Capítulo 4

Modelo de muestreo aleatorio

En este capítulo presentamos la modelación probabilista del *muestreo aleatorio*, y estudiamos las propiedades de los *estadísticos* más relevantes asociados a una muestra aleatoria: la media muestral y la (cuasi)varianza muestral, junto con un breve análisis del máximo y el mínimo.

El muestreo aleatorio de variables normales será particularmente relevante: el resultado principal al respecto es el teorema 4.6 de Fisher–Cochran.

4.1. Modelo de muestreo aleatorio y estadísticos

Decimos que la variable aleatoria Y es un **clon** de la variable aleatoria X si X e Y tienen la misma distribución; esto es, si para cualquier boreliano A de $\mathbb{R}$,

$$\mathbf{P}(X \in A) = \mathbf{P}(Y \in A).$$

Por supuesto, si Y es un clon de X, entonces X es un clon de Y.

En particular, si Y es un clon de X, las variables X e Y tienen la misma función de distribución:

$$F_X \equiv F_Y\,.$$

Escribiremos habitualmente que $X \stackrel{\mathrm{d}}{=} Y$.

Por **muestra aleatoria** de tamaño n de la variable X entendemos un vector $(X_1, \dots, X_n)$ en el que cada X_j es un clon de X y además las $X_1, \dots, X_n$ son (completamente) independientes. Denotamos con $\mathbb{X}$ al vector (aleatorio) columna

$$\mathbb{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

Éste es nuestro modelo para describir procesos de sorteos (independientes) y de extracción de muestras. El vector $\mathbb{X}$ contiene *todas* las posibles muestras $(x_1, \dots, x_n)$ (de X y de tamaño n), cada una con sus "probabilidades". Más concretamente,

- si X es discreta, la probabilidad de que se obtenga una determinada muestra $(x_1, \dots, x_n) \in \mathbb{R}^n$ es

  $$\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbf{P}(X = x_i),$$

  tal y como prescribe la función de masa conjunta de $\mathbb{X}$;
- mientras que si X es una variable continua, entonces los cálculos se realizan con la función de densidad conjunta de $\mathbb{X}$, que es

  $$f_{\mathbb{X}}(x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i), \quad \text{para cualquier } (x_1, \dots, x_n) \in \mathbb{R}^n,$$

  donde f_X es la función de densidad de la variable X de referencia.

Como ejemplo, digamos que X sigue una distribución geométrica de parámetro p, $X \sim \text{GEOM}(p)$; esto es, X toma los valores $1, 2, \dots$ con probabilidades $\mathbf{P}(X = j) = p(1-p)^{j-1}$. Si $n = 3$, entonces las posibles muestras son triples de enteros positivos (i, j, k), y la probabilidad de obtener una determinada muestra (i, j, k) es $p^3(1-p)^{i+j+k-3}$.

4.1.1. Estadísticos

Para los propósitos de este curso, interesará considerar (y analizar) variables aleatorias que se escriben como *funciones* de las variables de la muestra $\mathbb{X}$.

Cualquier función (medible) H definida en $\mathbb{R}^n$ y con valores en $\mathbb{R}$ aplicada a $(X_1, X_2, \dots, X_n)$ define un **estadístico** T,

$$T = H(X_1, X_2, \dots, X_n)\,,$$

de la variable X. Todo estadístico T es asimismo una variable aleatoria.

Interesará disponer de cuanta información sobre cada estadístico T seamos capaces de obtener: medias y varianzas, o incluso la propia distribución de probabilidad, lo que no resultará sencillo en cuanto H sea una función moderadamente complicada, como ya adelantamos en el apartado 2.2. Siguiendo el ejemplo de antes, supongamos que $X \sim \textsc{geom}(p)$, que $n = 3$, y que el estadístico de interés es $T = X_1 + X_2 + X_3$, la suma de los valores de la muestra. La probabilidad de que T tome el valor, digamos, 7, exige contar cuántos triples (i, j, k) de enteros positivos son tales que $i + j + k = 7$ (hay $\binom{6}{2}$ de ellos), para luego multiplicar esta cantidad por $p^3(1-p)^4$. Véase el ejemplo 4.2.3.

En lo que sigue, el tamaño de la muestra aleatoria será siempre un genérico n. No reflejaremos ese n en el símbolo que represente al estadístico (y por tanto escribiremos, por ejemplo, T en lugar de T_n). Más adelante, por ejemplo en el apartado 5.4, analizaremos el comportamiento de estadísticos cuando el tamaño de la muestra crece, y será necesario precisar algo más la notación.

Los estadísticos más habituales, y a los que dedicaremos más atención, son:

- la **media muestral**, que se denota por $\overline{X}$:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{j=1}^{n} X_j\,;$$

- y la **cuasivarianza muestral**, que se denota por S^2 (y requiere $n \geq 2$):

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\big(X_j - \overline{X}\big)^2\,.$$

Nótese que en S^2 se divide por $n-1$, y no por el tamaño de la muestra n. La variable S^2 es una función de las X_i más complicada que la $\overline{X}$. Concretamente, desarrollando cuadrados,

$$S^2 = \frac{1}{n-1}\Big[\sum_{j=1}^{n} X_j^2 - n\,\overline{X}^2\Big] = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - \frac{1}{n(n-1)}\sum_{i\neq j} X_i X_j$$

(véase el detalle de este cálculo en el apartado 4.3).

Ocasionalmente, trataremos otros estadísticos como la **cuasidesviación típica muestral**, que se define como $S = \sqrt{S^2}$; la **varianza muestral**,

$$D^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$$

(obsérvese que ahora se divide por n, y no por $n-1$), y la **desviación típica muestral**, dada por $\sqrt{D^2}$.

También nos interesarán

- el **máximo** de la muestra, $\text{máx}\,(X_1, X_2, \dots, X_n)$
- y el **mínimo** de la muestra, $\text{mín}\,(X_1, X_2, \dots, X_n)$,

además de otros estadísticos relacionados con la ordenación de (los valores de) la muestra, como el **rango**: $\text{máx}\,(X_1, X_2, \dots, X_n) - \text{mín}\,(X_1, X_2, \dots, X_n)$, y los **estadísticos de orden** $X_{(r:n)}$, definidos, para cada entero $1 \le r \le n$, como sigue: se ordena la muestra $(X_1, X_2, \dots, X_n)$

$$X_{(1:n)} \le X_{(2:n)} \le \dots \le X_{(n:n)}\,,$$

y se toma el valor que ocupa la posición r-ésima en esta ordenación de menor a mayor. Por ejemplo, $X_{(1:n)}$ es el mínimo y $X_{(n:n)}$ es el máximo; $X_{(2:n)}$ sería el segundo más pequeño.

Importa recalcar que:

- los estadísticos son variables aleatorias;
- en el cálculo de cualquier estadístico interviene tan sólo la muestra aleatoria $(X_1, \dots, X_n)$ de X, y no interviene ningún parámetro de la distribución específica a la que se aplica.

