

Capítulo 3

La normal multidimensional y distribuciones asociadas

3.1. Resumen ejecutivo de álgebra lineal	2
3.2. La normal multidimensional	3
3.2.1. Polinomios cuadráticos y funciones de densidad	3
3.2.2. La normal multidimensional	4
3.3. Distribuciones asociadas a la normal multidimensional	16
3.3.1. La distribución $\chi^2(n)$ (chi-cuadrado con n grados de libertad)	16
3.3.2. Distribución $F_{n,m}$ con n y m grados de libertad	18
3.3.3. Distribución t_n de Student	20
3.3.4. Distribución de Rayleigh	22

Los vectores normales (que siguen una distribución normal multidimensional) desempeñan un papel central en los modelos probabilistas en los que se apoya la modelización matemática de la inferencia estadística que iremos desarrollando en capítulos posteriores.

En este capítulo estudiamos diversas propiedades de estos vectores normales (apartado 3.2), así como algunas distribuciones de probabilidad (unidimensionales) asociadas y de utilidad como las distribuciones χ^2 (chi-cuadrado), las distribuciones t de Student y las distribuciones F de Fisher (apartado 3.3).

El estudio de la normal multidimensional se apoya en ciertas cuestiones de álgebra lineal, en concreto de formas cuadráticas, que recordamos telegráficamente en la primera sección. El lector interesado en el detalle de las demostraciones puede consultar el capítulo 8.

3.1. Resumen ejecutivo de álgebra lineal

A. De matrices simétricas y ortogonales

Una matriz $A = (a_{ij})$ (cuadrada, de dimensiones $n \times n$) se dice *simétrica* si coincide con su matriz traspuesta: $A^T = A$. Es decir, si $a_{ij} = a_{ji}$ para cada $i \neq j$.

Se dice que una matriz O es *ortogonal* si $O^T = O^{-1}$. Las columnas (o las filas) de O son vectores perpendiculares (dos a dos) y de norma 1.

Toda matriz A simétrica se *orto-diagonaliza*. Es decir, la matriz A , de dimensiones $n \times n$, tiene *autovalores* reales, en \mathbb{R}^n hay una base ortonormal de *autovectores* de A , y A se puede escribir en la forma

$$A = O D O^T,$$

donde

- O es una matriz ortogonal cuyas columnas son los autovectores de A ,
- D es matriz diagonal, cuya diagonal contiene los autovalores $\lambda_1, \dots, \lambda_n$ de A .

Véase el teorema 8.16, llamado espectral, y alrededores. Además (véase la proposición 8.2), $\det(A) = \prod_{j=1}^n \lambda_j$.

B. De formas cuadráticas

Una matriz simétrica A se dice *definida positiva* si la *forma cuadrática* asociada es definida positiva, es decir, si

$$\mathbf{x}^T A \mathbf{x} > 0, \quad \text{para todo } \mathbf{x} \neq \mathbf{0}.$$

Un par de caracterizaciones alternativas de ser definida positiva son las siguientes:

- Una matriz simétrica A es definida positiva si y sólo si sus autovalores son (estrictamente) positivos. Véase el teorema 8.27.
- Una matriz simétrica A es definida positiva si y sólo si admite *una raíz cuadrada* no singular: es decir, si A se factoriza en la forma

$$A = R^T R$$

para una cierta matriz cuadrada R no singular. Consúltase la proposición 8.39.

Finalmente, un par de propiedades útiles:

- Si la matriz A es simétrica y definida positiva, su inversa A^{-1} es asimismo simétrica y definida positiva: los autovalores de A^{-1} son los recíprocos de los de A . Véase la proposición 8.30.
- Si A es una matriz simétrica y definida positiva y si B es una matriz no singular, entonces

$$B^T A B$$

es asimismo simétrica y definida positiva. Consúltase el lema 8.25.

3.2. La normal multidimensional

Comenzamos dando respuesta a la pregunta de cuándo $e^{-H(x_1, \dots, x_n)}$, donde H es un polinomio cuadrático en las variables x_1, \dots, x_n , es una función de densidad. Esta discusión nos llevará a la definición de la normal multidimensional (sección 3.2.2).

3.2.1. Polinomios cuadráticos y funciones de densidad

Todo polinomio $H(\mathbf{x})$ en las variables $\mathbf{x} = (x_1, \dots, x_n)$ de grado 2 (polinomio cuadrático) se escribe en la forma

$$(3.1) \quad H(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} - 2\mathbf{x}^T \mathbf{b} + c, \quad \text{para todo } \mathbf{x} \in \mathbb{R}^n,$$

donde A es una matriz *simétrica* $n \times n$, \mathbf{b} es un vector (columna), $\mathbf{b} \in \mathbb{R}^n$, y c es una constante, $c \in \mathbb{R}$. Aquí, el factor -2 aparece por pura estética.

Supongamos que $\det(A) \neq 0$. Entonces, tomando

$$\mathbf{m} = A^{-1} \mathbf{b} \quad \text{y} \quad d = c - \mathbf{b}^T A^{-1} \mathbf{b},$$

podemos reescribir (3.1) como

$$(3.2) \quad H(\mathbf{x}) = (\mathbf{x} - \mathbf{m})^T A (\mathbf{x} - \mathbf{m}) + d, \quad \text{para todo } \mathbf{x} \in \mathbb{R}^n.$$

Esta reescritura se conoce como *completar cuadrados*.

Nota 3.2.1. En el caso $n = 1$, tenemos que $H(x) = ax^2 - 2bx + c$. Si $a \neq 0$, entonces tomamos $m = b/a$ y $d = c - b^2/a$ para escribir

$$H(x) = a(x - b/a)^2 + (c - b^2/a),$$

que nos da la usual fórmula cuadrática para resolver la ecuación de segundo grado.

Nos preguntamos ahora: dado un polinomio cuadrático H de la forma (3.1),

¿cuándo es $f(\mathbf{x}) = e^{-H(\mathbf{x})}$ una función de densidad en \mathbb{R}^n ?

La presencia de la exponencial nos garantiza que $f \geq 0$ en todo \mathbb{R}^n , así que, para que f sea función densidad, sólo se requiere que $\int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = 1$.

Diagonalizamos y escribimos $A = O^T D O$, donde O es ortogonal y D es diagonal. Recordemos que $\det(A) \neq 0$, de manera que A no tiene autovalores nulos.

Usando la representación de la ecuación (3.2), que $|\det(O)| = 1$, y cambios de variable naturales, calculamos

$$\begin{aligned} \int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} &= \int_{\mathbb{R}^n} e^{-H(\mathbf{x})} d\mathbf{x} = e^{-d} \int_{\mathbb{R}^n} e^{-(\mathbf{x}-\mathbf{m})^T A (\mathbf{x}-\mathbf{m})} d\mathbf{x} = e^{-d} \int_{\mathbb{R}^n} e^{-\mathbf{x}^T A \mathbf{x}} d\mathbf{x} \\ &= e^{-d} \int_{\mathbb{R}^n} e^{-(O\mathbf{x})^T D (O\mathbf{x})} d\mathbf{x} = e^{-d} \int_{\mathbb{R}^n} e^{-\mathbf{x}^T D \mathbf{x}} d\mathbf{x} \\ &= e^{-d} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\sum_{j=1}^n \lambda_j x_j^2} dx_1 \dots dx_n = e^{-d} \prod_{j=1}^n \int_{\mathbb{R}} e^{-\lambda_j x^2} dx. \end{aligned}$$

4 CAPÍTULO 3. LA NORMAL MULTIDIMENSIONAL Y DISTRIBUCIONES ASOCIADAS

Si algún λ_j fuera negativo, la integral ya sería infinita. Supongamos entonces que $\lambda_1, \dots, \lambda_n > 0$. Es decir, que la matriz A es (simétrica y) definida positiva. Entonces, como

$$\int_{\mathbb{R}} e^{-\lambda x^2} dx \stackrel{[\sqrt{2\lambda}x=y]}{=} \frac{1}{\sqrt{2\lambda}} \int_{\mathbb{R}} e^{-y^2/2} dy = \frac{\sqrt{2\pi}}{\sqrt{2\lambda}} = \sqrt{\frac{\pi}{\lambda}}, \quad \text{para todo } \lambda > 0,$$

se tiene que

$$\int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = e^{-d} \frac{\pi^{n/2}}{\sqrt{\prod_{j=1}^n \lambda_j}} = e^{-d} \frac{\pi^{n/2}}{\sqrt{\det(A)}}.$$

Así que para que $f(\mathbf{x})$ sea función de densidad, ha de ser de la forma

$$(3.3) \quad f(\mathbf{x}) = \frac{\sqrt{\det(A)}}{\pi^{n/2}} e^{-(\mathbf{x}-\mathbf{m})^T A (\mathbf{x}-\mathbf{m})}.$$

donde A es una matriz simétrica definida positiva.



