1 Multivariate Normal Distribution

Recall that a d-dimensional random vector $\boldsymbol{X}$ is said to have a d-dimensional normal distribution if $\boldsymbol{l}'\boldsymbol{X}$ has univariate normal distribution for all non-zero $\boldsymbol{l} \in \mathbb{R}^d$. Let $\boldsymbol{\mu}$ be the mean vector of $\boldsymbol{X}$ and Σ be the variance-covariance matrix of $\boldsymbol{X}$. If Σ is non-singular matrix then d-dimensional normal distribution possesses a PDF and it is given by

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

where $|\Sigma|$ the determinant of Σ.

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

Let X_i denote the i th component of $\boldsymbol{X}$. If $X \sim N_d(\boldsymbol{\mu}, \Sigma)$ then $X_i \sim N(\mu_i, \sigma_i^2)$, where μ_i is the ith component of μ and $\sigma_i^2 = \sigma_{ii}$ is the ith diagonal of Σ. The covariance between X_i and X_j is

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

where σ_{ij} is the (i, j)th element of Σ. The correlation between X_i and X_j is given by $\rho_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$. The covariance matrix may be specified implicitly through its diagonal entries σ_i^2 and correlation ρ_{ij}.

An alternative definition of d-dimensional normal distribution can be given as follows. A d-dimensional random vector $\boldsymbol{X}$ is said to have a d-dimensional normal distribution if it can be expressed in the form $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y}$, where A is a $d \times d$ matrix of real numbers, $\boldsymbol{Y} \sim N_d(\boldsymbol{0}, I_d)$. In this case $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and $Var(\boldsymbol{X}) = AA'$.

If the $d \times d$ symmetric matrix Σ is positive semidefinite but not positive definite then the rank of Σ is less than d, Σ fails to be invertible, and there is no normal density with covariance matrix Σ. In this case, we can define the normal distribution $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ as the distribution of $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}$ with $\boldsymbol{Z} \sim \mathcal{N}_d(0, I_d)$ for any $d \times d$ matrix A satisfying $AA' = \Sigma$. 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,
$$\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma) \Rightarrow A\boldsymbol{X} \sim \mathcal{N}_k(A\boldsymbol{\mu}, A\Sigma A'),$$
for any d-vector μ, $d \times d$ matrix Σ, and any $k \times d$ matrix A, for any k.

2. Marginal Distributions: If $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, then $X_i \sim N(\mu_i, \sigma_i^2)$, where $\sigma_i^2 = \Sigma_{ii}$, $i = 1, 2, \ldots, d$.

3. Conditioning Formula: Suppose the partitioned vector $\left(\boldsymbol{X}_{[1]}, \boldsymbol{X}_{[2]}\right)$ (where each $\boldsymbol{X}_{[i]}$ may itself be a vector) is multivariate normal with:

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

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

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

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

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

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

5. Independence: If $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ 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 $\boldsymbol{Z} \sim N_d(0, I_d)$ and $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z}$, then $X \sim N_d(\boldsymbol{\mu}, AA')$. Using any of the standard methods, we can generate independent standard normal random variables $Z_1, Z_2, \ldots, Z_d$ and assemble them into a vector $\boldsymbol{Z} = (Z_1,\, Z_2,\, \ldots,\, Z_d) \sim N_d(0, I_d)$. Thus, the problem of sampling from $N_d(\boldsymbol{\mu}, \Sigma)$ reduces to finding a matrix A for which $AA' = \Sigma$. The Cholesky factorization which is described below can be used for the same.

2.1.1 Cholesky Factorization

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

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

where $\boldsymbol{\mu} = (\mu_1,\, \mu_2,\, \ldots,\, \mu_d)$ and $A = (a_{ij})_{i,j=1,\,2,\,\ldots,\,d}$. 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:

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