En este capítulo no centraremos en el análisis de la media muestral $\overline{X}$ y de la cuasivarianza muestral S^2, con una breve excursión adicional por el mundo de los máximos y mínimos muestrales.

4.2. Sobre la media muestral

Partimos de una muestra aleatoria $\mathbb{X} = (X_1, \dots, X_n)^\intercal$ de tamaño n de una variable X. La variable media muestral viene dada por

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Este estadístico recoge todas las posibles medias muestrales del experimento que consiste en muestrear n veces la variable X.

4.2.1. Media y varianza de $\overline{X}$

Para cualquier variable X, la media y la varianza de la variable aleatoria $\overline{X}$ se expresan directamente en términos de la media y la varianza de la variable X.

Proposición 4.1 *Para cualquier variable aleatoria con* $\mathbf{E}(X^2) < +\infty$ *se tiene que*

$$\mathbf{E}(\overline{X}) = \mathbf{E}(X) \qquad y \qquad \mathbf{V}(\overline{X}) = \frac{1}{n}\,\mathbf{V}(X).$$

Demostración. Por un lado,

$$\mathbf{E}(\overline{X}) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}(X_j) = \frac{1}{n}\,n\,\mathbf{E}(X) = \mathbf{E}(X)\,;$$

y, por otro, usando la independencia,

$$\mathbf{V}(\overline{X}) = \frac{1}{n^2}\sum_{j=1}^{n}\mathbf{V}(X_j) = \frac{1}{n^2}\,n\mathbf{V}(X) = \frac{1}{n}\,\mathbf{V}(X).$$

(Aquí, por cierto, habría bastado con que las variables X_j fueran incorreladas). ■

4.2.2. Concentración de $\overline{X}$ alrededor de $\mathbf{E}(X)$

Si X es tal que $\mathbf{E}(X^2) < +\infty$, la desigualdad de Chebyshev (teorema 2.2), aplicada a la media muestral $\overline{X}$, nos dice que

$$\mathbf{P}\big(|\overline{X} - \mathbf{E}(X)| \ge \varepsilon\big) \le \frac{\mathbf{V}(X)}{n\varepsilon^2}\,, \tag{4.1}$$

es decir, que los valores de $\overline{X}$ se concentran cada vez más en torno a $\mathbf{E}(X)$ cuanto mayor sea n. La desigualdad anterior cuantifica cuán probable es que $\overline{X}$ se aleje del valor de la media $\mathbf{E}(X)$ de X.

Una escritura alternativa de (4.1) es la siguiente:

$$\mathbf{P}\Big(|\overline{X} - \mathbf{E}(X)| \ge \varepsilon\sqrt{\mathbf{V}(X)}\Big) \le \frac{1}{n\varepsilon^2}\,, \tag{4.2}$$

en la que registramos la magnitud de las desviaciones en torno a la media en la escala “natural”, la de la desviación típica.

De (4.1) se deduce la ley débil de los grandes números (para variables tales que $\mathbf{E}(X^2) < +\infty$): para todo $\varepsilon > 0$ se tiene que

$$\lim_{n\to\infty}\mathbf{P}\big(|\overline{X} - \mathbf{E}(X)| \ge \varepsilon\big) = 0\,.$$

4.2.3. Distribución (exacta y/o aproximada) de $\overline{X}$

En ciertas ocasiones, para determinados tipos de variables X, es posible obtener la distribución completa de la variable $\overline{X}$, lo que permite responder de manera exacta a cuestiones sobre concentración de los valores de $\overline{X}$.

Ejemplo 4.2.1. *Bernoulli y binomial.*

Si X es una $\text{BER}(p)$, entonces $n\overline{X} = X_1 + \cdots + X_n$ es $\text{BIN}(n, p)$. Es decir,

$$\mathbf{P}(n\overline{X} = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{para cada } k = 0, 1, \ldots, n.$$

Por lo tanto, $\overline{X}$ toma los valores $0, 1/n, 2/n, \ldots, 1$ con las mismas probabilidades con las que $n\overline{X}$ toma los valores $0, 1, \ldots, n$.

Más generalmente, si X es una binomial $\text{BIN}(m, p)$, entonces $n\overline{X} \sim \text{BIN}(nm, p)$, esto es,

$$\mathbf{P}(n\overline{X} = k) = \binom{nm}{k} p^k (1-p)^{nm-k} \qquad \text{para cada } k = 0, 1, \ldots, nm,$$

de manera que $\overline{X}$ toma los valores los valores $0, 1/n, 2/n, \ldots, m$ con las mismas probabilidades con las que $n\overline{X}$ toma los valores $0, 1, \ldots, nm$. ♣

Ejemplo 4.2.2. *Poisson.*

Si X es una $\text{POISS}(\lambda)$ con $\lambda > 0$, esto es,

$$\mathbf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad \text{para cada } k = 0, 1, 2 \ldots,$$

entonces (véase la nota 4.2.1) la variable $n\overline{X}$ es $\text{POISS}(n\lambda)$, es decir,

$$\mathbf{P}(n\overline{X} = k) = e^{-\lambda n} \frac{(\lambda n)^k}{k!} \qquad \text{para cada } k = 0, 1, 2 \ldots,$$

y por tanto $\overline{X}$ toma los valores $0, 1/n, 2/n, 3/n, \ldots$ con las mismas probabilidades (escritas arriba) con las que $n\overline{X}$ toma los valores $0, 1, 2, 3 \ldots$. ♣

☞ **Nota** 4.2.1. Sean $X \sim \text{POISS}(\mu)$ e $Y \sim \text{POISS}(\nu)$ independientes. Comprobamos ahora que $X + Y$ es $\text{POISS}(\mu + \nu)$. Para cada entero $k \geq 0$ se tiene que

$$\begin{aligned}\mathbf{P}(X + Y = k) &= \sum_{j=0}^{k} \mathbf{P}(X = j, Y = k - j) = \sum_{j=0}^{k} \mathbf{P}(X = j)\, \mathbf{P}(Y = k - j) \\ &= \sum_{j=0}^{k} e^{-\mu} \frac{\mu^j}{j!} e^{-\nu} \frac{\nu^{k-j}}{(k-j)!} = e^{-\mu-\nu} \frac{1}{k!} \sum_{j=0}^{k} \binom{k}{j} \mu^j \nu^{k-j} = e^{-(\mu+\nu)} \frac{1}{k!} (\mu + \nu)^k .\end{aligned}$$

En el segundo paso hemos usado la independencia, y en el último, el binomio de Newton. De aquí se deduce la observación usada antes de que $n\overline{X} \sim \text{POISS}(n\lambda)$.

Ejemplo 4.2.3. *Geométrica y binomial negativa.*

Si X es una GEO(p), entonces la suma $n\overline{X}$ es una binomial negativa BINNEG(n,p).

Una BINNEG(n,p) cuenta el número de lanzamientos (de una moneda con probabilidad de cara p) hasta obtener n caras.