Nota 3.2.2. Si $\det(A) = 0$, entonces $e^{-H(\mathbf{x})}$ nunca puede ser función de densidad. Veamos por qué. Ahora, obsérvese, no podemos completar cuadrados para eliminar el término lineal. Así que usamos (3.1) y argumentamos como sigue:

$$\begin{aligned} \int_{\mathbb{R}^n} e^{-H(\mathbf{x})} d\mathbf{x} &= e^{-c} \int_{\mathbb{R}^n} e^{-\mathbf{x}^T A \mathbf{x} + 2\mathbf{x}^T \mathbf{b}} d\mathbf{x} = e^{-c} \int_{\mathbb{R}^n} e^{-\mathbf{x}^T O^T D O \mathbf{x} + 2\mathbf{x}^T \mathbf{b}} d\mathbf{x} \\ &\stackrel{[\mathbf{y}=O\mathbf{x}]}{=} e^{-c} \int_{\mathbb{R}^n} e^{-\mathbf{y}^T D \mathbf{y} + 2\mathbf{y}^T O^T \mathbf{b}} d\mathbf{y} \stackrel{[\mathbf{h}:=O^T \mathbf{b}]}{=} e^{-c} \int_{\mathbb{R}^n} e^{-\mathbf{y}^T D \mathbf{y} + 2\mathbf{y}^T \mathbf{h}} d\mathbf{y} \\ &= e^{-c} \prod_{j=1}^n \int_{\mathbb{R}} e^{-\lambda_j y_j^2 + 2y_j h_j} dy_j. \end{aligned}$$

Como algún λ_j es nulo, la integral en la correspondiente variable y_j vale $+\infty$, sea cual sea el valor del correspondiente h_j .

3.2.2. La normal multidimensional

Dados

- un vector $\mathbf{m} \in \mathbb{R}^n$;
- y una matriz V de dimensiones $n \times n$ simétrica y definida positiva,

decimos que el vector (columna, aleatorio) \mathbb{X} sigue una distribución **normal multidimensional** (de dimensión n) con parámetros \mathbf{m} y V , lo que denotaremos por

$$\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V),$$

si, para todo $\mathbf{x} \in \mathbb{R}^n$, su función de densidad viene dada por

$$(3.4) \quad f_{\mathbb{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(V)}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T V^{-1}(\mathbf{x}-\mathbf{m})}$$

En esta expresión, \mathbf{x} es un vector columna. Nótese que en el exponente aparece la matriz V^{-1} , y no la matriz V . El caso $n = 1$ nos da la habitual expresión (2.19) para la densidad de la normal unidimensional con $\mathbf{m} = \mu \in \mathbb{R}$ y $V = \sigma^2 > 0$.

A los parámetros \mathbf{m} y V nos referiremos como

- el **vector de medias**
- y la **matriz de covarianzas**

de \mathbb{X} . En el teorema 3.1 se justifica esta terminología.

Nota 3.2.3. Observe, lector, que ésta es la expresión general que se ha obtenido en la ecuación (3.3) con $A = \frac{1}{2}V^{-1}$. Esta forma de escribir la función de densidad anticipa que es la matriz $V = (2A)^{-1}$ la que directamente tiene ese papel de matriz de covarianzas (y que el vector \mathbf{m} es un vector de medias).

En lo que sigue utilizaremos repetidamente la siguiente descomposición. Como V es simétrica y definida positiva, la podremos escribir como

$$V = UU^T,$$

con U una cierta matriz no singular. Obsérvese que $\det(V) = \det(U)^2$.

La matriz U *no* es una raíz cuadrada de V , en la terminología del apartado 3.1. Pero como

$$V^{-1} = (U^{-1})^T U^{-1},$$

resulta que U^{-1} es una raíz cuadrada de V^{-1} .

Con estas observaciones, podemos reescribir la densidad $f_{\mathbb{X}}$ dada en (3.4) en la forma alternativa

$$(3.5) \quad f_{\mathbb{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(U)|} e^{-\frac{1}{2} \|U^{-1}(\mathbf{x}-\mathbf{m})\|^2}$$

Nota 3.2.4. Re-comprobación de que $f_{\mathbb{X}}$ es función de densidad.

Por (3.3) ya sabemos que $f_{\mathbb{X}}$ es función de densidad. Pero usemos (3.5) para recomprobarlo. El cambio de variables $\mathbf{x} = \mathbf{m} + U\mathbf{y}$ nos da que

$$\begin{aligned} \int_{\mathbf{x} \in \mathbb{R}^n} f_{\mathbb{X}}(\mathbf{x}) d\mathbf{x} &= |\det(U)| \int_{\mathbf{y} \in \mathbb{R}^n} f_{\mathbb{X}}(\mathbf{m} + U\mathbf{y}) d\mathbf{y} \\ &= |\det(U)| \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(U)|} \int_{\mathbf{y} \in \mathbb{R}^n} e^{-\frac{1}{2} \|\mathbf{y}\|^2} d\mathbf{y} = \left(\frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} y^2} dy \right)^n = 1. \end{aligned}$$

El siguiente resultado describe completamente al vector aleatorio \mathbb{X} , sus marginales y la relación entre ellas.

Teorema 3.1 Si $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$, entonces cada una de las variables X_j componentes de \mathbb{X} es una variable normal. De hecho,

$$X_j \sim \mathcal{N}(m_j, v_{jj}), \quad 1 \leq j \leq n,$$

6 CAPÍTULO 3. LA NORMAL MULTIDIMENSIONAL Y DISTRIBUCIONES ASOCIADAS

y además

$$\text{cov}(X_i, X_j) = v_{ij}, \quad \text{para } 1 \leq i \neq j \leq n.$$

En otras palabras, las marginales de \mathbb{X} son variables normales, y además se tiene que $\mathbf{E}(\mathbb{X}) = \mathbf{m}$ y $\text{cov}(\mathbb{X}) = V$, como anticipábamos al comienzo de esta sección. La demostración del teorema 3.1 aparece más adelante.

Antes, en los apartados A y B siguientes, obtenemos varias propiedades útiles de un vector aleatorio $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$. Obsérvese primero que

$$\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V) \quad \text{si y sólo si} \quad \mathbb{X} - \mathbf{m} \sim \mathcal{N}(\mathbf{0}, V),$$

como se comprueba con un simple cambio de variables.

A. Normal multidimensional estándar

Si $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$ (donde I denota la matriz identidad) se dice que \mathbb{Y} sigue una **normal (multidimensional) estándar**.

Este caso particular es muy especial (por sus propiedades y su relevancia).

