1. Moments and Moment Generating Function

Moments: Let $\mathbf{X} = (X_1, \ldots, X_p)$ be a p-dimensional random vector of either discrete type, or (absolutely) continuous type. Let $f_{\mathbf{X}}$ and $S_{\mathbf{X}} = \{\mathbf{X} \in \mathbb{R}^p : f_{\mathbf{X}}(\mathbf{x}) > 0\}$ denote the pmf (or, pdf) and support of $\mathbf{X}$ (or, $f_{\mathbf{X}}$). Further, let f_{X_i} and $S_{X_i} = \{x \in \mathbb{R} : f_{X_i}(x) > 0\}$ denote the pmf (or, pdf) and support of X_i (or, f_{X_i}) for $i = 1, \ldots, p$.

Let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a function such that $\mathbf{E}[\psi(\mathbf{X})]$ exists (i.e., $\mathbf{E}|\psi(\mathbf{X})| < \infty$).

- If $\mathbf{X}$ is of discrete type, then

$$\mathbf{E}(\psi(\mathbf{X})) = \sum_{\mathbf{x} \in S_{\mathbf{X}}} \psi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}).$$

- If $\mathbf{X}$ is of absolutely continuous type, then

$$\mathbf{E}(\psi(\mathbf{X})) = \int_{\mathbb{R}^p} \psi(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

- For non-negative integers $k_1, \ldots, k_p$, let $\psi(\mathbf{x}) = x_1^{k_1} \cdots x_p^{k_p}$. Then,

$$\mu'_{k_1,\ldots,k_p} = \mathbf{E}\left(X_1^{k_1} \cdots X_p^{k_p}\right)$$

  is called a joint raw moment of order $k_1 + \cdots + k_p$ of $\mathbf{X}$.
- For non-negative integers $k_1, \ldots, k_p$, let $\psi(\mathbf{x}) = (x_1 - \mathbf{E}(X_1))^{k_1} \cdots (x_p - \mathbf{E}(X_p))^{k_p}$. Then

$$\mu_{k_1,\ldots,k_p} = \mathbf{E}\left((X_1 - \mathbf{E}(X_1))^{k_1} \cdots (X_p - \mathbf{E}(X_p))^{k_p}\right)$$

  is called a joint central moment of order $k_1 + \cdots + k_p$ of $\mathbf{X}$.
- Let $\psi(\mathbf{x}) = (x_i - \mathbf{E}(X_i))(x_j - \mathbf{E}(X_j))$ for $i, j = 1, \ldots, p$. Then, the covariance between X_i and X_j is

$$\begin{aligned} \operatorname{Cov}(X_i, X_j) &= \mathbf{E}\left((X_i - \mathbf{E}(X_i))(X_j - \mathbf{E}(X_j))\right) \\ &= \mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\mathbf{E}(X_j). \end{aligned}$$

Let $\mathbf{X} = (X_1, X_2, \ldots, X_{p_1})$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{p_2})$ be random vectors, and let $a_1, \ldots, a_{p_1}$ and $b_1, \ldots, b_{p_2}$ be real constants. Assume that the involved expectations exist. Then,

(i) $\mathbf{E}\left(\sum_{i=1}^{p_1} a_i X_i\right) = \sum_{i=1}^{p_1} a_i \mathbf{E}(X_i)$

(ii) $\operatorname{Cov}\left(\sum_{i=1}^{p_1} a_i X_i, \sum_{j=1}^{p_2} b_j Y_j\right) = \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} a_i b_j \operatorname{Cov}(X_i, Y_j)$.

In particular,

$$\operatorname{Var}\left(\sum_{i=1}^{p_1} a_i X_i\right) = \sum_{i=1}^{p_1} a_i^2 \operatorname{Var}(X_i) + \sum_{i=1}^{p_1} \sum_{\substack{j=1 \\ j \neq i}}^{p_1} a_i a_j \operatorname{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{p_1} a_i^2 \operatorname{Var}(X_i) + 2 \sum_{1 \leq i < j \leq p_1} a_i a_j \operatorname{Cov}(X_i, X_j).$$

We now state a property of expectations related to independence.