De hecho, una GEO(p) es una BINNEG$(1,p)$ y si X e Y son, respectivamente, BINNEG(k,p) y BINNEG(m,p) (mismo parámetro p) e independientes, entonces $X+Y$ es una BINNEG$(k+m,p)$. Por tanto, además, si X es BINNEG(m,p), entonces $n\overline{X}$ es BINNEG(nm,p). Véanse los detalles en el ejercicio 4.3. ♣

Ejemplo 4.2.4. *Exponencial y Gamma.*

Si X es EXP(λ) con $\lambda > 0$, entonces $n\overline{X}$ es una variable GAMMA(λ, n).

Esto se sigue de que X es GAMMA$(\lambda, 1)$ y que, por tanto, $n\overline{X}$ es GAMMA(λ, n).

En general, si X es GAMMA(λ, t), entonces $n\overline{X}$ es GAMMA(λ, nt). Véase la sección 2.3.4, y en particular la proposición 2.15. ♣

☞ **Nota** 4.2.2. Como caso particular del ejemplo anterior, si tomamos $\lambda = 1/2$, entonces $n\overline{X}$ es una variable GAMMA$(1/2, 2n/2)$, es decir, una χ^2_{2n}. Esto supone que la suma de n exponenciales independientes de parámetro $\lambda = 1/2$ se distribuye de la misma manera que la suma de $2n$ normales estándar al cuadrado independientes. Asombroso.

El caso en el que X es una variable normal es sin duda el más relevante.

Ejemplo 4.2.5. *Normal.*

Si X es $\mathcal{N}(\mu, \sigma^2)$, entonces $n\overline{X}$ es una variable $\mathcal{N}(n\mu, n\sigma^2)$ y, de hecho,

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

Esto se sigue de la proposición 2.10. ♣

Salvo en casos muy particulares, como los que hemos visto, obtener (exactamente) la distribución de probabilidad de $\overline{X}$ en fórmula cerrada suele ser imposible.

Ejemplo 4.2.6. *Lanzamiento de dados.*

El experimento consiste en lanzar un dado (regular) n veces. Cada lanzamiento es una variable X que toma los valores $1, 2, \dots, 6$ con probabilidad $1/6$ cada uno de ellos. Interesa entender $\overline{X}$, o quizás $n\overline{X}$.

Digamos, por ejemplo, que $n = 100$. La variable $n\overline{X}$ toma valores entre 6 y 600. ¿Pero cuál es la probabilidad de obtener, por ejemplo, una suma total entre 320 y 360? Parece un cálculo inabordable, incluso en este caso en el que la distribución de base es uniforme, y por tanto el cálculo es pura combinatoria. ♣

Sin embargo, disponemos de ese asombroso y maravilloso resultado conocido como el teorema central del límite[1], al que nos referiremos a veces con el acrónimo TCL,

[1] O teorema del límite central, en función de la religión que profese.

que en su versión más habitual dice que, si $X_1, X_2, \ldots$ es una sucesión de variables idénticas (con varianza finita) e independientes, entonces, la variable suma $\sum_{i=1}^n X_i$ (o quizás la variable promedio) tiende en distribución, tras la pertinente tipificación, y cuando $n \to \infty$, a una normal estándar.

Para ulteriores referencias, recogemos este resultado en el siguiente:

Teorema 4.2 (TCL) *Sea X una variable aleatoria con $\mathbf{E}(X^2) < \infty$. Consideremos una sucesión $X_1, X_2, \ldots$ de variables aleatorias, todas ellas clones independientes de X. Llamemos $\mathbf{E}(X) = \mu$ y $\mathbf{V}(X) = \sigma^2$. Entonces*

$$(\star) \qquad \sqrt{n}\,|\overline{X} - \mu| \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2) \qquad \textit{cuando } n \to \infty.$$

El símbolo $\xrightarrow{\text{d}}$ indica convergencia en distribución.

Una manera alternativa de escribir el resultado anterior, que quizás le resulte más familiar al lector, es la siguiente:

$$(\star\star) \qquad \frac{\overline{X} - \mathbf{E}(\overline{X})}{\sigma(\overline{X})} = \frac{\overline{X} - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)/n}} \xrightarrow{\text{d}} \mathcal{N}(0,1) \qquad \text{cuando } n \to \infty,$$

que nos dice que la función de distribución de la versión tipificada de $\overline{X}$ tiende, cuando $n \to \infty$, a la función de distribución Φ de la normal estándar: para cada $t \in \mathbb{R}$,

$$(\star\star\star) \qquad \mathbf{P}\Big(\frac{\overline{X} - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)/n}} \le t\Big) \longrightarrow \Phi(t) \qquad \text{cuando } n \to \infty.$$

Las tres expresiones, $(\star)$, $(\star\star)$ y $(\star\star\star)$, significan lo mismo.

En cualquier de ellas intervienen únicamente $\mathbf{E}(X)$ y $\mathbf{V}(X)$, y no la distribución explícita de X. Aunque, claro, se trata de resultados asintóticos.

Le habremos de sacar mucho jugo a este teorema más adelante. Por ahora, una par de aplicaciones directas.

Ejemplo 4.2.7. *Lanzamos la moneda* (*regular*) 10 000 *veces. Queremos estimar la probabilidad de que el promedio de caras obtenido esté entre el* 48 % *y el* 52 %.

La variable X de referencia es una $\textsc{ber}(1/2)$, cuya media es $\mathbf{E}(X) = 1/2$ y cuya varianza vale $\mathbf{V}(X) = 1/4$. La variable promedio $\overline{X} = \frac{1}{10000}(X_1 + \cdots + X_{10000})$ tiene media $\mathbf{E}(\overline{X}) = 1/2$ y varianza $\mathbf{V}(\overline{X}) = 1/40\,000$.

La pregunta del enunciado se refiere a la probabilidad

$$\mathbf{P}(48\,\% \le \overline{X} \le 52\,\%).$$

La distribución de $n\overline{X}$ es conocida (ejemplo 4.2.1): es una $\textsc{bin}(10\,000; 1/2)$. Así que

$$\mathbf{P}(48\,\% \le \overline{X} \le 52\,\%) = \mathbf{P}(4800 \le X_1 + \cdots + X_{10000} \le 5200) = \frac{1}{2^{10000}} \sum_{j=4800}^{5200} \binom{10000}{j}.$$

Se trata de una fórmula *exacta*, pero calcular su valor es computacionalmente muy exigente. Piense, lector, que por ejemplo $\binom{10000}{4800}$ es un número con más de 3000 cifras decimales. Por cierto, el valor (calculado con un ordenador, por supuesto) de la probabilidad pedida, redondeada a cinco decimales, es 99.99394 %.

La estimación que nos daría la desigualdad de Chebyshev sería

$$\begin{aligned}\mathbf{P}(48\,\% \le \overline{X} \le 52\,\%) &= \mathbf{P}(-2\,\% \le \overline{X} - 50\,\% \le 2\,\%) = \mathbf{P}(|\overline{X} - \tfrac{1}{2}| \le \tfrac{2}{100})\\ &= 1 - \mathbf{P}(|\overline{X} - \tfrac{1}{2}| > \tfrac{2}{100}) \ge 1 - \frac{\mathbf{V}(\overline{X})}{(2/100)^2} = \frac{15}{16} \approx 93.75\,\%,\end{aligned}$$

lo que al menos nos da idea de que la probabilidad es muy alta.