Proposición 3.2 *El vector aleatorio $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$ si y sólo si las componentes Y_j de \mathbb{Y} son variables normales **estándar** (unidimensionales) e **independientes**.*

Hacemos notar que el teorema 3.1 nos daría para este caso que las componentes Y_j son normales, y que $\mathbf{E}(\mathbb{Y}) = \mathbf{0}$ y $\text{Cov}(\mathbb{Y}) = I$. Pero aquí, atención, las componentes son además *independientes*.

DEMOSTRACIÓN. Si $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$, entonces

$$(3.6) \quad f_{\mathbb{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\mathbf{y}^\top \cdot \mathbf{y}} = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^n y_j^2} = \prod_{j=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} \right).$$

Así que la función de densidad conjunta se factoriza (lo que nos da la independencia, por la observación (2.7)) en un producto de funciones de densidad, cada una de las cuales es la de una normal (unidimensional) estándar (lo que nos da las marginales).

La prueba de la implicación en el otro sentido consiste en recorrer la fórmula (3.6) de derecha a izquierda. ■

Si las componentes de un vector \mathbb{Y} cualquiera son independientes, entonces son también incorreladas. Esto ocurre siempre. En el caso en el que $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$, vale también el recíproco: el que las componentes sean incorreladas implica también que son independientes. Asombroso.



Nota 3.2.5. Una versión más general de la proposición 3.2 reza así: $\mathbb{Y} \sim \mathcal{N}(\mathbf{m}, D)$, donde D es una matriz diagonal con entradas d_1, \dots, d_n *positivas* en la diagonal, si y sólo si las componentes Y_j de \mathbb{Y} son normales $\mathcal{N}(m_i, d_i)$ independientes. La demostración es análoga a la de la proposición 3.2.



Nota 3.2.6. Por insistir sobre el tema: si las componentes de $\mathbb{Y} = (Y_1, \dots, Y_n)^\top$ son normales (estándar) independientes, entonces $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.

Sin embargo, puede ocurrir que $\mathbb{X} = (X_1, \dots, X_n)^\top$ cumpla que $\mathbf{E}(\mathbb{X}) = \mathbf{m}$, que $\mathbf{cov}(\mathbb{X}) = V$, y que las X_j sean normales, y que a pesar de eso el vector \mathbb{X} no siga una normal multidimensional de parámetros \mathbf{m} y V . Esto es, el recíproco del teorema 3.1 no es cierto en general.

Si por ejemplo X es una variable normal, el vector $(X, -X)$ (cuyas componentes son normales) no es un vector aleatorio normal. Otro ejemplo: sea X una variable normal estándar y sea $a > 0$. Definimos Y (en el mismo espacio que X) como $Y = X$ donde $|X| > a$ y como $Y = -X$ donde $|x| \leq a$. La variable Y es normal estándar, pero (X, Y) no sigue distribución normal.

B. Tipificación

Proposición 3.3 Sea V una matriz simétrica y definida positiva. Escribimos $V = UU^\top$ (esto es, U^{-1} es raíz cuadrada de V^{-1}).

Se tiene que

$$\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V) \iff \mathbb{X} = \mathbf{m} + U\mathbb{Y},$$

donde \mathbb{Y} es una normal estándar, $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.



Nota 3.2.7. En el caso $n = 1$, tenemos que si $X \sim \mathcal{N}(\mu, \sigma^2)$, entonces $X = \mu + \sigma Y$, con $Y \sim \mathcal{N}(0, 1)$, como ya sabíamos.

Usaremos constantemente la parte de “representación” del teorema anterior: todo vector $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$ se puede escribir como $\mathbf{m} + U\mathbb{Y}$, con $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$. La “tipificación” de \mathbb{X} consistiría en la transformación inversa, en escribir $\mathbb{Y} = U^{-1}(\mathbb{X} - \mathbf{m})$.

DEMOSTRACIÓN. Sea $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$, y llamemos $\mathbb{Y} = U^{-1}(\mathbb{X} - \mathbf{m})$. Consideremos la transformación lineal de \mathbb{R}^n en \mathbb{R}^n dada por

$$\mathbf{y} \mapsto \mathbf{x} = \mathbf{m} + U\mathbf{y}.$$

Entonces

$$f_{\mathbb{X}}(\mathbf{x}) |\det(U)| = f_{\mathbb{Y}}(\mathbf{y}).$$

Por la ecuación (3.5) de $f_{\mathbb{X}}$ tenemos que

$$f_{\mathbb{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(U)|} |\det(U)| e^{-\|\mathbf{y}\|^2/2} = \frac{1}{(2\pi)^{n/2}} e^{-\|\mathbf{y}\|^2/2},$$

es decir, que $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.

Invirtiendo los pasos del argumento anterior, se deduce que si $\mathbb{X} = \mathbf{m} + U\mathbb{Y}$, con $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$, entonces $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$. ■

C. Demostración del teorema 3.1

Usaremos en esta demostración la proposición 3.3 y el hecho de que cualquier combinación lineal de variables normales independientes es una variable normal (proposición 2.11).

8 CAPÍTULO 3. LA NORMAL MULTIDIMENSIONAL Y DISTRIBUCIONES ASOCIADAS

Sea $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$. Probamos primero que las componentes son variables normales. Escribimos, siguiendo la proposición 3.3,

$$\mathbb{X} = \mathbf{m} + U\mathbb{Y},$$

donde $V = UU^\top$ e $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.

Leyendo esta relación componente a componente, tenemos que, para $1 \leq j \leq n$,

$$(\star) \quad X_j = m_j + \sum_{k=1}^n u_{jk} Y_k,$$

lo que nos da, usando la proposición 2.11 (pues las Y_k son normales independientes), que cada X_j es normal.

Queda determinar los parámetros de esas normales, y las covarianzas entre ellas.

Obsérvese que, como $UU^\top = V$, se tiene que

$$\sum_{k=1}^n u_{jk}^2 = v_{jj}, \quad \text{para } 1 \leq j \leq n, \quad \text{y} \quad \sum_{k=1}^n u_{ik} u_{jk} = v_{ij}, \quad \text{para } 1 \leq i \neq j \leq n,$$

donde las entradas de V se escriben como v_{ij} , y las de U como u_{ij} .

De (\star) se deduce inmediatamente que

$$\mathbf{E}(X_j) = m_j \quad \text{y} \quad \mathbf{V}(X_j) = \sum_{k=1}^n u_{jk}^2 = v_{jj},$$

pues cada $Y_k \sim \mathcal{N}(0, 1)$. En cuanto a las covarianzas, tenemos que, para $1 \leq i \neq j \leq n$,

$$\begin{aligned} \text{cov}(X_i, X_j) &= \mathbf{E}((X_i - m_i)(X_j - m_j)) = \mathbf{E}\left[\left(\sum_{k=1}^n u_{ik} Y_k\right)\left(\sum_{\ell=1}^n u_{j\ell} Y_\ell\right)\right] \\ &= \mathbf{E}\left[\sum_{k=1}^n \sum_{\ell=1}^n u_{ik} u_{j\ell} Y_k Y_\ell\right] = \sum_{k=1}^n \sum_{\ell=1}^n u_{ik} u_{j\ell} \mathbf{E}(Y_k Y_\ell) = \sum_{k=1}^n u_{ik} u_{jk} = v_{ij}, \end{aligned}$$

puesto que $\mathbf{E}(Y_k Y_\ell) = 0$ salvo cuando $k = \ell$; en este caso, $\mathbf{E}(Y_k^2) = 1$ (recuérdese que las variables Y_j son normales estándar independientes).

