

1 Multivariate Normal Distribution

Definition: A d -dimensional random vector \mathbf{X} is said to have a d -dimensional normal distribution if $\mathbf{l}'\mathbf{X}$ has univariate normal distribution for all non-zero $\mathbf{l} \in \mathbb{R}^d$.

Let X_i be the i -th component of \mathbf{X} , and $\mu_i = E(X_i)$ for $i = 1, 2, \dots, d$. Then the mean vector of \mathbf{X} is

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d).$$

The covariance between X_i and X_j is

$$\sigma_{ij} = Cov[X_i, X_j] = E[(X_i - \mu_i)(X_j - \mu_j)].$$

Note that for $i = j$, $\sigma_{ii} = Var(X_i)$. The correlation coefficient between X_i and X_j is given by $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$. The variance-covariance matrix of \mathbf{X} is given by

$$\boldsymbol{\Sigma} = ((\sigma_{ij}))_{i=1, 2, \dots, d; j=1, 2, \dots, d}$$

Notice that the covariance matrix may be specified implicitly through its diagonal entries σ_{ii} 's and correlation coefficients ρ_{ij} 's.

Let $\boldsymbol{\mu}$ be the mean vector of \mathbf{X} and $\boldsymbol{\Sigma} = ((\sigma_{ij}))$ be the variance-covariance matrix of \mathbf{X} . If $\boldsymbol{\Sigma}$ is non-singular matrix then d -dimensional normal distribution possesses a PDF and it is given by

$$\phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \quad \text{for } \mathbf{x} \in \mathbb{R}^d,$$

where $|\boldsymbol{\Sigma}|$ the determinant of $\boldsymbol{\Sigma}$.

When $\boldsymbol{\Sigma}$ is a non-singular matrix, a d -dimensional normal distribution is characterized by its mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Therefore, we use the notation $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote the fact that \mathbf{X} has a d -dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. A standard d -dimensional normal distribution is a special case where $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = I_d$, where I_d the $d \times d$ identity matrix.

An alternative definition: A d -dimensional random vector \mathbf{X} is said to have a d -dimensional normal distribution if it can be expressed in the form $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$, where A is a $d \times d$ matrix of real numbers, $\mathbf{Y} \sim N_d(\mathbf{0}, I_d)$.

In this case, $E(\mathbf{X}) = \boldsymbol{\mu}$ and $Var(\mathbf{X}) = AA'$. The resulting distribution is independent of which such A is chosen.

2 Some Properties of Multivariate Normal Distribution

1. Linear Transformation Property: Any linear transformation of a normal vector is again normal,

$$\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow A\mathbf{X} \sim \mathcal{N}_k(A\boldsymbol{\mu}, A\boldsymbol{\Sigma}A'),$$

for any d -vector μ , $d \times d$ matrix Σ , and any $k \times d$ matrix A , for any k .

2. Marginal Distributions: If $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$, then $X_i \sim N(\mu_i, \sigma_i^2)$, where $\sigma_i^2 = \Sigma_{ii}$, $i = 1, 2, \dots, d$.
3. Conditioning Formula: Suppose the partitioned vector $(\mathbf{X}_{[1]}, \mathbf{X}_{[2]})$ (where each $\mathbf{X}_{[i]}$ may itself be a vector) is multivariate normal with:

$$\begin{pmatrix} \mathbf{X}_{[1]} \\ \mathbf{X}_{[2]} \end{pmatrix} \sim \mathcal{N}_d \left(\begin{pmatrix} \mu_{[1]} \\ \mu_{[2]} \end{pmatrix}, \begin{pmatrix} \Sigma_{[11]} & \Sigma_{[12]} \\ \Sigma_{[21]} & \Sigma_{[22]} \end{pmatrix} \right)$$

and suppose $\Sigma_{[22]}$ has full rank. Then,

$$\left(\mathbf{X}_{[1]} \middle| \mathbf{X}_{[2]} = \mathbf{x} \right) \sim \mathcal{N}_m \left(\mu_{[1]} + \Sigma_{[12]} \Sigma_{[22]}^{-1} (\mathbf{x} - \mu_{[2]}), \Sigma_{[11]} - \Sigma_{[12]} \Sigma_{[22]}^{-1} \Sigma_{[21]} \right),$$

where m is the order of the vector $\mathbf{X}_{[1]}$. This equation gives the distribution of $\mathbf{X}_{[1]}$ conditional on $\mathbf{X}_{[2]} = \mathbf{x}$.

4. Moment Generating Function: If $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$, then

$$E [\exp(\theta' \mathbf{X})] = \exp \left(\mu' \theta + \frac{1}{2} \theta' \Sigma \theta \right).$$

5. Independence: If $\mathbf{X} \sim \mathcal{N}_d(\mu, \Sigma)$ and $\Sigma_{ij} = 0$ ($i \neq j$), then X_i and X_j are independent random variables.

2.1 Generating from Multivariate Normal Distribution

To generate from multivariate normal distribution, we can use the alternative definition. If $\mathbf{Z} \sim N_d(0, I_d)$ and $\mathbf{X} = \mu + A\mathbf{Z}$, then $\mathbf{X} \sim N_d(\mu, AA')$. Using any of the standard methods, we can generate independent standard normal random variables Z_1, Z_2, \dots, Z_d and assemble them into a vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d) \sim N_d(0, I_d)$. Thus, the problem of sampling from $N_d(\mu, \Sigma)$ reduces to finding a matrix A for which $AA' = \Sigma$.

Among all such A , a lower triangular one is particularly convenient because it reduces the calculation $\mu + A\mathbf{Z}$ to the following:

$$\begin{aligned} X_1 &= \mu_1 + a_{11}Z_1 \\ X_2 &= \mu_2 + a_{21}Z_1 + a_{22}Z_2 \\ \dots &= \dots \\ X_d &= \mu_d + a_{d1}Z_1 + a_{d2}Z_2 + \dots + a_{dd}Z_d, \end{aligned}$$

where $\mu = (\mu_1, \mu_2, \dots, \mu_d)$ and $A = (a_{ij})_{i,j=1,2,\dots,d}$. The Cholesky factorization of the matrix Σ may be used for finding A .

In the 2×2 case, the covariance matrix Σ is represented as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

Assuming $\sigma_1 > 0$ and $\sigma_2 > 0$, the Cholesky factorization is given by:

$$A = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1 - \rho^2}\sigma_2 \end{pmatrix}.$$

Thus, we can sample from a bivariate normal distribution by setting:

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 Z_1, \\ X_2 &= \mu_2 + \rho\sigma_2 Z_1 + \sqrt{1 - \rho^2}\sigma_2 Z_2. \end{aligned}$$

For the case of a $d \times d$ covariance matrix Σ we get:

$$\begin{aligned} a_{ij} &= \frac{\left(\sigma_{ij} - \sum_{k=1}^{j-1} a_{ik}a_{jk}\right)}{a_{jj}}, \quad j < i, \\ a_{ii} &= \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} a_{ik}^2}. \end{aligned}$$