Note: This document is a part of the lectures given to students of IIT Guwahati during the Jan-May 2018 Semester.

A d-dimensional normal distribution is characterized by a d-vector μ and a $d \times d$ covariance matrix Σ. We abbreviate it as $\mathcal{N}(\mu, \Sigma)$, To qualify as a covariance matrix, Σ must be symmetric (i.e, Σ and $\Sigma^\top$ are equal) and positive semidefinite (meaning that $x^\top \Sigma x \geq 0$ for all $x \in \mathbb{R}^d$). This is equivalent to the requirement that all eigenvalues of Σ be nonnegative (as a symmetric matrix Σ automatically has real eigenvalues). If Σ is positive definite (meaning that strict inequality $x^\top \Sigma x > 0$ holds for all $x \in \mathbb{R}^d$) or equivalently that all eigenvalues of Σ are positive, then the normal distribution $\mathcal{N}(\mu, \Sigma)$ has the density:

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

with $|\Sigma|$ the determinant of Σ. The standard d- dimensional normal $\mathcal{N}(0, I_d)$ with I_d the $d \times d$ identity matrix is the special case:

$$\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}x^\top x\right).$$

If $X \sim \mathcal{N}(\mu, \Sigma)$ (i.e, the random vector X has multivariate normal distribution) the its i-th component X_i has distribution $\mathcal{N}(\mu_i, \sigma_i^2)$ with $\sigma_i^2 = \Sigma_{ii}$. The i-th and the j-th components have covariance:

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

which justifies calling Σ the covariance matrix. The correlation between X_i and $X_j$is given by $\rho_{ij} = \Sigma_{ij}/\sigma_i\sigma_j$. In specifying a multivariate distribution it is sometimes convenient to use the definition is opposite direction; specify the marginal standard deviation σ_i , $i = 1, 2 \ldots, d$ and correlations ρ_{ij} from which the covariance matrix $\Sigma_{ij} = \sigma_i\sigma_j\rho_{ij}$ is then determined. If the $d \times d$ matrix Σ is positive semidefinite but not positive definite then the rank of Σ is less than d and Σ fails to be invertible and there is no normal density with covariance matrix Σ. In this case we can define the normal distribution $\mathcal{N}(\mu, \Sigma)$ as the distribution of $X = \mu + AZ$ with $Z \sim \mathcal{N}(0, I_d)$ for any $d \times d$ matrix A satisfying $AA^\top = \Sigma$. The resulting distribution is independent of which such A is chosen.

Some properties of Multivariate Normal Distribution:

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

$$X \sim \mathcal{N}(\mu, \Sigma) \Rightarrow AX \sim \mathcal{N}(A\mu, A\Sigma A^\top),$$

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

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

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

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

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

This equation gives the distribution of $X_{[1]}$ conditioned on $X_{[2]} = x$.

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

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

Generating Multivariate Moments:

A multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$ is specified by its mean vector μ and covariance matrix Σ. The covariance matrix may be specified implicitly through its diagonal entries σ_i^2 and correlation ρ_{ij}. From Linear Transformation property we know that if $Z \sim \mathcal{N}(0, I)$ and $X = \mu + AZ$, then $X \sim \mathcal{N}(\mu, AA^\top)$. Using any of the standard method we can generate independent standard normal random variables $Z_1, Z_2, \ldots, Z_d$ and assemble them into a vector $Z \sim \mathcal{N}(0, I)$. Thus the problem of sampling from $\mathcal{N}(\mu, \Sigma)$ reduces to finding a matrix A for which $AA^\top = \Sigma$.

Cholesky Factorization:

Among all such A a lower triangular one is particularly convenient be cause it reduces the calculation $\mu + AZ$ to the following:

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

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:

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