Esta segunda parte de la demostración, en términos matriciales (y en un par de líneas), se escribe como sigue: primero,

$$\mathbf{E}(\mathbb{X}) = \mathbf{E}(\mathbf{m} + U\mathbb{Y}) = \mathbf{m} + U\mathbf{E}(\mathbb{Y}) = \mathbf{m}.$$

Para las varianzas y covarianzas, usando la expresión (2.11), tenemos por un lado que

$$I = \mathbf{Cov}(\mathbb{Y}) = \mathbf{E}(\mathbb{Y} \cdot \mathbb{Y}^\top),$$

y por otro,

$$\begin{aligned} \mathbf{Cov}(\mathbb{X}) &= \mathbf{E}((\mathbb{X} - \mathbf{m}) \cdot (\mathbb{X} - \mathbf{m})^\top) = \mathbf{E}((U\mathbb{Y}) \cdot (U\mathbb{Y})^\top) = \mathbf{E}(U\mathbb{Y} \cdot \mathbb{Y}^\top U^\top) \\ &= U \mathbf{E}(\mathbb{Y} \cdot \mathbb{Y}^\top) U^\top = UU^\top = V. \end{aligned}$$

¡Voilà!

D. Propiedades adicionales

Siguen ahora unos cuantos corolarios de los resultados anteriores, que registramos para su uso posterior.

Lema 3.4 (Cambio de escala y traslación) Sea \mathbb{X} un vector normal aleatorio $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$, sea B una matriz $n \times n$ no singular y sea $\mathbf{h} \in \mathbb{R}^n$.

Entonces, el vector aleatorio \mathbb{Z} dado por

$$\mathbb{Z} = \mathbf{h} + B\mathbb{X}.$$

sigue una $\mathcal{N}(\mathbf{h} + B\mathbf{m}, BV B^\top)$.

DEMOSTRACIÓN. Basta escribir, usando la proposición 3.3, y con $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$, que

$$\mathbb{Z} = \mathbf{h} + B\mathbb{X} = \mathbf{h} + B(\mathbf{m} + U\mathbb{Y}) = \mathbf{h} + B\mathbf{m} + (BU)\mathbb{Y}.$$

La prueba termina usando la parte recíproca de la misma proposición 3.3, pues $BU(BU)^\top = BUU^\top B^\top = BV B^\top$. ■

Lema 3.5 (Rotación) Si \mathbb{Y} es una variable normal (multidimensional) estándar y si O es una matriz ortogonal, entonces $\mathbb{Z} = O\mathbb{Y}$ es, asimismo, normal (multidimensional) estándar.

DEMOSTRACIÓN. Se sigue del lema 3.4 y de que $OO^\top = I$. ■

Proposición 3.6 Si $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$ y si $\mathbf{a} = (a_1, a_2, \dots, a_n)^\top \in \mathbb{R}^n$ es un vector no nulo, entonces

$$\mathbf{a}^\top \mathbb{X} = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

es una variable normal (unidimensional), que viene especificada por su esperanza

$$\mathbf{E}\left(\sum_{j=1}^n a_j X_j\right) = \sum_{j=1}^n a_j m_j = \mathbf{a}^\top \mathbf{m},$$

y su varianza

$$\mathbf{V}\left(\sum_{j=1}^n a_j X_j\right) = \sum_{j=1}^n a_j^2 v_{jj} + \sum_{1 \leq i \neq j \leq n} a_i a_j v_{ij} = \mathbf{a}^\top V \mathbf{a}.$$

Nótese que $\mathbf{V}(\sum_{j=1}^n a_j X_j) = \mathbf{a}^\top V \mathbf{a} > 0$, pues V es definida positiva y $\mathbf{a} \neq \mathbf{0}$.

DEMOSTRACIÓN. Escribiendo

$$\mathbb{X} = \mathbf{m} + U\mathbb{Y},$$

tenemos que

$$(\star) \quad \mathbf{a}^\top \mathbb{X} = \mathbf{a}^\top \mathbf{m} + (\mathbf{a}^\top U)\mathbb{Y}.$$

10 CAPÍTULO 3. LA NORMAL MULTIDIMENSIONAL Y DISTRIBUCIONES ASOCIADAS

En virtud de la proposición 2.11, se tiene que $(\mathbf{a}^\top U)\mathbb{Y}$ es una variable normal. Por tanto, por (\star) se tiene que $\mathbf{a}^\top \mathbb{X}$ es una variable normal. Como $\mathbf{E}(\mathbb{Y}) = \mathbf{0}$, es claro que

$$\mathbf{E}(\mathbf{a}^\top \mathbb{X}) = \mathbf{a}^\top \mathbf{m}.$$

Por otro lado,

$$\mathbf{V}\left(\sum_{j=1}^n a_j X_j\right) = \sum_{j=1}^n a_j^2 \mathbf{V}(X_j) + \sum_{1 \leq i \neq j \leq n} a_i a_j \text{cov}(X_i, X_j) = \sum_{j=1}^n a_j^2 V_{jj} + \sum_{1 \leq i \neq j \leq n} a_i a_j V_{ij}. \blacksquare$$



Nota 3.2.8. ¡Atención! No es cierto que las combinaciones lineales de variables normales (definidas en un mismo espacio de probabilidad, por supuesto) hayan de ser normales. Por ejemplo, si X es variable normal estándar e $Y = -X$, entonces Y también es normal estándar, pero la suma $X + Y \equiv 0$ no es una variable normal.

Corolario 3.7 Si $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$ y si $\mathbf{a} = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$ es un vector no nulo, entonces la combinación lineal

$$\mathbf{a}^\top \mathbb{Y} = a_1 Y_1 + \dots + a_n Y_n$$

es una variable normal (unidimensional), que viene especificada por los valores de su esperanza y su varianza:

$$\mathbf{E}\left(\sum_{j=1}^n a_j Y_j\right) = 0 \quad y \quad \mathbf{V}\left(\sum_{j=1}^n a_j Y_j\right) = \sum_{j=1}^n a_j^2.$$

Esta es de nuevo, la proposición 2.11 que, ¡atención!, ya hemos usado en la demostración del teorema 3.1.

E. La normal bidimensional

La normal bidimensional aparece con particular frecuencia. Pongamos que (X, Y) es normal bidimensional, y que X e Y (que son variables normales) están tipificadas: $\mathbf{E}(X) = \mathbf{E}(Y) = 0$ y $\mathbf{V}(X) = \mathbf{V}(Y) = 1$.

Recordemos que la distribución de (X, Y) está determinada por su vector de medias, que en este caso es $\mathbf{m} = \mathbf{0}$, y su matriz V , que en este caso tiene unos en la diagonal:

$$V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

El registro ρ es la covarianza de X e Y , que en este caso (tipificado) es también el coeficiente de correlación de X e Y dado por $\rho = \mathbf{E}(XY)$. Obsérvese que $\rho \in (-1, 1)$.

El determinante de V es $\det(V) = 1 - \rho^2$. La matrix V^{-1} es

$$V^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$$

3.2. La normal multidimensional

11

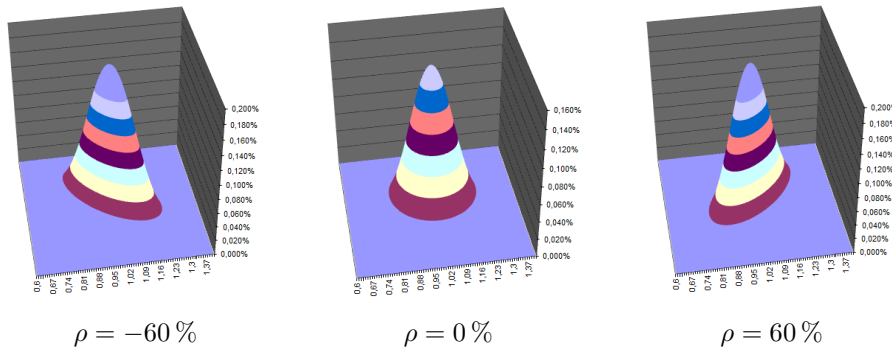
Como

$$\begin{pmatrix} x & y \end{pmatrix} V^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{1-\rho^2} (x^2 + y^2 - 2\rho xy),$$

la función de densidad $f_{(X,Y)}$ de (X,Y) viene dada por

$$(3.7) \quad f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{x^2 + y^2 - 2\rho xy}{1-\rho^2}\right).$$

Las figuras que siguen muestran el aspecto de esta función de densidad para tres valores de ρ :



Nota 3.2.9. Obsérvese que si (U,V) son variables normales estándar independientes y $\rho \in (-1,1)$, entonces

$$X = U \quad \text{e} \quad Y = \rho U + \sqrt{1-\rho^2} V$$

conforman un vector aleatorio (X,Y) de normales estándar y con correlación ρ . Alternativamente, si escribimos $\rho = \cos(2\theta)$ para un cierto ángulo $\theta \in (0, \pi)$, entonces las variables

$$X = \cos(\theta) U + \sin(\theta) V \quad \text{e} \quad Y = \cos(\theta) U - \sin(\theta) V$$

dan un vector aleatorio (X,Y) de normales estándar y con correlación ρ .

