Normal Multivariada

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Un repaso de algunas nociones de Álgebra Lineal

Transpuesta de una matriz

Dada una matriz $A \in \mathbb{R}^{m\times n}$, su matriz transpuesta $A^t \in \mathbb{R}^{n\times m}$ se obtiene intercambiando las filas y las columnas.

$$A = \left(\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \end{array} \right) \in \mathbb{R}^{2\times 3} \Rightarrow A^t \left(\begin{array}{cc} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{array} \right) \in \mathbb{R}^{3\times 2}$$

La operación de transponer tiene algunas propiedades interesantes:

$$(A+B)^t = A^t + B^t, \quad (A\cdot B)^t = B^t \cdot A^t, \quad \det(A^t) = \det(A)$$

Vamos a escribir los vectores como columnas. El producto escalar lo podemos escribir así:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \in \mathbb{R}^n \Rightarrow \langle x, y \rangle = x^t \cdot y$$

Matrices Simétricas y Ortogonales

- $A \subset \mathbb{R}^{n\times n}$ se dice simétrica si $A^t = A$.
- $P \subset \mathbb{R}^{n\times n}$ se dice ortogonal si $P^t \cdot P = P \cdot P^t = I$.

Teorema

Si $A \subset \mathbb{R}^{n\times n}$ es simétrica, entonces existe P ortogonal tal que

$$D = P^t A P = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

es diagonal, siendo los $\lambda_k \in \mathbb{R}$ los autovalores de la matriz A.

Formas Cuadráticas

Una forma cuadrática en las variables $x_1, x_2, \ldots x_n$ es un polinomio homogéneo de segundo grado en ellas, por ejemplo

$$q_2(x_1, x_2) = x_1^2 + 2x_2^2 - 6x_1x_2$$

$$q_3(x_1, x_2, x_3) = x_1^2 + 4x_2^2 - x_3^2 - 6x_1x_2 - 8x_1x_3$$

son formas cuadráticas en 2 y 3 variables respectivamente.

Dada una matriz simétrica $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, podemos asociarle la forma cuadrática en n variables

$$q_A(x) = \langle Ax, x \rangle = x^t \cdot A \cdot x = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

Formas Cuadráticas (2)

Recíprocamente, cada forma cuadrática está asociada a una única matriz simétrica. Veamos cómo:

$$\begin{aligned} q_2(x_1, x_2) &= x_1^2 + 2x_2^2 - 6x_1x_2 \\ &= x_1^2 + 2x_2^2 - 3x_1x_2 - 3x_2x_1 \end{aligned}$$

$$\Rightarrow q_2 = q_A \text{ con } A = \begin{pmatrix} 1 & -3 \\ -3 & 2 \end{pmatrix}$$

Similarmente

$$q_3(x_1, x_2, x_3) = x_1^2 + 4x_2^2 - x_3^2 - 6x_1x_2 - 8x_1x_3 = q_B(x)$$

con

$$B = \begin{pmatrix} 1 & -3 & -4 \\ -3 & 4 & 0 \\ -4 & 0 & -1 \end{pmatrix}$$

Formas Cuadráticas Definidas Positivas

- Una forma cuadrática $q_A(x)$ y la correspondiente matriz simétrica A se dicen semi-definidas positivas si

$$q_A(x) \geq 0 \text{ para todo } x \in \mathbb{R}^n$$

Ejemplo:

$$q_A(x) = x_1^2 - 2x_1x_2 + x_2^2 = (x_1 - x_2)^2, \quad A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

es semifinida positiva.

- Una forma cuadrática $q_A(x)$ y la correspondiente matriz simétrica A se dicen definidas positivas si

$$q_A(x) > 0 \text{ para todo } x \neq \vec{0} \in \mathbb{R}^n$$

Ejemplo

$$q_A(x) = x_1^2 - x_1x_2 + x_2^2 = \left(x_1 - \frac{1}{2}x_2\right)^2 + \frac{3}{4}x_2^2 \quad A = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$$

Formas Cuadráticas Definidas Positivas (2)

Teorema

Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica, (λ_k) sus autovalores, y q_A su forma cuadrática asociada.

