P_X(D_X) = 1$.

(ii) Se dice que X es absolutamente continua *si existe una función* $f : \mathbb{R} \to [0, +\infty)$ *integrable y tal que*

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$$

para todo $a \in \mathbb{R}$*. La función f se denomina* función de densidad de X.

Nótese que la distribución de probabilidad de una variable aleatoria queda totalmente determinada en el momento en que se dé su función de masa, en el caso discreto, o su función de densidad, en el caso absolutamente continuo.

También cabe remarcar que esta clasificación no es exhaustiva: existen variables aleatorias que no son ni discretas ni absolutamente continuas, aunque no son objeto de estudio de esta asignatura.

Proposición 57. *Sea X una variable aleatoria en un espacio de probabilidad* $(\Omega, \mathcal{A}, P)$.

(i) Si X es una variable aleatoria discreta y $D_X = \{x_i\}_{i \in I}$*, con I un conjunto numerable, entonces*

$$\sum_{i \in I} p(x_i) = 1$$

(ii) Si X es una variable aleatoria absolutamente continua, entonces

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

Proposición 58 (Cambio de variable discreto). *Sea X una variable aleatoria discreta con función de masa p_X y sea* $g : \mathbb{R} \to \mathbb{R}$ *una función medible. Entonces* $Y = g \circ X$ *es una variable aleatoria discreta, y su función de masa viene dada por*

$$p_Y(y) = \sum_{x \in g^{-1}(\{y\})} p_X(x), \quad y \in \mathbb{R}$$

Proposición 59 (Cambio de variable absolutamente continuo). *Sea X una variable aleatoria absolutamente continua con función de densidad f_X y sea $\varphi: I \to \mathbb{R}$ una función de clase 1 y estrictamente monótona definida en un intervalo $I \subset \mathbb{R}$ tal que $X(\Omega) \subset I$. Entonces $Y = \varphi \circ X$ es una variable aleatoria absolutamente continua con función de densidad dada por*

$$f_Y(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|, \quad y \in \varphi(I)$$

A.5. Características numéricas

Definición 48. *Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.*

(i) Se define la esperanza de X *como*

$$E(X) := \begin{cases} \displaystyle\sum_{x \in D_X} x p(x) & \text{si } X \text{ es discreta} \\ \displaystyle\int_{-\infty}^{+\infty} x f(x) dx & \text{si } X \text{ es absolutamente continua} \end{cases}$$

(ii) Se define el momento de orden n respecto al origen de X *como*

$$\alpha_n := E(X^n) = \begin{cases} \displaystyle\sum_{x \in D_X} x^n p(x) & \text{si } X \text{ es discreta} \\ \displaystyle\int_{-\infty}^{+\infty} x^n f(x) dx & \text{si } X \text{ es absolutamente continua} \end{cases}$$

Si $n = 1$, suele denotarse $\alpha_1 = \mu$.

(iii) Se define la varianza de X *como*

$$\operatorname{Var}(X) := E(X^2) - E(X)^2 = \alpha_2 - \mu^2$$

(iv) Se define el momento de orden r respecto a la media de X *como*

$$\mu_r := E((X-\mu)^r) = \begin{cases} \displaystyle\sum_{x \in D_X} (x-\mu)^r p(x) & \text{si } X \text{ es discreta} \\ \displaystyle\int_{-\infty}^{+\infty} (x-\mu)^r f(x) dx & \text{si } X \text{ es absolutamente continua} \end{cases}$$

La posibilidad de que alguna suma o integral sea infinita no es asunto de esta asignatura. Será habitual suponer que las características numéricas con las que se trabaje sean finitas, aunque no se mencione.

Proposición 60. *Se verifican las siguientes propiedades:*

(i) Si X e Y son variables aleatorias, entonces

$$E(X+Y) = E(X) + E(Y)$$

(ii) Si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces

$$E(aX+b) = aE(X) + b$$

(iii) Si X e Y son variables aleatorias independientes, entonces

$$E(XY) = E(X)E(Y)$$

(iv) Si $g: I \to \mathbb{R}$ es continua y X es una variable aleatoria con $X(\Omega) \subset I$, entonces

$$E(g \circ X) = \begin{cases} \displaystyle\sum_{x \in D_X} g(x)p(x) & \text{si } X \text{ es discreta} \\ \displaystyle\int_{-\infty}^{+\infty} g(x)f(x)\,dx & \text{si } X \text{ es absolutamente continua} \end{cases}$$

(v) Si X es una variable aleatoria, entonces

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

para todos $a, b \in \mathbb{R}$.