En el caso general, no tipificado, la función de densidad de la normal bidimensional es

$$(3.8) \quad f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right]\right).$$

Ahora, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, y las variables X e Y tienen correlación ρ . En la terminología general de (3.4),

$$\mathbf{m} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{y} \quad V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

F. Proyección de normal (multidimensional)

Supongamos que $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$ es normal multidimensional de dimensión n . Ya hemos visto que cada una de las coordenadas X_j de \mathbb{X} es una variable normal. Vamos a comprobar ahora que cualquier proyección de \mathbb{X} sobre un subespacio de dimensión inferior (no sólo de dimensión 1) es asimismo una variable normal multidimensional.

Para $1 \leq k \leq n$, seleccionamos un (sub)conjunto de k índices $J = \{j_1, j_2, \dots, j_k\}$, con $1 \leq j_1 < j_2 < \dots < j_k \leq n$. Consideramos el vector $\mathbf{m}_J = (m_{j_1}, \dots, m_{j_k})$ (que se obtiene al seleccionar en \mathbf{m} las coordenadas m_j con $j \in J$) y la matriz V_J , de dimensiones $k \times k$, que se obtiene de V quedándonos únicamente con filas y columnas etiquetadas con índices de J .

El objeto de interés es el vector $\mathbb{X}_J = (X_{j_1}, X_{j_2}, \dots, X_{j_k})^\top$ de dimensión k .

Proposición 3.8 *Con las notaciones anteriores, y para cualquier subconjunto de índices J de tamaño k , si $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$, entonces $\mathbb{X}_J \sim \mathcal{N}(\mathbf{m}_J, V_J)$.*

El caso $k = 1$ es el de una única coordenada, que ya ha sido probado (teorema 3.1). El caso $k = n$ es la propia \mathbb{X} ; aquí no hay nada que probar.

DEMOSTRACIÓN. La comprobación de que el vector de medias de \mathbb{X}_J es \mathbf{m}_J y que la matriz de covarianzas de las variables que conforman \mathbb{X}_J es V_J es inmediata: son, simplemente, las varianzas y covarianzas que tenían como parte del vector \mathbb{X} original. Lo que resta, pues, es comprobar que \mathbb{X}_J es una normal multidimensional de dimensión k .

Sin pérdida de generalidad¹, para esta verificación podemos suponer y suponemos que $\mathbf{m} = \mathbf{0}$, que $k = n - 1$ y que $j_k = n - 1$. Es decir, suponemos que $\mathbb{X} \sim \mathcal{N}(\mathbf{0}, V)$ y queremos ver que $(X_1, X_2, \dots, X_{n-1})^\top$ es normal. Conviene observar que si $V = I$, es decir si \mathbb{X} es normal estándar, el resultado es inmediato.

Escribamos $\mathbb{X} = U \mathbb{Y}$, donde U es no singular (y $UU^\top = V$), y donde $\mathbb{Y} \sim \mathcal{N}(\mathbf{0}, I)$.

Por la proposición 3.6, cualquier combinación lineal

$$a_1 X_1 + \dots + a_{n-1} X_{n-1}$$

(en la que no todos los coeficientes sean nulos) es una variable normal. Esto nos dice, en particular, que las variables X_1, X_2, \dots, X_{n-1} son linealmente independientes².

Sean Z_1, Z_2, \dots, Z_{n-1} obtenidas de X_1, X_2, \dots, X_{n-1} mediante ortonormalización de Gram-Schmidt: es decir, con el mismo argumento que en Gram-Schmidt logramos que

¹Es decir, si comprobamos este caso, tenemos el caso general. Es evidente que podemos suponer que el vector de medias es nulo. Y si comprobamos que la proyección sobre una dimensión menos es una normal, bastará iterar el argumento para proyecciones en menos dimensiones. Por último, escoger la última coordenada para proyectar obedece, simplemente, a razones de estética tipográfica.

²La independencia lineal de variables aleatorias se define de manera análoga a la del álgebra lineal: la única combinación lineal de las variables que la variable aleatoria nula es la que tiene todos los coeficientes 0.

- cada Z_j sea combinación lineal de X_1, X_2, \dots, X_j ;
- para cada $1 \leq j \leq n-1$ se tenga que $\mathbf{E}(Z_j) = 0$ y que $\mathbf{E}(Z_j^2) = \mathbf{V}(Z_j) = 1$;
- si $1 \leq i < j \leq n-1$, entonces $\mathbf{E}(Z_i Z_j) = 0$.

Nótese que como cada Z_i es combinación lineal de Y_1, Y_2, \dots, Y_n , y como éstas son variables independientes, del corolario 3.7 se deduce que cada Z_j es una variable normal. Escribamos matricialmente esas combinaciones lineales:

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix} = M \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

para cierta matriz M de dimensiones $(n-1) \times n$, cuyas filas son ortonormales, pues por construcción $\mathbf{E}(Z_i Z_j) = \delta_{ij}$, para $1 \leq i, j \leq n-1$. (La notación δ_{ij} es la delta de Kronecker: δ_{ij} vale 1 si $i = j$ y 0 en el resto de los casos).

Añadimos a M una última fila (de norma 1 y perpendicular a las anteriores) para formar una matriz ortogonal \widehat{M} . Obtenemos

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_{n-1} \\ Z_n \end{pmatrix} = \widehat{M} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

donde surge una variable aleatoria adicional Z_n . Por el lema 3.5, como \widehat{M} es una matriz ortogonal, el vector $(Z_1, \dots, Z_n)^\top$ es un vector normal multidimensional estándar, es decir, las Z_j son normales (unidimensionales) estándar independientes.

Y, por consiguiente, y esto es clave, $(Z_1, Z_2, \dots, Z_{n-1})^\top \sim \mathcal{N}(\mathbf{0}, I_{n-1})$.

Sea A la matriz $(n-1) \times (n-1)$ que nos da la ortonormalización Gram-Schmidt:

$$A \begin{pmatrix} X_1 \\ \vdots \\ X_{n-1} \end{pmatrix} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix}.$$

La matriz A es triangular inferior y, sobre todo, no singular, porque los Z_j son linealmente independientes.