- *A es semi-definida positiva si y sólo si $\lambda_k \geq 0$ para todo k.*
- *A es definida positiva si y sólo si $\lambda_k > 0$ para todo k.*

Corolario

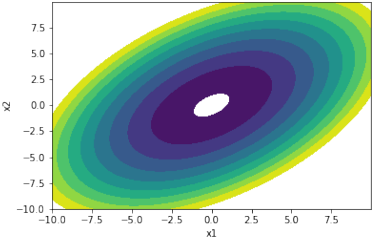
- *Si A es semi-definida positiva, $\det(A) \geq 0$.*
- *Si A es definida positiva, $\det(A) > 0$.*

Esto es inmediato, pues $\det(A) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$.

Curvas de nivel

En $n = 2$, las curvas de nivel de una forma cuadrática definida positiva son elipses. Veámoslo en el ejemplo

$$q_A(x) = x_1^2 - x_1 x_2 + x_2^2 = \left(x_1 - \frac{1}{2}x_2\right)^2 + \frac{3}{4}x_2^2$$



En $n = 3$ las superficies de nivel de una forma cuadrática definida positiva serán elipsoides.

Parte II

Espeanza de un vector aleatorio y Matriz de covariancias

Espeanza de un vector aleatorio y Matriz de covariancias

Consideramos un vector aleatorio X. Su esperanza se define componente a componente, y es un nuevo vector (no aleatorio)

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix} \in \mathbb{R}^n \Rightarrow \mu_X = E[X] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \dots \\ E[X_n] \end{pmatrix} \in \mathbb{R}^n$$

Definimos su matriz de covariancias $\Sigma = \Sigma_X = \text{Cov}$ por $\Sigma_{i,j} = \text{Cov}(X_i, X_j)$.

$$\Sigma_X = \text{Cov}(X) = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(2_1, X_2) & \dots & \text{Cov}(X_2, X_n) \\ \dots & & & \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Cov}(X_n, X_n) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Notamos que es una matriz simétrica. También podemos escribir:

$$\text{Cov}(X) = E[(X - \mu_X) \cdot (X - \mu_X)^t]$$

Espeanza de un vector aleatorio y Matriz de covariancias (2)

Notamos que en la diagonal de la matriz de covariancias $\text{Cov}(X)$ aparecen las variancias

$$\sigma_{X_i}^2 = \text{Cov}(X_i, X_i) = \text{Var}(X_i)$$

Otra observación interesante es que si las componentes del vector X son independientes, entonces serán no correlacionadas

$$\text{Cov}(X_i, X_j) = 0 \text{ si } i \neq j$$

por lo que la matriz $\text{Cov}(X)$ será diagonal.

Un ejemplo que ya vivmos: Distribución normal multivariada estándar

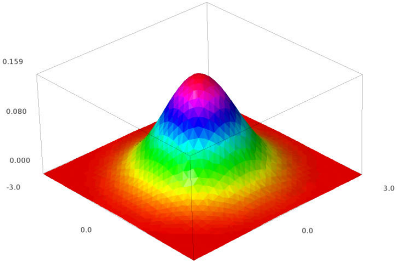
Si X es un vector con componentes $X_i \sim N(0,1)$ independientes, su densidad conjunta vendrá dada por

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

Tenemos

$$E[X] = \vec{0}, \quad \mathrm{Cov}(X) = I \text{ (matriz identidad)}$$

Por ejemplo si $n = 2$ tenemos la distribución normal bivariada estándar



Efecto de un cambio lineal sobre la esperanza y la matriz de covariancias

Si hacemos un cambio lineal $Y = A \cdot X + b$ donde ahora $b \in \mathbb{R}^n$ es un vector no aleatorio, y $A \in \mathbb{R}^{n \times n}$ es una matriz no aleatoria, encontramos que:

$$E[Y] = AE[X] + E[b] = AE[X] + b = A \cdot \mu_X + b$$

mientras que:

$$\begin{aligned}
\mathrm{Cov}[Y] &= E[(X - \mu_Y) \cdot (Y - \mu_Y)^t] \\
&= E[((A \cdot \mu_X + b)) - (A \cdot \mu_X + b)) \cdot (A \cdot X + b - (A \cdot \mu_X + b))^t] \\
&= E[(A \cdot (X - \mu)) \cdot (A \cdot (X - \mu_X))^t] \\
&= E[A \cdot (X - \mu)) \cdot (X - \mu_X)^t \cdot A^t] \\
&= A \cdot E[(X - \mu)) \cdot (X - \mu_X)^t] \cdot A^t \\
&= A \cdot \mathrm{Cov}(X) \cdot A^t
\end{aligned}$$