Lemma 1. *Let X, Y be random variables on a common probability space. If X and Y are independent, then $\mathbf{E}[H_1(X)H_2(Y)] = \mathbf{E}[H_1(X)]\mathbf{E}[H_2(Y)]$ for any functions $H_1, H_2 : \mathbb{R} \to \mathbb{R}$ (for which the expectations exist). In particular, $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Proof. Independence means that the joint density (analogous statements for pmf omitted) of (X, Y) is of the form $f(t, s) = g(t)h(s)$, where $g(t)$ is the density of X and $h(s)$ is the density of Y. Hence,

$$\mathbf{E}[H_1(X)H_2(Y)] = \iint H_1(t)H_2(s)f(t,s)dtds = \left(\int_{-\infty}^{\infty} H_1(t)g(t)dt\right)\left(\int_{-\infty}^{\infty} H_2(s)h(s)ds\right)$$

which is precisely $\mathbf{E}[H_1(X)]\mathbf{E}[H_2(Y)]$. ■

Moment Generating Function: Let $\mathbf{X} = (X_1, \ldots, X_p)$ be a p-dimensional random vector, and

$$A = \left\{\mathbf{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p : \mathbf{E}\left(e^{\sum_{i=1}^p t_i X_i}\right) < \infty\right\}.$$

Define the function $M_{\mathbf{X}} : A \to \mathbb{R}$ by

$$M_{\mathbf{X}}(\mathbf{t}) = \mathbf{E}\left(e^{\sum_{i=1}^p t_i X_i}\right), \quad \mathbf{t} = (t_1, t_2, \ldots, t_p) \in A.$$

The function $M_{\mathbf{X}} : A \to \mathbb{R}$ is called the joint moment generating function (mgf) of random vector $\mathbf{X}$. For $\mathbf{a} = (a_1, a_2, \ldots, a_p) \in \mathbb{R}^p$, $-\mathbf{a} = (-a_1, -a_2, \ldots, -a_p)$ and $(-\mathbf{a}, \mathbf{a}) = \{\mathbf{t} \in \mathbb{R}^p : -a_i < t_i < a_i \text{ for } i = 1, \ldots, p\}$. As in the one-dimensional case, many properties of probability distribution of $\mathbf{X}$ can be studied through the joint mgf of $\mathbf{X}$. Some of the results, which may be useful in this direction, are provided below (without their proofs). Note that $M_{\mathbf{X}}(0_p) = 1$, where 0_p is the vector of 0s.

If $X_1, \ldots, X_p$ are independent, then

$$M_{\mathbf{X}}(\mathbf{t}) = \mathbf{E}\left(e^{\sum_{i=1}^p t_i X_i}\right) = \mathbf{E}\left(\prod_{i=1}^p e^{t_i X_i}\right) = \prod_{i=1}^p \mathbf{E}\left(e^{t_i X_i}\right) = \prod_{i=1}^p M_{X_i}(t_i) \text{ for } \mathbf{t} \in \mathbb{R}^p.$$

Suppose that $M_{\mathbf{X}}(\mathbf{t})$ exists in a rectangle $(-\mathbf{a}, \mathbf{a}) \subseteq \mathbb{R}^p$. Then, $M_{\mathbf{X}}(\mathbf{t})$ possesses partial derivatives of all orders in $(-\mathbf{a}, \mathbf{a})$. Furthermore, for positive integers $k_1, \ldots, k_p$

$$\mathbf{E}\left(X_1^{k_1} X_2^{k_2} \cdots X_p^{k_p}\right) = \left[\frac{\partial^{k_1+k_2+k_3+\cdots+k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\mathbf{X}}(\mathbf{t})\right]_{\mathbf{t}=0_p}.$$