Alternativa (y aproximadamente), podemos proceder como sigue:

$$\begin{aligned}\mathbf{P}(48\,\% \le \overline{X} \le 52\,\%) &= \mathbf{P}(-2\,\% \le \overline{X} - \mathbf{E}(X) \le 2\,\%)\\ &= \mathbf{P}\Big(\frac{-2\,\%}{\sqrt{1/40000}} \le \frac{\overline{X} - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)/10000}} \le \frac{-2\,\%}{\sqrt{1/40000}}\Big)\\ &\approx \mathbf{P}(-4 \le \mathcal{N}(0,1) \le 4) = \Phi(4) - \Phi(-4) = 99.99367\,\%,\end{aligned}$$

redondeando de nuevo a cinco decimales. ♣

Ejemplo 4.2.8. *Lanzamiento de dados, segunda parte.*

Retomamos el ejemplo 4.2.6. Para calcular la probabilidad de que tras $n = 100$ lanzamientos de un dado regular obtengamos una suma total de puntos entre 320 y 360 procedemos como sigue. Calculamos primero $\mathbf{E}(X) = 7/2$ y $\mathbf{V}(X) = 35/12$, de manera que $\mathbf{E}(n\overline{X}) = 350$ y $\mathbf{V}(n\overline{X}) = 875/3$. Luego,

$$\begin{aligned}\mathbf{P}\Big(320 \le \sum_{i=1}^{100} X_i \le 360\Big) &= \mathbf{P}\Big(-\frac{30}{\sqrt{875/3}} \le \frac{\sum_{i=1}^{100} X_i - 350}{\sqrt{875/3}} \le \frac{10}{\sqrt{875/3}}\Big)\\ &\approx \Phi(0.5855) - \Phi(-1.7566) = 72.091\,\% - 3.949\,\% = 68.142\,\%.\end{aligned}$$

Aquí estamos suponiendo que $n = 100$ es lo suficientemente grande como para que el teorema central del límite se pueda aplicar. ♣

☞ **Nota** 4.2.3. En estos dos ejemplos hemos utilizado valores específicos de la función Φ de distribución de la normal estándar. En Excel, los valores de $\Phi(x)$ se pueden calcular con la instrucción `=distr.norm.estand(x)`.

4.3. Sobre la cuasivarianza muestral

Partimos, de nuevo, de una muestra aleatoria $\mathbb{X} = (X_1, \ldots, X_n)^\intercal$ de tamaño n de la variable X. La variable cuasivarianza muestral

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{4.3}$$

es un estadístico más complejo que $\overline{X}$ y, en general, la información de que se dispone sobre S^2 es más pobre.

Tras expandir el cuadrado en (4.3), escribimos S^2 en la forma alternativa

$$(n-1)S^2 = \sum_{j=1}^{n} X_j^2 - n\overline{X}^2 , \tag{4.4}$$

o también, usando que $\overline{X} = (X_1 + \cdots + X_n)/n$, en la forma

$$n(n-1)S^2 = (n-1)\sum_{j=1}^{n} X_j^2 - \sum_{i\neq j} X_i X_j , \tag{4.5}$$

en la que se exhibe explícitamente cómo se combinan las variables $X_1, \ldots, X_n$, y que será útil en ciertos cálculos. Por cierto, en forma matricial, (4.5) se escribe

$$nS^2 = \mathbb{X}^{\intercal}\, \Omega_n\, \mathbb{X},$$

donde Ω_n es la matriz simétrica con coeficientes

$$\omega_{i,j} = \begin{cases} 1\,, & \text{si } i = j\,, \\ -1/(n-1)\,, & \text{si } i \neq j\,. \end{cases}$$

Obsérvese que $\mathbf{x}^{\intercal}\, \Omega_n\, \mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$ (es decir, Ω_t es semidefinida positiva); de hecho, $\mathbf{x}^{\intercal}\, \Omega_n\, \mathbf{x} = 0$ si y solo si $\mathbf{x}$ es un vector constante, apelando a (4.3).

4.3.1. Media de S^2

Proposición 4.3 *Para cualquier variable X con $\mathbf{E}(X^2) < +\infty$ se tiene que*

$$\mathbf{E}(S^2) = \mathbf{V}(X)\,.$$

Es decir, *en media*, S^2 vale la varianza de X. Esta proposición 4.3 muestra por qué en S^2 se divide por $n-1$ y no por n.

Conviene remarcar, y lo remarcamos, que

- para el estadístico D^2 (varianza muestral), que se define como S^2, pero dividiendo por n en lugar de $n-1$, se tiene que $\mathbf{E}(D^2) = \frac{n-1}{n}\,\mathbf{E}(S^2) < \mathbf{V}(X)$;
- y que $\mathbf{E}(S) \neq \sqrt{\mathbf{V}(X)}$.

☞ **Nota** 4.3.1. De hecho, por la desigualdad de Jensen (teorema 2.1), $\mathbf{E}(S)^2 \leq \mathbf{E}(S^2) = \mathbf{V}(X)$, así que $\mathbf{E}(S) \leq \sqrt{\mathbf{V}(X)}$. La igualdad en $\mathbf{E}(S) \leq \sqrt{\mathbf{V}(X)}$ se tiene sólo si S^2 es constante, es decir, cuando X es constante. (Una forma de comprobar esto último es apelar al corolario 4.5.)

DEMOSTRACIÓN DE LA PROPOSICIÓN 4.3. Usando (4.4) se tiene que

$$(n-1)\mathbf{E}(S^2) = n\,\mathbf{E}(X^2) - n\,\mathbf{E}(\overline{X}^2)\,,$$

que podemos expresar en términos sólo de X usando que

$$\mathbf{E}(\overline{X}^2) = \mathbf{V}(\overline{X}) + \mathbf{E}(\overline{X})^2 = \frac{1}{n}\,\mathbf{V}(X) + \mathbf{E}(X)^2\,,$$

para obtener que

$$\begin{aligned}(n-1)\mathbf{E}(S^2) &= n\,\mathbf{E}(X^2) - n\left(\frac{1}{n}\,\mathbf{V}(X) + \mathbf{E}(X)^2\right)\\ &= n\left(\mathbf{E}(X^2) - \mathbf{E}(X)^2\right) - \mathbf{V}(X) = (n-1)\mathbf{V}(X)\,,\end{aligned}$$

y, finalmente, que

$$\mathbf{E}(S^2) = \mathbf{V}(X)\,.$$

Una prueba alternativa, algebraicamente más directa, partiendo de (4.5) en lugar de (4.4), va como sigue:

$$\begin{aligned}n(n-1)\mathbf{E}(S^2) &= (n-1)\sum_{i=1}^{n}\mathbf{E}(X_i^2) - \sum_{i\neq j}\mathbf{E}(X_i\cdot X_j)\\ &= (n-1)n\,\mathbf{E}(X^2) - n(n-1)\,\mathbf{E}(X)^2 = n(n-1)\,\mathbf{V}(X)\,,\end{aligned}$$

donde hemos usado la independencia de las X_i (para escribir que $\mathbf{E}(X_i\cdot X_j) = \mathbf{E}(X_i)\,\mathbf{E}(X_j)$) y que todas las X_i con clones de X. ■

☞ **Nota** 4.3.2. Supongamos que $\mu = \mathbf{E}(X)$ es *conocida*, y formemos

$$T^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2\,.$$

Obsérvese que se divide por n y no por $n-1$. Se tiene

$$nT^2 = \sum_{j=1}^{n}X_j^2 - n\mu^2\,,$$

así que $n\,\mathbf{E}(T^2) = n\,\mathbf{E}(X^2) - n\mu^2 = n\mathbf{V}(X)$, y por tanto $\mathbf{E}(T^2) = \mathbf{V}(X)$.