o sea:

$$\Sigma_Y = A \cdot \Sigma_X \cdot A^t$$

La matriz de covariancias es siempre definida positiva

Teorema

- *Si* X *es un vector aleatorio n-dimensional, su matriz se covarinncias* $\text{Cov}(X)$ *es una matriz simétrica semi-definida positiva.*
- *Además, es definida positiva, salvo en el caso en que la distribución del vector* X *está concentrada en un hiperplano afin* H*, es decir cuando existe un hiperplano afín*

$$H = \{x \in \mathbb{R}^n : \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 + \ldots + \alpha_n \cdot x_n = b\}$$

tal que

$$P\{X \in H\} = 1$$

Demostración

Sea $\mu_X = E[X]$. Entonces ya observamos que $\text{Cov}(X - \mu_X) = \text{Cov}(X)$. Por lo que podemos suponer sin pérdida de generalidad que $\mu_X = \vec{0}$.
Entonces $\text{Cov}(X) = E[X \cdot X^t]$. Consideremos la expresión

$$q(\alpha) = E[(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n \cdot X_n)^2] \quad \alpha \in \mathbb{R}^n$$

Notamos que $q(\alpha) \geq 0$ y que

$$q(\alpha) = E\left[\sum_{i,j=1}^{n} X_i X_j \alpha_i \alpha_j\right] = \sum_{i,j=1}^{n} E[X_i X_j] \alpha_i \alpha_j = \sum_{i,j=1}^{n} \text{Cov}(X_i, X_j) \alpha_i \alpha_j$$

Entonces $q(\alpha)$ es la forma cuadrática asociada a la matriz $\text{Cov}(X)$. Deducimos que $\text{Cov}(X)$ es semidefinida positiva. Finalmente si para algún $\alpha \in \mathbb{R}^n$,

$$q(\alpha) = 0 \Rightarrow \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n \cdot X_n = 0 \text{ con probabilidad } 1$$

y esto dice que la distribución del vector X está concentrada en un hiperplano.

Parte III

Distribución normal multivariada en general

Planteo del problema

En el ejercicio 25 de la práctica 7 se plantea el siguiente problema, supongamos que X es un vector con distribución normal multivariada estándar como vimos antes, hacemos un cambio lineal de variable

$$Y = A \cdot X + b$$

donde A es una matriz no singular. ¿cuál es la distribución de Y?

Esta distribución se llamará distribución normal multivariada, y es sumamente útil en las aplicaciones a la estadística. Generalizará a n dimensiones la distribución normal.

En la práctica se considera el caso especial en que $n = 2$ (distribución normal bivariada), pero las cuentas son igualmente fáciles en general (con la notación adecuada).

Algunas observaciones

- Para simplificar, vamos a considerar primero el caso especial donde $b = \vec{0}$. (distribución normal multivariada centrada en el origen)
- Ya vimos que entonces la esperanza y varianza de Y serán

$$\mu_Y = A \cdot \mu_X = \vec{0}$$

$$\Sigma_Y = A \cdot \Sigma_X \cdot A^t = A \cdot A^t$$

[Ojo: ¡en esta expresión el orden importa, no siempre una matriz A conmuta con su transpuesta A^t !]
Esta es una matriz matriz simétrica definida positiva asociada a la forma cuadrática $q(x) = \|A^t \cdot x\|^2$ pues

$$q(x) = (A^t \cdot x)^t \cdot (A^t \cdot x) = x^t \cdot (A \cdot A^t) \cdot x$$

y como A es no singular, A^t también con lo que $q(x) = 0$ si y sólo si $x = 0$.

Fórmula de la densidad conjunta en la normal multivariada

Usando el teorema de cambio de variable que vimos en la clase 11 con $y = \varphi(x) = A \cdot x$, $\varphi^{-1}(y) = A^{-1} \cdot x$, tenemos que la densidad conjunta de Y se relaciona con la de X por

$$\begin{aligned} f_Y(y) &= f_X(A^{-1}y) \cdot |det(A^{-1})| \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\|A^{-1}y\|^2/2} \cdot |det(A^{-1})| \end{aligned}$$