Por consiguiente,

$$A^{-1} \begin{pmatrix} Z_1 \\ \vdots \\ Z_{n-1} \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_{n-1} \end{pmatrix},$$

y por el lema 3.4, por ejemplo, obtenemos, como conclusión, que $(X_1, \dots, X_{n-1})^\top$ es un vector normal. ■

Como complemento, registramos el siguiente:

Corolario 3.9 *Supongamos que el vector aleatorio \mathbb{X} sigue una normal multidimensional de dimensión n , con parámetros $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$. Sea $1 \leq k \leq n$, y consideremos una matriz A de dimensiones $k \times n$ y de rango máximo. Definimos el vector \mathbb{Z} (de dimensión k) como $\mathbb{Z} = A\mathbb{X}$. Entonces, $\mathbb{Z} = \mathcal{N}(A\mathbf{m}, AVA^T)$.*

Obsérvese que el caso $k = 1$ de este resultado es la proposición 3.6, que analiza combinaciones lineales de las componentes de \mathbb{X} , y que el caso $k = n$ es el lema 3.4 sobre transformaciones lineales de vectores normales.

Además, la proposición 3.8 es también un caso particular, en el que la matriz A es simplemente la que representa la proyección sobre las coordenadas correspondientes al conjunto de k índices prefijado.

DEMOSTRACIÓN. La comprobación de los parámetros es directa:

$$\begin{aligned} \mathbf{E}(\mathbb{Z}) &= \mathbf{E}(A\mathbb{X}) = A\mathbf{E}(\mathbb{X}) = A\mathbf{m}, \\ \text{Cov}(\mathbb{Z}) &= \mathbf{E}[(\mathbb{Z} - A\mathbf{m}) \cdot (\mathbb{Z} - A\mathbf{m})^T] = \mathbf{E}[(A(\mathbb{X} - \mathbf{m})) \cdot (A(\mathbb{X} - \mathbf{m}))^T] \\ &= A\mathbf{E}[(\mathbb{X} - \mathbf{m}) \cdot (\mathbb{X} - \mathbf{m})^T]A^T = A\text{Cov}(\mathbb{X})A^T = AVA^T. \end{aligned}$$

Queda, pues, verificar que \mathbb{Z} es un vector normal de dimensión k . Veamos. Tomamos la matriz A y la ampliamos, añadiendo $n - k$ filas al final de la misma, de manera que la matriz ampliada, que llamaremos \tilde{A} , sea invertible (usamos aquí que A es de rango máximo). Tenemos entonces que

$$\tilde{\mathbb{Z}} = \tilde{A}\mathbb{X},$$

donde $\tilde{\mathbb{Z}}$ es un vector aleatorio de dimensión n cuyas primeras k componentes coinciden con las de \mathbb{Z} . Por el lema 3.4, sabemos que $\tilde{\mathbb{Z}}$ es un vector normal, pues \mathbb{X} lo es, y \tilde{A} es invertible. Pero ahora \mathbb{Z} se obtiene a partir de $\tilde{\mathbb{Z}}$ por una proyección (en este caso, sobre las primeras k coordenadas). Usando la proposición 3.8, concluimos que \mathbb{Z} es un vector normal. ■

EJEMPLO 3.2.1. Sean Y, X_1, X_2, \dots, X_n variables normales estándar independientes. Para $\rho \in (0, 1)$, definimos el vector aleatorio $\mathbb{Z} = (Z_1, \dots, Z_n)^T$ mediante

$$(\star) \quad Z_j = \sqrt{\rho}Y + \sqrt{1-\rho}X_j$$

para cada $j = 1, \dots, n$.

El vector $\mathbb{Z} = (Z_1, \dots, Z_n)^T$ se puede escribir como

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} \sqrt{\rho} & \sqrt{1-\rho} & 0 & 0 & \cdots & 0 \\ \sqrt{\rho} & 0 & \sqrt{1-\rho} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \sqrt{\rho} & 0 & 0 & 0 & \cdots & \sqrt{1-\rho} \end{pmatrix} \begin{pmatrix} Y \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix},$$

donde la matriz, a la que llamaremos A , es de dimensiones $n \times (n+1)$.

3.2. La normal multidimensional

15

Aplicando el corolario 3.9, con $\mathbb{X} = (Y, X_1, X_2, \dots, X_n) \sim \mathcal{N}(\mathbf{0}, I_{n+1})$, se deduce que el vector \mathbb{Z} sigue una normal $\mathcal{N}(A\mathbf{0}, AIA^T) = \mathcal{N}(\mathbf{0}, V)$, donde la matriz V tiene entradas

$$v_{ij} = \begin{cases} 1, & \text{si } i = j, \\ \rho, & \text{si } i \neq j. \end{cases}$$

Para comprobar este hecho podemos calcular el producto matricial AA^T , o alternativamente, calcular las medias, varianzas y covarianzas “a mano”, usando (\star) , la independencia de las variables Y, X_1, \dots, X_n , y el que son normales estándar. Por un lado,

$$\mathbf{E}(Z_j) = \mathbf{E}(\sqrt{\rho}Y + \sqrt{1-\rho}X_j) = \sqrt{\rho}\mathbf{E}(Y) + \sqrt{1-\rho}\mathbf{E}(X_j) = 0,$$

$$\mathbf{V}(Z_j) = \mathbf{V}(\sqrt{\rho}Y + \sqrt{1-\rho}X_j) = \rho\mathbf{V}(Y) + (1-\rho)\mathbf{V}(X_j) = \rho + (1-\rho) = 1,$$

mientras que

$$\begin{aligned} \mathbf{E}(Z_i Z_j) &= \mathbf{E}((\sqrt{\rho}Y + \sqrt{1-\rho}X_i)(\sqrt{\rho}Y + \sqrt{1-\rho}X_j)) \\ &= \rho\mathbf{E}(Y^2) + \sqrt{\rho(1-\rho)}\mathbf{E}(Y)\mathbf{E}(X_i) + \sqrt{\rho(1-\rho)}\mathbf{E}(Y)\mathbf{E}(X_j) + (1-\rho)\mathbf{E}(X_i)\mathbf{E}(X_j) \\ &= \rho \end{aligned}$$

Lo interesante de la transformación descrita en (\star) es que las variables Z_1, \dots, Z_n resultantes son normales estándar, conjuntamente son una normal multidimensional, y además la correlación entre cualesquiera dos de ellas *es fija* e igual ρ (un valor positivo). ♣

Nota 3.2.10. Resumimos en esta nota los resultados obtenidos hasta aquí sobre combinaciones lineales de variables normales (que conjuntamente son una normal multidimensional), junto con algún caso adicional que no se ha contemplado. Partimos de un vector $\mathbb{X} = (X_1, \dots, X_n)^T$ que sigue una normal n -dimensional con parámetros $\mathbb{X} \sim \mathcal{N}(\mathbf{m}, V)$. Consideramos una matriz A de dimensiones $k \times n$ y definimos el vector \mathbb{Z} (de dimensiones $k \times 1$) como $\mathbb{Z} = A\mathbb{X}$.

En esta situación se tiene que el vector de medias de \mathbb{Z} es $\mathbf{E}(\mathbb{Z}) = A\mathbf{m}$ y la matriz de varianzas/covarianzas es $\mathbf{Cov}(\mathbb{Z}) = AVA^T$. Además, salvo en el caso en el que alguna fila de A sea nula, tenemos que todas las coordenadas Z_j son variables aleatorias normales. Todo esto está en la proposición 3.6.

En cuanto a la posible normalidad del vector \mathbb{Z} , hemos visto (corolario 3.9) que si $1 \leq k \leq n$ y si A es de rango máximo (k , en este caso), entonces \mathbb{Z} es un vector normal.

Quedan dos casos por analizar: cuando $1 \leq k \leq n$ y A no es de rango máximo, y cuando $k > n$. Obsérvese que, en ambas situaciones, el rango de A es menor que k . En cualquiera de estos casos, \mathbb{Z} *nunca* es un vector normal.

La razón es que, como el rango de A es menor que k , existe una combinación lineal (con coeficientes no todos nulos) de las variables Z_1, \dots, Z_k que es igual a 0. La proposición 3.6 impide entonces que $\mathbb{Z} = (Z_1, \dots, Z_k)^T$ sea un vector normal.