(vi) Si $X_1, \ldots, X_n$ son independientes, entonces

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$$

Proposición 61 (Desigualdad de Chebyshev). *Dada una variable aleatoria X, para todo $\varepsilon > 0$ se verifica*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\mathrm{Var(X)}}{\varepsilon^2}$$

A.6. Función generatriz de momentos

Definición 49. *Sea X una variable aleatoria en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Se define la* función generatriz de momentos asociada a X *como*

$$M_X(t) := E(e^{tX}), \quad t \in \mathbb{R}$$

Proposición 62. *Se verifican las siguientes propiedades:*

(i) X e Y son variables aleatorias independientes si y solo si

$$M_{X+Y}(t) = M_X(t) M_Y(t), \quad t \in \mathbb{R}$$

(ii) Si $a, b \in \mathbb{R}$ y X es una variable aleatoria, entonces

$$M_{aX+b}(t) = e^{bt} M_X(at), \quad t \in \mathbb{R}$$

(iii) Si X es una variable aleatoria y $k \in \mathbb{N}$,

$$M_X^{(k)}(0) = E(X^k)$$

En particular,

$$M_X''(0) - \mu^2 = \sigma^2,$$

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

(iv) La función generatriz de momentos determina de forma única la distribución de una variable aleatoria. En otras palabras, si X e Y son variables aleatorias con $M_X(t) = M_Y(t)$ para todo $t \in \mathbb{R}$, entonces X e Y siguen la misma distribución de probabilidad.

Definición 50. *Dado un vector aleatorio $(X_1, \ldots, X_n)$, se define la* función de distribución de momentos conjunta asociada a $(X_1, \ldots, X_n)$ *como*

$$M_{(X_1,\ldots,X_n)}(t_1, \ldots, t_n) := E(e^{\sum_{i=1}^n t_i X_i}), \quad (t_1, \ldots, t_n) \in \mathbb{R}^n$$

Evidentemente, las propiedades vistas en la proposición anterior que sean susceptibles de generalizarse al caso multidimensional van a seguir siendo ciertas, pero, como ya se adelantó, nos centramos en el caso de una variable.

A.7. Convergencia de una sucesión de variables aleatorias

Definición 51. *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias en un mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$.*

(i) Se dice que $\{X_n\}_{n\in\mathbb{N}}$ converge en probabilidad a la variable aleatoria *X si*

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$$

para todo $\varepsilon > 0$. En ese caso, se denota $X_n \xrightarrow{P} X$.

(ii) Se dice que $\{X_n\}_{n\in\mathbb{N}}$ converge de forma casi segura a la variable aleatoria *X si*

$$P(\{\omega \in \Omega\colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1,$$

y, en ese caso, se denota $X_n \xrightarrow{\text{c.s.}} X$.

(iii) Se dice que $\{X_n\}_{n\in\mathbb{N}}$ converge en ley a la variable aleatoria *X si*

$$\lim_{n\to\infty} F_n(x_0) = F(x_0) \text{ para todo } x_0 \in \{x \in \mathbb{R}\colon F \text{ es continua en } x\},$$

donde F_i es la función de distribución de X_i y F la de X. En ese caso, se denota $X_n \xrightarrow{\mathcal{L}} X$.

Teorema 5 (Teorema de Kolmogorov). *Considérese una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ en un espacio de probabilidad independientes e idénticamente distribuidas. Si $E(X_i) = \mu < \infty$ para todo $i = 1, \dots, n$, entonces*

$$\overline{X}_n \xrightarrow{\text{c.s.}} \mu,$$

donde

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

Teorema 6 (Teorema de Lindeberg-Lévy). *Sea $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias en un mismo espacio de probabilidad independientes, idénticamente distribuidas y de forma que $E(X_i) = \mu < \infty$ y $\mathrm{Var}(X_i) = \sigma^2 < \infty$ para todo $i = 1, \dots, n$. Si $S_n = X_1 + \ldots + X_n$ para cada $n \in \mathbb{N}$, entonces*

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{L}} Z \sim N(0,1),$$

A.8. Distribuciones de probabilidad discretas

Para la definición siguiente, se recuerda que la distribución de probabilidad de una variable aleatoria discreta queda totalmente determinada por su función de masa.