Vamos a reescribir esta fórmula en términos de la matriz de covariancias

$$\Sigma = \Sigma_Y = A \cdot A^t$$

Notamos que

$$\|A^{-1}y\|^2 = (A^{-1}y)^t \cdot (A^{-1}y) = y^t \cdot (A^{-1})^t \cdot A^{-1} \cdot y$$

Como por otra parte:

$$\Sigma^{-1} = (A \cdot A^t)^{-1} = (A^t)^{-1} \cdot A^{-1} = (A^{-1})^t \cdot A^{-1}$$

vemos que esta expresión es la forma cuadrática $q_{\Sigma^{-1}}$ asociada a Σ^{-1}

Fórmula de la densidad conjunta en la normal multivariada

Hasta ahora vimos que

$$f_Y(y) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}q(y)} \cdot |det(A^{-1})|$$

donde

$$q(y) = q_{\Sigma^{-1}}(y) = y^t \Sigma^{-1} y$$

Finalmente, veamos que relación tiene del determinante de A^{-1} con el de Σ. Como $\Sigma = A \cdot A^t$, entonces

$$\det(\Sigma) = \det(A) \cdot \det(A^t) = \det(A)^2 \Rightarrow |det(A^{-1})| = \det(\Sigma)^{-1/2}$$

y obtenemos

Densidad normal multivariada centrada en el origen

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2} y^t \Sigma^{-1} y}$$

Distribuciones marginales de la normal multivariada

Como

$$Y_j = \sum_{i=1}^{n} A_{i,j} \cdot X_j$$

y las $X_j \sim N(0,1)$ independientes,

$$A_{i,j} \cdot X_j \sim N(0, A_{i,j}^2)$$

Usando el teorema que vimos en la clase 11 sobre la suma de variables normales independientes, obtenemos que:

$$Y_j \sim N(0, \sigma_j^2) \text{ donde } \sigma_j^2 = \sum_{i=1}^{n} A_{i,j}^2$$

Notamos que esto es consistente con la fórmula

$$\Sigma_Y = A^t \cdot A$$

que obtuvimos antes (Las σ_j^2 aparecen en la diagonal de la matriz Σ_Y).

Un caso especial

Un caso de especial interés es cuando la matriz A con la que hacemos el cambio de variable es ortogonal $Y = A \cdot X$, lo que siginifca que $\Sigma = A \cdot A^t = I$. También $\det(\Sigma) = 1$. Por lo que obtenemos que $f_Y = f_X$, o sea:

Proposición

Si X tiene distribución normal multivariada estándar, y hacemos un cambio de variable $Y = A \cdot X$ con A una matriz ortogonal, Y también tiene distribución normal multivariada estándar.

Caso general

Si $b \in \mathbb{R}^n$ es cualquiera y $A \in \mathbb{R}^{n \times n}$ es una matriz no singular, a partir de $Y = A \cdot X + b$ obtendríamos que

$$\mu_Y = E[Y] = b$$

y que

$$\Sigma = \mathrm{Cov}(Y) = A \cdot A^t$$

mientras que la densidad de Y sera

Densidad normal multivariada centrada en el origen

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2}(y-\mu)^t \Sigma^{-1}(y-\mu)}$$

Algunas observaciones

Proposición

Si X tiene distribución normal multivariada, son equivalentes:

- *Las componentes X_j de X son independientes.*
- *Las X_j no están correlacionadas, o sea*

$$\operatorname{Cov}(X_i, X_j) = 0 \text{ si } i \neq j$$

  o sea la matriz $\operatorname{Cov}(X)$ es diagonal.

Esta propiedad NO es cierta para vectores aleatorios en general (como ya vimos en la clase 6).

El caso especial $n = 2$, distribución normal bivariada

En el ejercicio 25 de la práctica 7 se considera el caso especial $n = 2$, con un ligero cambio de notación: ahora el vector aleatorio se denota (X, Y), no (Y_1, Y_2). La densidad conjunta es:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)\right]\right\}}$$

donde

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \text{y} \qquad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

y ρ es el coeficiente de correlación entre X e Y.
Esta fórmula sale de que en este caso

$$\det(\Sigma) = \sigma_X^2\sigma_Y^2(1-\rho^2), \quad \mathrm{Adj}(\Sigma) = \begin{pmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_X & \sigma_Y^2 \end{pmatrix}$$

$$\Sigma^{-1} = \frac{\mathrm{Adj}(A)^t}{\det(A)} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_X} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y} \end{pmatrix}$$