4.3.2. Varianza de S^2

Arrancamos con la expresión de la varianza de S^2 para variables tipificadas:

Proposición 4.4 *Si X es una variable tipificada (es decir, $\mathbf{E}(X) = 0$ y $\mathbf{E}(X^2) = \mathbf{V}(X) = 1$) con $\mathbf{E}(X^4) < +\infty$, entonces*

$$\mathbf{V}(S^2) = \frac{1}{n}\,\mathbf{E}(X^4) - \frac{n-3}{n(n-1)}\,.$$

☞ **Nota** 4.3.3. En la expresión anterior suponemos que $n \geq 2$. Nótese cómo de ella se deduce, de pasada, que $\mathbf{E}(X^4) > (n-3)/(n-1)$ para todo $n \geq 2$ si la variable X está tipificada. Obsérvese que $(n-3)/(n-1)$ tiende a 1 cuando $n \to \infty$.

De la proposición 4.4, como comprobaremos en un momento, se deduce que:

Corolario 4.5 *Para cualquier variable X con $\mathbf{E}(X^4) < +\infty$ se tiene que*

$$\mathbf{V}(S^2) = \frac{1}{n}\,\mathbf{E}\big((X-\mathbf{E}(X))^4\big) - \frac{n-3}{n(n-1)}\,\mathbf{E}\big((X-\mathbf{E}(X))^2\big)^2 .$$

Así que, cuando $n \to \infty$, se tiene que $\mathbf{V}(S^2) \to 0$.

DEMOSTRACIÓN DEL COROLARIO 4.5. Sea Y la versión tipificada de X, es decir, $Y := (X-\mu)/\sigma$, donde $\mu = \mathbf{E}(X)$ y $\sigma^2 = \mathbf{V}(X)$. Obsérvese que $X = \sigma Y + \mu$, y también que $\overline{X} = \sigma\,\overline{Y} + \mu$.

Esto nos dice que

$$S_X^2 \equiv \sigma^2 S_Y^2 .$$

Por tanto, aplicando la proposición 4.4,

$$\mathbf{V}(S_X^2) = \sigma^4\mathbf{V}(S_Y^2) = \sigma^4\Big(\frac{1}{n}\mathbf{E}(Y^4) - \frac{n-3}{n(n-1)}\Big) = \sigma^4\Big(\frac{1}{n}\mathbf{E}((X-\mu)^4)\frac{1}{\sigma^4} - \frac{n-3}{n(n-1)}\Big) .$$

■

DEMOSTRACIÓN DE LA PROPOSICIÓN 4.4. Como la variable X está tipificada, es decir, como $\mathbf{E}(X) = 0$ y $\mathbf{V}(X) = \mathbf{E}(X^2) = 1$, aplicando la proposición 4.3 obtenemos que

$$(\star) \qquad \mathbf{V}(S^2) = \mathbf{E}(S^4) - \mathbf{E}(S^2)^2 = \mathbf{E}(S^4) - \mathbf{V}(X)^2 = \mathbf{E}(S^4) - 1,$$

Solo resta calcular $\mathbf{E}(S^4)$. Para este cálculo usamos la expresión dada en (4.5),

$$n(n-1)S^2 = (n-1)\sum_{j=1}^n X_j^2 - \sum_{i\neq j} X_i X_j .$$

Hemos de elevar al cuadrado y aplicar esperanzas:

$$n^2(n-1)^2\,\mathbf{E}(S^4) = \mathbf{E}\Big[\Big((n-1)\sum_{j=1}^n X_j^2 - \sum_{i\neq j} X_i X_j\Big)^2\Big]$$

Pero no desarrollamos completamente el cuadrado de la derecha sino que anticiparemos aquellos sumandos del desarrollo que, como consecuencia de que $\mathbf{E}(X) = 0$ y de que las X_j son independientes, sabemos que al tomar esperanzas se anularán.

a) El término con el cuadrado de la primera suma nos da:

$$\begin{aligned}\mathbf{E}\Big[\Big((n-1)\sum_{j=1}^n X_j^2\Big)^2\Big] &= (n-1)^2\,\mathbf{E}\Big[\Big(\sum_{j=1}^n X_j^2\Big)^2\Big] \\ &= (n-1)^2\left[n\mathbf{E}(X^4) + n(n-1)\mathbf{E}(X^2)^2\right] = (n-1)^2 n\mathbf{E}(X^4) + (n-1)^2 n(n-1).\end{aligned}$$

b) Vamos con el término con el cuadrado de la segunda suma. Al desarrollar

$$\mathbf{E}\Big(\Big(\sum_{i\neq j} X_i X_j\Big)^2\Big),$$

los únicos sumandos no nulos son aquellos en los que se multiplican términos X_iX_j con X_iX_j, o términos X_iX_j con X_jX_i. Por tanto,

$$\mathbf{E}\Big(\Big(\sum_{i\neq j} X_iX_j\Big)^2\Big) = 2n(n-1)\mathbf{E}(X^2)^2 = 2n(n-1).$$

c) Por último, todos los sumandos del término cruzado

$$\mathbf{E}\Big(\Big((n-1)\sum_{k=1}^{n} X_k^2\Big)\cdot\Big(\sum_{i\neq j} X_iX_j\Big)\Big)$$

son nulos, pues todos ellos son combinaciones de variables del tipo $X_k^2X_iX_j$, con $i\neq j\neq k$, o bien del tipo $X_k^3X_j$, con $j\neq k$, y en cualquiera de ellos aparece (al menos) un factor $\mathbf{E}(X_i)$, que es 0.