For $i \neq j$ with $i, j \in \{1, \ldots, p\}$, define

$$\begin{aligned}
\operatorname{Cov}(X_i, X_j) &= \mathbf{E}(X_i X_j) - \mathbf{E}(X_i)\,\mathbf{E}(X_j) \\
&= \left[\frac{\partial^2}{\partial t_i \partial t_j} M_{\mathbf{X}}(\mathbf{t})\right]_{\mathbf{t}=0_p} - \left[\frac{\partial}{\partial t_i} M_{\mathbf{X}}(\mathbf{t})\right]_{\mathbf{t}=0_p} \left[\frac{\partial}{\partial t_j} M_{\mathbf{X}}(\mathbf{t})\right]_{\mathbf{t}=0_p} \\
&= \left[\frac{\partial^2}{\partial t_i \partial t_j} \Psi_{\mathbf{X}}(\mathbf{t})\right]_{\mathbf{t}=0_p},
\end{aligned}$$

where $\Psi_{\mathbf{X}}(\mathbf{t}) = \ln M_{\mathbf{X}}(\mathbf{t})$.
We also have $M_{\mathbf{X}}(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0) = \mathbf{E}\left(e^{t_i X_i + t_j X_j}\right) = M_{X_i, X_j}(t_i, t_j)$ for $i, j \in \{1, \ldots, p\}$.

Identically distributed: Let $\mathbf{X}$ and $\mathbf{Y}$ be two p-dimensional random vectors, defined on the same probability space. Then, $\mathbf{X}$ and $\mathbf{Y}$ are said to have the same distribution (written as $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$) if

$$F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{Y}}(\mathbf{x}) \ \forall\ \mathbf{x} \in \mathbb{R}^p \text{ (i.e., they have the same distribution function).}$$

If $\mathbf{X}$ and $\mathbf{Y}$ are p-dimensional random vectors of discrete type with joint pmf $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$, respectively. Then, $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$ if and only if $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{x}) \ \forall\ \mathbf{x} \in \mathbb{R}^p$.

If $\mathbf{X}$ and $\mathbf{Y}$ are p-dimensional random vectors of absolutely continuous type with joint pdf $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$, respectively. Then, $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$ if and only if $f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{x}) \ \forall\ \mathbf{x} \in \mathbb{R}^p$.

Let $\mathbf{X}$ and $\mathbf{Y}$ be p-dimensional random vectors with $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$. Then, for any function $h : \mathbb{R}^p \to \mathbb{R}$, we have

(i) $h(\mathbf{X}) \stackrel{D}{=} h(\mathbf{Y})$,

(ii) $\mathbf{E}[h(\mathbf{X})] = \mathbf{E}[h(\mathbf{Y})]$ (provided the expectations exist).

Let $\mathbf{X}$ and $\mathbf{Y}$ be two random vectors having mgfs $M_{\mathbf{X}}$ and $M_{\mathbf{Y}}$ that are finite on a rectangle $(-\mathbf{a}, \mathbf{a})$ for some $\mathbf{a} = (a_1, a_2, \ldots, a_p) \in \mathbb{R}^p$. Suppose that

$$M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{t}) \quad \forall \quad \mathbf{t} \in (-\mathbf{a}, \mathbf{a}).$$

Then, $\mathbf{X} \stackrel{D}{=} \mathbf{Y}$.

If $X_1, X_2, \ldots, X_p$ are independent and identically distributed (i.i.d.), i.e., $X_i \overset{D}{=} X_1$ for $i = 2, \ldots, p$, then

$$M_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^{p} M_{X_1}(t_i) \text{ for } \mathbf{t} \in \mathbb{R}^p.$$

Define $Y = \sum_{i=1}^{p} X_i$ and $\bar{X} = \frac{1}{p} \sum_{i=1}^{p} X_i$, then

$$M_Y(t) = [M_{X_1}(t)]^p \text{ and } M_{\bar{X}}(t) = [M_{X_1}(t/p)]^p \text{ for } t \in \mathbb{R}.$$

Exercise 2. Let $X_1, X_2, \ldots, X_p$ be independent random variables such that $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ with $-\infty < \mu_i < \infty$ and $\sigma_i > 0$ for $i = 1, \ldots, p$. If $a_1, \ldots, a_p$ are real constants (such that not all of them are zero), then show that

$$\sum_{i=1}^{p} a_i X_i \sim N\left(\sum_{i=1}^{p} a_i \mu_i, \sum_{i=1}^{p} a_i^2 \sigma_i^2\right).$$

2. Covariance and Correlation

Covariance: Let X, Y be random variables on a common probability space. The *covariance* of X and Y is defined as $\text{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$. It can also be written as $\text{Cov}(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$.

Correlation: Let X, Y be random variables on a common probability space. Their *correlation* is defined as $\text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$.