3.3. Distribuciones asociadas a la normal multidimensional

Describimos ahora unas cuantas familias de distribuciones de probabilidad (unidimensionales) asociadas a la normal multidimensional (o a colecciones de normales independientes), que usaremos a menudo en los capítulos venideros.

3.3.1. La distribución $\chi^2(n)$ (chi-cuadrado con n grados de libertad)

Decimos que Z es una **variable χ^2 con n grados de libertad** si

$$Z = X_1^2 + \cdots + X_n^2,$$

donde X_1, \dots, X_n son variables normales estándar independientes. Escribiremos que $Z \sim \chi_n^2$. El símbolo χ^2 se lee «chi cuadrado» o «ji cuadrado».

Obsérvese que la variable Z toma valores en $[0, +\infty)$.

A. Media y varianza

- Media:

$$(3.9) \quad \boxed{\mathbf{E}(Z) = n}$$

Basta observar que $\mathbf{E}(X_j^2) = 1$, para cada X_j .

- Varianza:

$$(3.10) \quad \boxed{\mathbf{V}(Z) = 2n}$$

Veamos. Si X es una normal estándar, resulta que $\mathbf{E}(X^4) = 3$ (nota 2.3.4), y por tanto $\mathbf{V}(X^2) = \mathbf{E}(X^4) - \mathbf{E}(X^2)^2 = 3 - 1 = 2$. Ahora argumentamos:

- Como Z es una suma de variables independientes, se tiene que

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

- Alternativamente, desarrollando el cuadrado,

$$\begin{aligned} \mathbf{E}(Z^2) &= \mathbf{E}\left(\left(\sum_{j=1}^n X_j^2\right)^2\right) = \mathbf{E}\left(\sum_{j=1}^n X_j^4 + \sum_{i \neq j} X_i^2 X_j^2\right) \\ &= n\mathbf{E}(X^4) + n(n-1)\mathbf{E}(X^2)^2 = 3n + n(n-1) = n(n+2), \end{aligned}$$

$$\text{y, por tanto, } \mathbf{V}(Z) = \mathbf{E}(Z^2) - \mathbf{E}(Z)^2 = n(n+2) - n^2 = 2n.$$

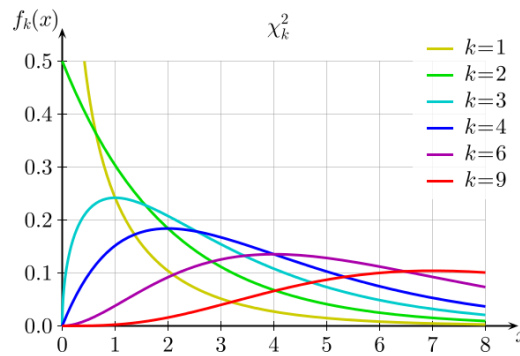
En realidad ya obtuvimos estos dos momentos en la sección 2.3.4, aprovechando los cálculos generales para las distribuciones Gamma; véase el apartado siguiente.

B. Función de densidad

La función de densidad de $Z \sim \chi_n^2$ viene dada para $x > 0$ por

$$(3.11) \quad f_Z(x) = c_n x^{n/2-1} e^{-x/2} = \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2}$$

(para $x \leq 0$ es $f_Z(x) = 0$). El valor de la constante c_n se obtiene de exigir que $\int f_Z = 1$ (recuérdese la proposición 2.14).



La comprobación de (3.11) es directa, usando las propiedades de la familia de variables Gamma: si X_1 es normal estándar, entonces, por el ejemplo 2.2.4,

$$f_{X_1^2}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-x/2}.$$

Comparando con la función de densidad (2.22) de las variables Gamma, observamos que $X_1^2 \sim \text{GAMMA}(1/2, 1/2)$.

Por la proposición 2.15 sobre suma de variables Gamma independientes,

una variable χ_n^2 es una variable $\text{GAMMA}(1/2, n/2)$,

lo que nos da directamente (3.11) usando de nuevo la fórmula (2.22).

Nota 3.3.1. (Re)comprobación por inducción de la expresión para la densidad f_{Z_n} de la χ_n^2 . Pongamos $Z_n = X_1^2 + \dots + X_n^2$, donde las X_j son normales estándar independientes. Damos siempre las f_{Z_n} solo en $x \in (0, +\infty)$ (fuera valen 0). Para $n = 1$ ya lo hemos visto (ejemplo 2.2.4): $f_{Z_1}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-x/2} = c_1 x^{1/2-1} e^{-x/2}$.

De n a $n+1$:

$$\begin{aligned} f_{Z_{n+1}}(x) &= \int_0^\infty f_{Z_n}(y) f_{Z_1}(x-y) dy = \int_0^x f_{Z_n}(y) f_{Z_1}(x-y) dy \\ &= c_n c_1 \int_0^x y^{n/2-1} e^{-y/2} (x-y)^{-1/2} e^{-(x-y)/2} dy \\ &\stackrel{[\text{cambio } y=xu]}{=} c_n c_1 e^{-x/2} x^{n/2+1} \int_0^1 u^{n/2-1} (1-u)^{-1/2} du \\ &= c_n c_1 \underbrace{\left(\int_0^1 u^{n/2-1} (1-u)^{-1/2} du \right)}_{=c_{n+1}} e^{-x/2} x^{n/2+1}. \end{aligned}$$

Más adelante necesitaremos los valores de las esperanzas de ciertas transformaciones adicionales de una chi cuadrado:

Lema 3.10 Sea $Z \sim \chi_n^2$ y sea $\alpha \in \mathbb{R}$. Entonces

$$(3.12) \quad \mathbf{E}(Z^\alpha) = 2^\alpha \frac{\Gamma(n/2 + \alpha)}{\Gamma(n/2)}$$

siempre que $n/2 + \alpha > 0$.

En particular, para todo $n \geq 3$,

$$\mathbf{E}(1/Z) = \frac{1}{n-2}.$$

(Para $n = 1, 2$ la esperanza $\mathbf{E}(1/Z)$ es infinita).

Y para todo $n \geq 1$,

$$\mathbf{E}(\sqrt{Z}) = \sqrt{2} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}.$$

DEMOSTRACIÓN. Usando la expresión (3.11) de la función de densidad de la χ_n^2 y las propiedades de la función Gamma de Euler,

$$\begin{aligned} \mathbf{E}(Z^\alpha) &= \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} \int_0^\infty t^{\alpha+n/2-1} e^{-t/2} dt \\ &= \frac{1}{\Gamma(n/2)} \frac{2^{\alpha+n/2-1}}{2^{n/2}} 2 \int_0^\infty u^{\alpha+n/2-1} e^{-u} du = 2^\alpha \frac{\Gamma(n/2 + \alpha)}{\Gamma(n/2)}. \quad \blacksquare \end{aligned}$$



Nota 3.3.2. Si $Z \sim \chi_n^2$, entonces la variable \sqrt{Z} recoge el resultado de sortear un punto en \mathbb{R}^n , de manera que sus coordenadas se sortean con normales estándar independientes, y medir su *distancia al origen*. El lema 3.10 nos dice que, en media, esa distancia es $\sqrt{2} \Gamma((n+1)/2)/\Gamma(n/2)$. Por ejemplo, si $n = 2$, la distancia media al origen de los puntos generados con este procedimiento es $\sqrt{\pi}/2 \approx 1.25$. En tres dimensiones, esa distancia media es $2\sqrt{2/\pi} \approx 1.596$. Usando la aproximación de Stirling, se puede comprobar que $\mathbf{E}(\sqrt{Z}) \sim \sqrt{n}$ cuando $n \rightarrow \infty$. Véase el ejercicio 3.8.