Poniendo todo esto junto obtenemos que

$$n^2(n-1)^2\mathbf{E}(S^4) = (n-1)^2n\,\mathbf{E}(X^4) + (n-1)^2n(n-1) + 2n(n-1)\,,$$

es decir,

$$\mathbf{E}(S^4) = \frac{1}{n}\,\mathbf{E}(X^4) + \frac{(n-1)^2+2}{n(n-1)}\,.$$

De manera que, recordando $(\star)$, deducimos que

$$\mathbf{V}(S^2) = \mathbf{E}(S^4) - 1 = \frac{1}{n}\,\mathbf{E}(X^4) + \frac{(n-1)^2+2}{n(n-1)} - 1 = \frac{1}{n}\,\mathbf{E}(X^4) - \frac{n-3}{n(n-1)}\,,$$

como queríamos. ■

4.3.3. Concentración de S^2 alrededor de σ^2

Sea X una variable aleatoria X con $\mathbf{E}(X^4) < +\infty$. Llamemos $\mu = \mathbf{E}(X)$ y $\sigma^2 = \mathbf{V}(X)$. De la variable aleatoria S^2 ya sabemos que $\mathbf{E}(S^2) = \sigma^2$ (proposición 4.3). Del corolario 4.5 se deduce inmediatamente que, si $n\geq 3$,

$$\mathbf{V}(S^2) \leq \frac{1}{n}\,\mathbf{E}\big((X-\mathbf{E}(X))^4\big)\,. \tag{4.6}$$

Usando esto y la desigualdad de Chebyshev (teorema 2.2), aplicada a S^2, obtenemos que, para todo $\varepsilon>0$,

$$\mathbf{P}\big(|S^2-\sigma^2|\geq\varepsilon\big) \leq \frac{\mathbf{V}(S^2)}{\varepsilon^2} \leq \frac{\mathbf{E}\big((X-\mu)^4\big)}{n\varepsilon^2} \tag{4.7}$$

(compárese con (4.1)), y, en particular que, para todo $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbf{P}\big(|S^2-\sigma^2|\geq\varepsilon\big) = 0\,.$$

Una versión alternativa de (4.7) sería

$$\mathbf{P}\big(|S^2-\sigma^2| \geq \varepsilon \sqrt{\mathbf{E}((X-\mu)^4)}\ \big) \leq \frac{1}{n\varepsilon^2},$$

en la que se exhibe la unidad "natural" de desviación de S^2 en torno a su media σ^2, que es ahora la raíz cuadrada del cuarto momento (compárese con (4.2)).

4.3.4. Distribución de S^2

Determinar explícitamente la distribución de la variable S^2 es, salvo en algún caso particular, una tarea complicada.

EJEMPLO 4.3.1. *Veamos el caso en el que X es* BER(p).

Sabemos que $n\overline{X} \sim$ BIN(n,p). Pero como además en este caso X^2 también es BER(p), resulta que

$$(n-1)S^2 = \sum_{j=1}^n X_j^2 - n\overline{X}^2 = \sum_{j=1}^n X_j - n\overline{X}^2 = n\overline{X} - n\overline{X}^2 = n\overline{X}(1-\overline{X})\,,$$

de manera que

$$n(n-1)S^2 = Z(n-Z), \qquad \text{donde } Z = n\overline{X} \text{ es una BIN}(n,p).$$

Es decir, la variable S^2 toma los valores

$$0, \frac{n-1}{n(n-1)}, \frac{2(n-2)}{n(n-1)}, \frac{3(n-3)}{n(n-1)}, \dots, \frac{n-1}{n(n-1)}, 0$$

con probabilidades

$$\mathbf{P}\Big(S^2 = \frac{j(n-j)}{n(n-1)}\Big) = \binom{n}{j} p^j\,(1-p)^{n-j} \qquad j=0,1,2,\dots,n.$$

♣

Más adelante determinaremos también la distribución explícita de S^2 en el caso en el que X sea una variable normal (véase el teorema 4.6).

4.4. Sobre máximos y mínimos

Dada una muestra aleatoria $(X_1,\dots,X_n)$ de X, consideramos los estadísticos

$$M = \text{máx}\{X_1,X_2,\dots,X_n\} \qquad \text{y} \qquad m = \text{mín}\{X_1,X_2,\dots,X_n\}.$$

En cuanto a la función de distribución del máximo, para $t\in\mathbb{R}$, como

$$\begin{aligned}\mathbf{P}(M\leq t) &= \mathbf{P}(\text{máx}(X_1,\dots,X_n)\leq t) = \mathbf{P}(X_1\leq t,\dots,X_n\leq t)\\ &= \mathbf{P}(X_1\leq t)\cdots\mathbf{P}(X_n\leq t) = \mathbf{P}(X\leq t)^n\end{aligned} \tag{4.8}$$

tenemos que

$$F_M(t) = F_X(t)^n. \tag{4.9}$$

Por lo tanto, en el caso de que X sea una variable continua, se tiene que la función de densidad del máximo es, para $t \in \mathbb{R}$,

$$f_M(t) = n\, F_X(t)^{n-1}\, f_X(t). \tag{4.10}$$

Para el cálculo análogo correspondiente al estadístico del mínimo, conviene pasar al complementario y argumentar como sigue:

$$\begin{aligned} \mathbf{P}(m > t) &= \mathbf{P}(\text{mín}(X_1, \dots, X_n) > t) = \mathbf{P}(X_1 > t, \dots, X_n > t) \\ &= \mathbf{P}(X_1 > t) \cdots \mathbf{P}(X_n > t) = \mathbf{P}(X > t)^n = (1 - F_X(t))^n, \end{aligned} \tag{4.11}$$

lo que nos da que

$$F_m(t) = 1 - (1 - F_X(t))^n \tag{4.12}$$

para $t \in \mathbb{R}$, y derivando, para el caso de X continua, que

$$f_m(t) = n\,(1 - F_X(t))^{n-1}\, f_X(t). \tag{4.13}$$

Obtener la distribución del máximo o el mínimo es algo directo; sin embargo, obtener fórmulas para los momentos, digamos, media y varianza, ya no lo es tanto. Por ejemplo, la media de la variable máximo se calcularía como

$$\mathbf{E}(M) = n \int_{-\infty}^{\infty} t\, F_X(t)^{n-1}\, f_X(t)\, dt,$$

que puede (suele) ser una integral complicada.

Consúltese también el ejercicio 4.5 para el caso de los estadísticos de orden.

Ejemplo 4.4.1. *La variable X tiene función de densidad dada por $f(x) = 2x$ para $x \in [0,1]$ (y es 0 para los restantes valores de x).*

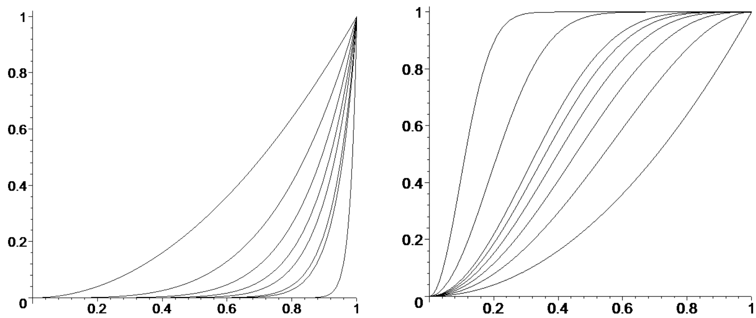
La función de distribución de X es

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ x^2 & \text{si } x \in [0,1], \\ 1 & \text{si } x > 1. \end{cases}$$

De manera que, apelando a (4.9) y (4.12), las respectivas funciones de distribución del máximo M y el mínimo m de muestras de tamaño n vienen dadas por

$$F_M(x) = \begin{cases} 0 & \text{si } x < 0, \\ x^{2n} & \text{si } x \in [0,1], \\ 1 & \text{si } x > 1. \end{cases} \quad \text{y} \quad F_m(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - (1 - x^2)^n & \text{si } x \in [0,1], \\ 1 & \text{si } x > 1. \end{cases}$$

cuyo aspecto (en el intervalo $(0,1)$ y para valores cada vez mayores de n) se aprecia en las dos siguientes gráficas:



Obsérvese cómo la masa de la distribución se concentra cerca del 1 (para el máximo, figura de la izquierda), y cerca del 0 para el mínimo.