Measures of association: The marginal distributions of X and Y do not determine the joint distribution of (X, Y). In particular, giving the means and standard deviations of X and Y does not tell anything about possible relationships between the two.
Read this: http://probability.ca/jeff/teaching/uncornor.html.

Covariance is the quantity that is used to measure the "association" of Y and X. Correlation is a dimension free quantity that measures the same. For example, we shall see that if $Y = X$, then $\text{Corr}(X, Y) = +1$, if $Y = -X$ then $\text{Corr}(X, Y) = -1$. Further, if X and Y are independent, then $\text{Corr}(X, Y) = 0$. In general, if an increase in X is likely to mean an increase in Y, then the correlation is positive and if an increase in X is likely to mean a decrease in Y then the correlation is negative.

Properties of covariance and variance: Let X, Y, X_i, Y_i be random variables on a common probability space. Small letters a, b, c etc. will denote scalars.

(1) (Bilinearity): $\text{Cov}(aX_1 + bX_2, Y) = a\text{Cov}(X_1, Y) + b\text{Cov}(X_2, Y)$ and $\text{Cov}(X, aY_1 + bY_2) = a\text{Cov}(X, Y_1) + b\text{Cov}(X, Y_2)$.

(2) (Symmetry): $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

(3) (Positivity): $\text{Cov}(X, X) \geq 0$ with equality if and only if X is a constant random variable. Indeed, $\text{Cov}(X, X) = \text{Var}(X)$.

Exercise 3. If X and Y are independent, then show that $\text{Cov}(X, Y) = 0$ and hence, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

However, $\text{Cov}(X, Y) = 0$ does not necessarily imply that X and Y are independent!

Example 4. Let $\mathbf{X} = (X_1, X_2)$ be a bivariate random vector of absolutely continuous type with pdf given by

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} 1, & \text{if } 0 < |x_2| \leq x_1 < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\mathbf{E}\left(X_1X_2\right) = \int_0^1 \int_{-x_1}^{x_1} x_1x_2 \,\mathrm{d}x_2 \,\mathrm{d}x_1 = 0,$$

$$\mathbf{E}\left(X_1\right) = \int_0^1 \int_{-x_1}^{x_1} x_1 \,\mathrm{d}x_2 \,\mathrm{d}x_1 = \frac{2}{3},$$

$$\mathbf{E}\left(X_2\right) = \int_0^1 \int_{-x_1}^{x_1} x_2 \,\mathrm{d}x_2 \,\mathrm{d}x_1 = 0,$$

and

$$\mathrm{Cov}\left(X_1, X_2\right) = \mathbf{E}\left(X_1X_2\right) - \mathbf{E}\left(X_1\right)\mathbf{E}\left(X_2\right) = 0$$

Therefore,

$$\mathrm{Corr}\left(X_1, X_2\right) = 0$$

i.e., X_1 and X_2 are uncorrelated.

Exercise 5. Show that

$$f_{\mathbf{X}}\left(x_1, x_2\right) \neq f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \quad \forall\ \left(x_1, x_2\right) \in \mathbb{R}^2.$$

Therefore, X_1 and X_2 are not independent.

Remark 6. Note that the properties of covariance are very much like properties of inner-products in vector spaces. In particular, we have the following analogue of the well-known inequality for vectors $(\mathbf{u} \cdot \mathbf{v})^2 \leq (\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v})$.

Cauchy-Schwarz inequality: If X and Y are random variables with finite variances, then

$$(\mathrm{Cov}(X, Y))^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y)$$

with equality if and only if $Y = aX + b$ for some scalars a, b.
Follow the proof of Cauchy-Schwarz inequality that you have seen for vectors. This just means that $\mathrm{Var}(X + tY) \geq 0$ for any scalar t and choose an appropriate t to get the Cauchy-Schwarz inequality.

Bivariate Normal Distribution: We say that (X, Y) follows a bivariate normal distribution if its pdf is given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-Q/2} \text{ for } \quad -\infty < x < \infty, \quad -\infty < y < \infty,$$

where

$$Q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

with $-\infty < \mu_i < \infty, \sigma_i > 0$ for $i = 1, 2$, and ρ satisfies $\rho^2 < 1$. Clearly, this function is positive everywhere in $\mathbb{R}^2$.