Definición 52. *Sea X una variable aleatoria discreta en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.*

(i) Dado $x_0 \in \mathbb{R}$, se dice que X sigue una distribución degenerada en el punto *x_0, y se denota $X \sim D(x_0)$, cuando para cada $x \in \mathbb{R}$ se tiene*

$$P(X = x) = \begin{cases} 1 & si\ x = x_0 \\ 0 & si\ x \neq x_0 \end{cases}$$

(ii) Dados n puntos $\{x_1, \dots, x_n\}$ distintos, se dice que X sigue una distribución uniforme sobre los *n*

puntos, *y se denota* $X \sim U(\{x_1,\dots,x_n\})$*, cuando para cada* $x \in \mathbb{R}$ *se tiene*

$$P(X=x)=\begin{cases}\dfrac{1}{n} & si\ x\in\{x_1,\dots,x_n\}\\[2ex] 0 & si\ x\in\{x_1,\dots,x_n\}^c\end{cases}$$

(iii) *Dado* $p \in (0,1)$*, se dice que* X *sigue una* distribución de Bernoulli de parámetro p*, y se denota* $X \sim Ber(p)$*, cuando*

$$P(X=0)=1-p,\quad P(X=1)=p \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto}$$

Usualmente se escribe $q=1-p$*.*

(iv) *Dados* $n\in\mathbb{N}$ *y* $p\in(0,1)$*, se dice que* X *sigue una* distribución binomial de parámetros n y p*, y se denota* $X \sim Bin(n,p)$*, cuando*

$$P(X=k)=\binom{n}{k}p^k q^{n-k} \ \text{ para cada } k\in\mathbb{N}\cup\{0\} \text{ con } k\le n, \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto,}$$

donde, de nuevo, se denota $q=1-p$*.*

(v) *Dado* $\lambda>0$*, se dice que* X *sigue una* distribución de Poisson de parámetro λ*, y se denota* $X\sim P(\lambda)$*, cuando*

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!} \ \text{ para cada } k\in\mathbb{N}\cup\{0\}, \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto}$$

(vi) *Dado* $p\in(0,1)$*, se dice que* X *sigue una* distribución geométrica de parámetro p*, y se denota* $X\sim Geo(p)$*, cuando*

$$P(X=n)=q^n p \ \text{ para cada } n\in\mathbb{N}\cup\{0\}, \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto}$$

(vii) *Dado* $N\in\mathbb{N}$ *y dados* $n,D\in\mathbb{N}$ *tales que* $n\le N$ *y* $D\le N$*, se dice que* X *sigue una* distribución hipergeométrica de parámetros N, D y n*, y se denota* $X\sim H(N,D,n)$*, cuando*

$$P(X=k)=\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}) \ \text{ para cada } k\in\mathbb{N}\cup\{0\}, \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto}$$

(viii) *Si* $n\in N$ *y* $p\in(0,1)$*, se dice que* X *sigue una* distribución binomial negativa de parámetros n y p*, y se denota* $X\sim BN(n,p)$*, cuando*

$$P(X=k)=\binom{k+n-1}{k}q^k p^n \ \text{ para cada } k\in\mathbb{N}\cup\{0\}, \quad \text{y} \quad P(X=x)=0 \ \text{ en el resto}$$

Ejemplo 32. Dada una variable aleatoria discreta, a cada distribución de probabilidad se le asocia de manera natural un experimento aleatorio con diversas características:

(*i*) Una distribución degenerada se asocia a un experimento aleatorio con solo un posible resultado, como lanzar un dado de una cara.

(*ii*) Una distribución uniforme se asocia a un experimento aleatorio con varios resultados de la misma probabilidad, como tirar un dado de seis caras.

(*iii*) Una distribución de Bernoulli de parámetro $p\in(0,1)$ se asocia a un experimento aleatorio con dos posibles resultados. Los valores que toma la variable aleatoria son 0 (normalmente denominado *fracaso*, de probabilidad $q=1-p$) o 1 (*éxito*, de probabilidad p).