Las funciones de densidad son

$$f_M(x) = 2n\,x^{2n-1} \qquad \text{y} \qquad f_m(x) = 2n\,x\,(1-x^2)^{n-1}$$

para $x \in [0,1]$, lo que nos da finalmente que

$$\mathbf{E}(M) = \int_0^1 x\,2n\,x^{2n-1}\,dx = \frac{2n}{2n+1}\,,$$

mientras que

$$\mathbf{E}(M) = 2n\int_0^1 x^2\,(1-x^2)^{n-1}\,dx$$

que, ups, ¿cuánto vale esta integral?

Véanse también los ejemplos 5.1.5 y 5.1.7. ♣

☞ **Nota** 4.4.1. Existe también un resultado asintótico, paralelo al teorema central del límite, para máximos y para mínimos. Para el máximo, por ejemplo, afirma que si la sucesión de máximos (de muestras aleatorias cada vez mayores), convenientemente normalizada, converge en distribución, entonces solo puede hacerlo a tres posibles distribuciones, conocidas como de Gumbel, de Fréchet y de Weibull. En concreto, reza así. Sea F la función de distribución de la X de referencia y llamemos $M_n = \text{máx}(X_1,\dots,X_n)$. Si existen un par de sucesiones de números $a_n > 0$ y b_n tales que

$$\mathbf{P}\Big(\frac{M_n - b_n}{a_n} \le x\Big) = \mathbf{P}(M_n \le a_n x + b) = F^n(a_n\,x + b_n) \xrightarrow{\ \mathrm{d}\ } G(x) \qquad \text{cuando } n \to \infty$$

para cada $x \in \mathbb{R}$, donde $G(x)$ es una función de distribución no degenerada, entonces G ha de ser una de las tres siguientes:

$$\underset{\text{(Gumbel)}}{\Lambda(x) = e^{e^{-x}};} \qquad \underset{\text{(Fréchet)}}{\Phi_\alpha(x) = e^{-x^{-\alpha}}, x \ge 0;} \qquad \underset{\text{(Weibull)}}{\Psi_\alpha(x) = e^{-(-x)^\alpha}, x < 0}$$

Nótese cómo, a diferencia del teorema central del límite, la convergencia en distribución de los máximos normalizados no está garantizada.

4.4.1. Concentración del máximo/mínimo

Sea X una variable aleatoria. Sea β el supremo (esencial) de X, es decir,

$$\mathbf{P}(X \le \beta) = 1 \qquad \text{y} \qquad \mathbf{P}(X \le r) < 1, \quad \text{para todo } r < \beta.$$

En otras palabras, $F_X(\beta) = 1$, pero $F_X(r) < 1$, para todo $r < \beta$. Obsérvese que β puede ser $+\infty$, como por ejemplo en el caso de la normal.

Para cada $r < \beta$,
$$\lim_{n\to\infty} \mathbf{P}(r < M \leq \beta) = 1\,.$$
Esto se sigue de que
$$\mathbf{P}(r < M \leq \beta) = 1 - \mathbf{P}(M \leq r) = 1 - F_M(r) = 1 - F_X(r)^n,$$
y de que $F_X(r) < 1$.

Si β es finito, escribimos entonces que, para todo $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbf{P}(\beta - \varepsilon < M \leq \beta) = 1\,.$$
Esto es, el máximo muestral se concentra alrededor del supremo (esencial) de X.

Análogamente, se tiene que el mínimo se concentra alrededor del ínfimo esencial α definido por
$$\mathbf{P}(X \leq r) = 0 \quad \text{para todo } r < \alpha \qquad \text{y} \qquad \mathbf{P}(X \leq \alpha) > 0.$$
Aquí, α podría ser $-\infty$.

4.5. Muestreo aleatorio de variables normales

Tratamos aquí, para el caso en el que
$$X \sim \mathcal{N}(\mu, \sigma^2),$$
la **distribución exacta de los estadísticos $\overline{X}$ y S^2**.

Recordemos (ejemplo 4.2.5 y proposición 2.10) que ya conocemos la distribución exacta de $\overline{X}$:
$$\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)\,.$$

Vamos a obtener ahora:

- la distribución exacta de S^2,
- y además, y esto es un hecho crucial (en cuanto a su aplicación práctica) y asombroso, que $\overline{X}$ y S^2 son variables aleatorias … *independientes*.

Nótese que las variables
$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \qquad \text{y} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$$
dependen (ambas) del mismo conjunto de variables, $X_1, \dots, X_n$, lo que no sugiere en absoluto posible independencia.

De hecho, y como piedra de toque, recordamos del ejemplo 4.3.1 que, cuando $X \sim \text{BER}(p)$, entonces $(n-1)S^2 = n\overline{X}(1-\overline{X})$. Así que $\overline{X}$ y S^2 están bien lejos de ser independientes. O, más sencillo, si X es la variable que toma los valores ± 1 con probabilidad $1/2$ entonces $(n-1)S^2 = n - n\overline{X}^2$, que no es independiente de $\overline{X}$.

4.5.1. Distribución (conjunta) de $\overline{X}$ y S^2 cuando X es normal

Veamos, como ilustración, el caso $X \sim \mathcal{N}(0,1)$ para una muestra $\mathbb{X} = (X_1, X_2)^\intercal$ de tamaño $n = 2$. El vector $\mathbb{X}$ sigue una $\mathcal{N}(\mathbf{0}, I)$, donde I es la matriz identidad 2×2.

En este caso tenemos que

$$\overline{X} = \frac{X_1 + X_2}{2},$$
$$S^2 = \big(X_1 - \tfrac{1}{2}(X_1 + X_2)\big)^2 + \big(X_2 - \tfrac{1}{2}(X_1 + X_2)\big)^2 = \frac{(X_1 - X_2)^2}{2}.$$

Consideramos ahora el vector aleatorio $\mathbb{Y} = (Y_1, Y_2)^\intercal$ dado por

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}}_{=O} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \qquad \text{es decir,} \qquad \begin{cases} Y_1 = \frac{1}{\sqrt{2}}\,(X_1 + X_2), \\ Y_2 = \frac{1}{\sqrt{2}}\,(X_1 - X_2). \end{cases}$$

Como la matriz O es ortogonal, resulta que el vector $\mathbb{Y}$ también sigue una $\mathcal{N}(\mathbf{0}, I)$, por el lema 3.5 (aunque aquí podríamos haber hecho la comprobación a mano, con el cambio de variables correspondiente). En particular Y_1 e Y_2 son *normales estándar independientes.*

Pero como
$$\overline{X} = \frac{1}{\sqrt{2}} Y_1 \qquad \text{y} \qquad S^2 = Y_2^2$$
deducimos, de una tacada, que $\overline{X} \sim \mathcal{N}(0, 1/2)$, que $S^2 \sim \chi_1^2$ (pues es una normal estándar al cuadrado), y que $\overline{X}$ y S^2 son variables independientes.