Remark 7. Note that we can derive this new formulation from the earlier one by defining

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

Exercise 8. Compute Σ^{-1} and $det(\Sigma)$.

This pdf has mgf given by (proof is given later):

$$M_{(X,Y)}(t_1, t_2) = \exp\left\{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}\left(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2\right)\right\}.$$

Thus, the mgf of X is

$$M_X(t_1) = M_{(X,Y)}(t_1, 0) = \exp\left\{t_1\mu_1 + \frac{1}{2}t_1^2\sigma_1^2\right\},$$

while the mgf of Y is

$$M_Y(t_2) = M_{(X,Y)}(0, t_2) = \exp\left\{t_2\mu_2 + \frac{1}{2}t_2^2\sigma_2^2\right\}.$$

Hence, X has a $N\left(\mu_1, \sigma_1^2\right)$ distribution. In the same way, Y has a $N\left(\mu_2, \sigma_2^2\right)$ distribution. Thus, μ_1 and μ_2 are the respective means of X and Y, while σ_1^2 and σ_2^2 are the respective variances of X and Y.

Exercise 9. For the parameter ρ, show that

$$\mathbf{E}(XY) = \frac{\partial^2 M_{(X,Y)}}{\partial t_1 \partial t_2}(0,0) = \rho\sigma_1\sigma_2 + \mu_1\mu_2.$$

Hence, $\mathrm{Cov}(X, Y) = \rho\sigma_1\sigma_2$. As the notation suggests, ρ is the correlation coefficient between X and Y (Check!).

Remark 10. We know that if X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. For the bivariate normal distribution, if $\rho = 0$ (equivalently, $\mathrm{Cov}(X, Y) = 0$), then the joint mgf of (X, Y) factors into the product of the marginal mgfs. Hence, X and Y are independent random variables. Thus, if (X, Y) has a bivariate normal distribution, then X and Y are independent if and only if they are uncorrelated (i.e., $\rho = 0$). Read below:
https://en.wikipedia.org/wiki/Normally_distributed_and_uncorrelated_does_not_imply_independent.

Check the file 'N_2distribution.R'.

Multivariate Normal Distribution: In this section, we generalize the bivariate normal distribution to the n-dimensional multivariate normal distribution. The derivation of the distribution is simplified by first discussing the standardized variable case, and then proceeding to the general case. Consider the random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)$, where $Z_1, \ldots, Z_n$ are i.i.d. $N(0, 1)$ random variables. Then, the density of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\mathbf{z}'\mathbf{z}\right\}$$

for $\mathbf{z} \in \mathbb{R}^n$.

Definition 11. We define $\mathbf{E}[\mathbf{X}]$ as the n-dimensional vector $(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_n])'$, and $\mathrm{Cov}[\mathbf{X}]$ as a $n \times n$ matrix with the (i, j)th element as $\mathrm{Cov}(X_i, X_j)$ for $1 \leq i, j \leq n$.

Note that the diagonal elements of $\mathrm{Cov}[\mathbf{X}]$ (a symmetric matrix) are the componentwise variances $\mathrm{Var}[X_i]$ for $1 \leq i \leq n$.

Exercise 12. The mean and covariance matrix of $\mathbf{Z}$ are

$$\mathbf{E}[\mathbf{Z}] = 0_n \text{ and } \mathrm{Cov}[\mathbf{Z}] = I_n,$$

where 0_n is the vector of 0s and I_n denotes the identity matrix of order n.

The mgf of Z_is evaluated at t_i is $\exp\left\{t_i^2/2\right\}$ for $i = 1, \ldots, n$. Since the Z_is are independent, the mgf of $\mathbf{Z}$ is

$$M_{\mathbf{Z}}(\mathbf{t}) = \mathbf{E}\left[\exp\left\{\mathbf{t}'\mathbf{Z}\right\}\right] = \mathbf{E}\left[\prod_{i=1}^{n} \exp\left\{t_i Z_i\right\}\right] = \prod_{i=1}^{n} \mathbf{E}\left[\exp\left\{t_i Z_i\right\}\right] = \exp\left\{\frac{1}{2}\sum_{i=1}^{n} t_i^2\right\} = \exp\left\{\frac{1}{2}\mathbf{t}'\mathbf{t}\right\}.$$

for all $\mathbf{t} \in \mathbb{R}^n$. We say that $\mathbf{Z}$ has a multivariate normal distribution with mean vector 0_n and covariance matrix I_n. We abbreviate this by saying that $\mathbf{Z}$ has a $N_n\left(0_n, I_n\right)$ distribution.