(*iv*) Una distribución binomial de parámetros $n\in\mathbb{N}$ y $p\in(0,1)$ se asocia a un experimento consistente en realizar n pruebas de Bernoulli, todas con probabilidad de éxito p. Los valores que toma la variable aleatoria son el número de éxitos (o de fracasos).

(v) Una distribución de Poisson de parámetro $\lambda > 0$ se suele asociar a un experimento aleatorio en el que se cuenta el número de ocurrencias de un fenómeno aleatorio de frecuencia de ocurrencia media λ durante un periodo de tiempo fijo o en una región fija del espacio. Los valores que toma la variable aleatoria son el número de ocurrencias por unidad de tiempo o espacio.

(vi) Una distribución geométrica de parámetro $p \in (0,1)$ está asociada a un experimento aleatorio consistente en realizar pruebas de Bernoulli hasta obtener un éxito, siendo todos los éxitos de probabilidad p. Los valores que toma la variable aleatoria son el número de fracasos antes del primer éxito.

(vii) Una distribución hipergeométrica de parámetros N, D y n se asocia a un experimento aleatorio en el que se dispone de una caja con N piezas, de las cuales D son defectuosas. Los valores que toma la variable aleatoria son el número de piezas defectuosas extraídas en una muestra de tamaño n.

($viii$) Una distribución binomial negativa de parámetros $n \in \mathbb{N}$ y $p \in (0,1)$ se asocia a un experimento consistente en realizar n pruebas de Bernoulli, todas con probabilidad de éxito p. Los valores que toma la variable aleatoria son el número de fracasos antes del n-ésimo éxito. Así, el suceso $X = k$ consiste en obtener k fracasos en las $n+k-1$ primeras pruebas, y un éxito en la prueba número $n+k$.

Proposición 63. *Se verifican las siguientes propiedades:*

(i) Dadas n variables aleatorias independientes $X_1,\dots,X_n$ con $X_i \sim Ber(p)$ para todo $i \in \{1,\dots,n\}$, se tiene que

$$\sum_{i=1}^{n} X_i \sim Bin(n,p)$$

(ii) Dadas n variables aleatorias independientes $X_1,\dots,X_n$ con $X_i \sim P(\lambda_i)$ para todo $i \in \{1,\dots,n\}$, se tiene que

$$\sum_{i=1}^{n} X_i \sim P\big(\sum_{i=1}^{n} \lambda_i\big)$$

A.9. Distribuciones de probabilidad absolutamente continuas

Al igual que en la sección anterior, se recuerda que basta dar la función de densidad de una variable aleatoria absolutamente continua para que quede determinada su distribución de probabilidad.

Definición 53. *Sea X una variable aleatoria absolutamente continua con función de densidad f en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$.*

(i) Dado un intervalo $[a,b] \subset \mathbb{R}$, se dice que X sigue una distribución uniforme en $[a,b]$, *y se denota $X \sim U([a,b])$, cuando*

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

(ii) Dados $\mu \in \mathbb{R}$ y $\sigma^2 > 0$, se dice que X sigue una distribución normal de parámetros μ y σ^2, *y se denota $X \sim N(\mu, \sigma^2)$, cuando*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{para cada } x \in \mathbb{R}$$

Cuando $\mu = 0$ y $\sigma^2 = 1$, se habla de distribución normal estándar.

(iii) *Si $\lambda > 0$, se dice que X sigue una* distribución exponencial de parámetro λ, *y se denota $X \sim Exp(\lambda)$, cuando*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

(iv) *Dados $\alpha > 0$ y $\beta > 0$, se dice que X sigue una* distribución gamma de parámetros α y β, *y se denota $X \sim Ga(\alpha, \beta)$, cuando*

$$f(x) = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{si } x > 0 \\ 0 & \text{en otro caso,} \end{cases}$$

donde

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x}\, dx$$

es la denominada función gamma.

(v) *Dados $\alpha > 0$ y $\beta > 0$, se dice que X sigue una* distribución beta de parámetros α y β, *y se denota $X \sim Be(\alpha, \beta)$, cuando*

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{si } x \in (0,1) \\ 0 & \text{en otro caso} \end{cases}$$