La generalización de este argumento da lugar al siguiente teorema:

Teorema 4.6 (Fisher–Cochran) *Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces*

1) *$\overline{X}$ es asimismo normal: $\overline{X}$ es $\mathcal{N}(\mu, \sigma^2/n)$.*
2) *$(n-1)S^2/\sigma^2$ es χ_{n-1}^2.*
3) *$\overline{X}$ y S^2 son variables aleatorias **independientes**.*

DEMOSTRACIÓN. La parte 1) ya es conocida (ejemplo 4.2.5), así que basta probar los otros dos enunciados.

El argumento tiene dos partes. Primero comprobamos que basta probar el resultado en el caso en el que X es una normal estándar.

Veamos. Dada $X \sim \mathcal{N}(\mu, \sigma^2)$, consideramos la variable normal estándar $Y = (X - \mu)/\sigma$ y la muestra $\mathbb{Y} = (Y_1, \dots, Y_n)^\intercal$. Obsérvese que

$$(\star) \quad \overline{Y} = (\overline{X} - \mu)/\sigma, \qquad \text{y que} \qquad (\star\star) \quad (n-1)S_Y^2 = (n-1)S_X^2/\sigma^2.$$

La independencia de $\overline{Y}$ y S_Y^2 nos daría la independencia de $\overline{X}$ y S_X^2. El que $\overline{Y}$ siguiera una $\mathcal{N}(0, 1/n)$ llevaría a que $\overline{X}$ fuera una $\mathcal{N}(\mu, \sigma^2/n)$, por $(\star)$, y si $(n-1)S_Y^2 \sim \chi_{n-1}^2$, entonces tendríamos que $(n-1)S_X^2/\sigma^2$ seguiría también una χ^2_{n-1}, por $(\star\star)$. Así que, como afirmábamos, basta probar 2) y 3) para el caso tipificado.

Digamos entonces que X es una normal estándar, y consideremos la muestra $\mathbb{X} = (X_1, \dots, X_n)^\intercal$, que sigue una $\mathcal{N}(\mathbf{0}, I)$, con I la matriz identidad $n \times n$.

Sea A una matriz $n \times n$ ortogonal, cuya primera fila es

$$\begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}$$

Esta fila es un vector de $\mathbb{R}^n$ de módulo 1. El resto de las filas de la matriz A se pueden obtener por ortonormalización de Gram–Schmidt de la base formada por esa primera fila y $n-1$ vectores de la base canónica.

Sea $\mathbb{Z}$ el vector columna aleatorio dado por

$$\mathbb{Z} = A\mathbb{X}.$$

Como A es ortogonal y $\mathbb{X} \sim (\mathbf{0}, I)$, $\mathbb{Z}$ es un vector normal estándar: sus componentes son normales estándar $Z_1, Z_2, \dots, Z_n$ independientes. Véase el lema 3.5.

Obsérvese que $Z_1 = \sqrt{n}\,\overline{X}$. De manera que $\overline{X} = \frac{1}{\sqrt{n}} Z_1$ es $\mathcal{N}(0, 1/n)$ (ésta es la parte 1) del teorema, de nuevo).

Por otro lado, en general, se tiene que

$$(n-1)S^2 = \sum_{j=1}^n X_j^2 - n\overline{X}^2.$$

Ahora, como A es matriz ortogonal,

$$\sum_{j=1}^n Z_j^2 = \mathbb{Z}^\intercal \mathbb{Z} = \mathbb{X}^\intercal A^\intercal A \mathbb{X} = \mathbb{X}^\intercal \mathbb{X} = \sum_{j=1}^n X_j^2,$$

de manera que usando que $Z_1 = \sqrt{n}\,\overline{X}$ obtenemos que

$$(n-1)S^2 = \sum_{j=1}^n X_j^2 - n\overline{X}^2 = \sum_{j=1}^n Z_j^2 - Z_1^2 = \sum_{j=2}^n Z_j^2 \quad \Longrightarrow \quad (n-1)S^2 = \sum_{j=2}^n Z_j^2,$$

En consecuencia, y de un sólo golpe, tenemos que

- $(n-1)S^2$ es una χ^2_{n-1}, pues las Z_j son normales estándar independientes;
- y las variables $(n-1)S^2$ y $\overline{X}$ son independientes, pues la primera depende de $Z_2, \dots, Z_n$, mientras que la segunda sólo depende de Z_1. ■

☞ **Nota** 4.5.1. Supongamos que X es normal estándar. El teorema 4.6 nos da la distribución de S^2 y nos dice que $(n-1)S^2$ es χ^2_{n-1}, y, por tanto, que

$$(n-1)^2 \mathbf{V}(S^2) = \mathbf{V}(\chi^2_{n-1}) = 2(n-1),$$

de manera que

$$\mathbf{V}(S^2) = \frac{2}{n-1}\,,$$

Por otro lado, la proposición 4.4 (válida en general, no sólo cuando X es normal) nos dice (usando que, para una variable X normal estándar, $\mathbf{E}(X^4) = 3$) que

$$\mathbf{V}(S^2) = \frac{1}{n}\mathbf{E}(X^4) - \frac{n-3}{n(n-1)} = \frac{3}{n} - \frac{n-3}{n(n-1)} = \frac{2}{n-1}\,,$$

como tenía que ser.

Sea ahora $X \sim \mathcal{N}(\mu, \sigma^2)$. Entonces, por el teorema 4.6 de Fisher–Cochran,

1.- $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = Y \sim \mathcal{N}(0,1)$ es normal estándar;

2.- $(n-1)S^2/\sigma^2 = Z_{n-1}$ es una χ^2_{n-1},

3.- además X y Z_{n-1} son independientes.

Consideramos ahora la variable

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

Para determinar su distribución de probabilidad, la reescribimos en términos de las variables Y (normal estándar) y Z_{n-1} (una χ^2_{n-1}) introducidas más arriba,

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{\sqrt{n}(\overline{X} - \mu)}{S} = \frac{\sqrt{n}(\overline{X} - \mu)/\sigma}{S/\sigma} = \frac{Y}{\sqrt{Z_{n-1}/(n-1)}},$$

y reconocemos (recordando la sección 3.3.3) que la variable en cuestión es una variable t de Student con $n-1$ grados de libertad. Para referencia posteriores, recogemos este resultado en la siguiente:

Proposición 4.7 *Si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces*

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim \text{STU}(n-1),$$

es decir, la variable $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ se distribuye como una t de Student con $n-1$ grados de libertad.