3.3.2. Distribución $F_{n,m}$ con n y m grados de libertad

Para enteros $n, m \geq 1$, del cociente

$$Z = \frac{(X_1^2 + X_2^2 + \cdots + X_n^2)/n}{(Y_1^2 + Y_2^2 + \cdots + Y_m^2)/m},$$

donde $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ son normales estándar (completamente) independientes, se dice que sigue la **distribución** $F_{n,m}$. Escribiremos que $Z \sim F_{n,m}$. Estas distribuciones F se conocen como de Fisher, de Snedecor, o de Fisher–Snedecor.

Obsérvese que Z toma valores sólo en $(0, +\infty)$, y que

$$Z = \frac{U_n/n}{V_m/m}$$

donde U_n y V_m son independientes y $U_n \sim \chi_n^2$ y $V_m \sim \chi_m^2$.

Obsérvese que si $Z \sim F_{n,m}$ entonces $1/Z \sim F_{m,n}$.

A. Media y varianza

Sea $Z \sim F_{n,m}$.

- Si $m \geq 3$, la esperanza de Z es

$$\mathbf{E}(Z) = \frac{m}{m-2},$$

pues por (3.9) y el lema 3.10, $\mathbf{E}(Z) = \mathbf{E}(U_n/n)\mathbf{E}(m/V_m) = \mathbf{E}(m/V_m) = m/(m-2)$.

Si $m = 1, 2$, entonces $\mathbf{E}(Z) = +\infty$.

- Si $m \geq 5$, entonces

$$\mathbf{V}(Z) = \frac{2(m+n-2)m^2}{n(m-4)(m-2)^2},$$

que se obtiene usando la expresión anterior para $\mathbf{E}(Z)$, junto con que $\mathbf{E}(U_n^2) = n(n+2)$ y que $\mathbf{E}(1/V_m^2) = 1/((m-2)(m-4))$.

Si $m \leq 4$, entonces $\mathbf{V}(Z) = +\infty$.

B. Función de densidad

La función de densidad de una variable $Z \sim F(n, m)$ es, para $x > 0$,

$$f_Z(x) = \frac{n}{m} \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \frac{(nx/m)^{n/2-1}}{(1+(nx/m))^{(n+m)/2}}.$$



Nota 3.3.3. Esta fórmula se sigue de que si U, V son independientes y $U \sim \chi_n^2$ y $V \sim \chi_m^2$, entonces, para $x > 0$

$$f_{U/V}(x) = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \frac{x^{n/2-1}}{(1+x)^{(n+m)/2}}.$$

Para ver esto, consideremos el cambio de variables $A = U/V$ y $B = V$:

$$f_{(A,B)}(a, b) \frac{1}{b} = f_U(u) f_V(v),$$

es decir,

$$f_{(A,B)}(a, b) = b f_U(ab) f_V(b).$$

Por tanto,

$$\begin{aligned} f_A(a) &= \int_0^\infty b f_U(ab) f_V(b) db = c_n c_m a^{n/2-1} \int_0^\infty b^{(n+m)/2-1} e^{-b(1+a)/2} db \\ &= c_n c_m \frac{a^{n/2-1}}{(1+a)^{(n+m)/2}} \int_0^\infty y^{(n+m)/2-1} e^{-y/2} dy \\ &= \frac{c_n c_m}{c_{n+m}} \frac{a^{n/2-1}}{(1+a)^{(n+m)/2}} = \frac{\Gamma((n+m)/2)}{\Gamma(n/2)\Gamma(m/2)} \frac{a^{n/2-1}}{(1+a)^{(n+m)/2}}. \end{aligned}$$

3.3.3. Distribución t_n de Student

Fijamos $n \geq 2$. Sean Y y Z_n dos variables aleatorias independientes, Y con distribución normal estándar y Z_n con distribución χ_n^2 . Se dice que el cociente

$$t_n := \frac{Y}{\sqrt{Z_n/n}}$$

sigue una **distribución t de Student con n grados de libertad**. Escribiremos que $t_n \sim \text{STU}(n)$.

Contraviniendo la regla de denotar las variables aleatorias con mayúsculas, pero siguiendo la tradición, usamos t_n para denotar variables de Student.



Nota 3.3.4. Obsérvese que si $t_n \sim \text{STU}(n)$, entonces $t_n^2 \sim F(1, n)$.

A. Media y varianza

Para la **esperanza** de t_n : como numerador y denominador son independientes y $\mathbf{E}(Y) = 0$ tenemos que

$$\mathbf{E}(t_n) = 0.$$

Para la **varianza**: si $n \geq 3$,

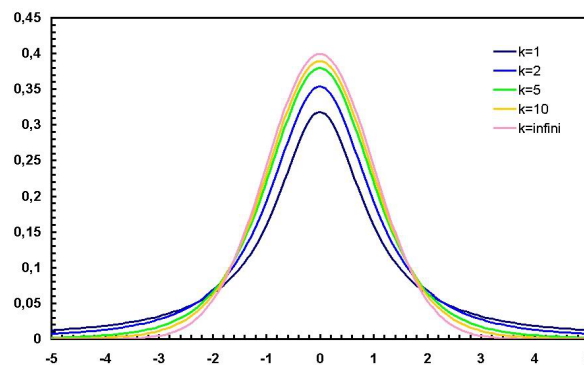
$$\mathbf{V}(t_n) = \mathbf{E}(t_n^2) = n \mathbf{E}(Y^2) \mathbf{E}(1/Z_n) = n \mathbf{E}(1/Z_n) = \frac{n}{n-2},$$

apelando al lema 3.10. Si $n = 1, 2$, la varianza $\mathbf{V}(t_2) = +\infty$.

B. Función de densidad

La función de densidad de una $t_n \sim \text{STU}(n)$ es

$$(3.13) \quad f_{t_n}(t) = D_n \left(\frac{1}{1+t^2/n} \right)^{(n+1)/2}, \quad \text{donde } D_n = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}.$$





Nota 3.3.5. He aquí el cálculo de esta función de densidad. La función de densidad f conjunta de (Y, Z_n) es

$$f(y, z) = \left(\frac{1}{\sqrt{2\pi}} e^{-y^2/2} \right) \left(\frac{1}{\Gamma(n/2)} \left(\frac{1}{2} \right)^{n/2} z^{n/2-1} e^{-z/2} \right) = C_n e^{-y^2/2} z^{n/2-1} e^{-z/2},$$

donde $C_n = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(n/2)} \left(\frac{1}{2} \right)^{n/2}$.

Hacemos el cambio de variables (y, z) a (w, t)

$$\begin{cases} w = z, \\ t = y/\sqrt{z/n}. \end{cases}$$

Como el jacobiano de esta transformación es $|\partial(w, t)/\partial(y, z)| = 1/\sqrt{z/n}$, tenemos que la función de densidad conjunta de W y t_n viene dada por

$$f_{W, t_n}(w, t) \left| \frac{\partial(w, t)}{\partial(y, z)} \right| = f(y, z),$$

que despejando la y y z en términos de w y t dice

$$f_{W, t_n}(w, t) = (C_n/\sqrt{n}) w^{(n-1)/2} e^{-\frac{w}{2}(1+t^2/n)}.$$

Si fijamos t e integramos respecto de w (haciendo el cambio $r = \frac{w}{2}(1+t^2/n)$), para así conseguir la densidad marginal de t_n , obtenemos que

$$f_{t_n}(t) = \left[(C_n/\sqrt{n}) \Gamma\left(\frac{n+1}{2}\right) 2^{(n+1)/2} \right] \left(\frac{1}{1+t^2/n} \right)^{(n+1)/2} = D_n \left(\frac{1}{1+t^2/n} \right)^{(n+1)/2},$$

donde $D_n = \frac{1}{\sqrt{n\pi}} \Gamma\left(\frac{n+1}{2}\right)/\Gamma\left(\frac{n}{2}\right)$.