For the *general case*, suppose Σ is a $n \times n$ symmetric and positive definite matrix. Then, from linear algebra, we can always decompose Σ as follows:

$$\Sigma = \Gamma' \Lambda \Gamma,$$

where Λ is the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues and the columns of Γ' (say, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$) are the corresponding eigenvectors of Σ. This decomposition is called the spectral decomposition of Σ. The matrix Γ is orthogonal, i.e., $\Gamma^{-1} = \Gamma'$ and hence, $\Gamma\Gamma' = I_n$.

We define the square root of the positive definite matrix Σ as

$$\Sigma^{1/2} = \Gamma' \Lambda^{1/2} \Gamma.$$

Note that $\Sigma^{1/2}$ is symmetric and positive definite. It is now easy to show that

$$\left(\Sigma^{1/2}\right)^{-1} = \Gamma' \Lambda^{-1/2} \Gamma.$$

We write the left side of this equation as $\Sigma^{-1/2}$. Suppose $\mathbf{Z}$ has a $N_n\,(0_n, I_n)$ distribution. Let Σ be a positive definite, symmetric matrix and μ be an $n \times 1$ vector of constants. Define the random vector $\mathbf{X}$ by

$$\mathbf{X} = \Sigma^{1/2}\mathbf{Z} + \mu.$$

We now have

$$\mathbf{E}[\mathbf{X}] = \mu \text{ and } \mathrm{Cov}[\mathbf{X}] = \Sigma^{1/2}\Sigma^{1/2} = \Sigma.$$

Further, the mgf of $\mathbf{X}$ is given by

$$\begin{aligned}
M_{\mathbf{X}}(\mathbf{t}) = \mathbf{E}\left[\exp\left\{\mathbf{t}'\mathbf{X}\right\}\right] &= \mathbf{E}\left[\exp\left\{\mathbf{t}'\Sigma^{1/2}\mathbf{Z} + \mathbf{t}'\mu\right\}\right] \\
&= \exp\left\{\mathbf{t}'\mu\right) \left[\exp\left\{\left(\Sigma^{1/2}\mathbf{t}\right)'\mathbf{Z}\right\}\right] \\
&= \exp\left\{\mathbf{t}'\mu\right\} \exp\left\{(1/2)\left(\Sigma^{1/2}\mathbf{t}\right)'\Sigma^{1/2}\mathbf{t}\right\} \\
&= \exp\left\{\mathbf{t}'\mu\right\} \exp\left\{(1/2)\mathbf{t}'\Sigma\mathbf{t}\right\} \\
&= \exp\left\{\mathbf{t}'\mu + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right\}.
\end{aligned}$$

The transformation between $\mathbf{X}$ and $\mathbf{Z}$ is one-to-one with the inverse transformation

$$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \mu)$$

and Jacobian $\left|\Sigma^{-1/2}\right| = |\Sigma|^{-1/2}$. Hence, upon simplification, the pdf of $\mathbf{X}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right\} \quad \text{for } \mathbf{x} \in \mathbb{R}^n.$$

The following theorem says that a linear transformation of a multivariate normal random vector has a multivariate normal distribution.

Theorem 13. *Suppose* $\mathbf{X}$ *has a* $N_n(\mu, \Sigma)$ *distribution. Let* $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$*, where* A *is an* $m \times n$ *matrix and* $\mathbf{b} \in \mathbb{R}^m$*. Then,* $\mathbf{Y}$ *has a* $N_m\left(A\mu + \mathbf{b}, A\Sigma A'\right)$ *distribution.*

Proof. For $\mathbf{t} \in \mathbb{R}^m$, the mgf of $\mathbf{Y}$ is

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= \mathbf{E}\left[\exp\left\{\mathbf{t}'\mathbf{Y}\right\}\right] \\
&= \mathbf{E}\left[\exp\left\{\mathbf{t}'(A\mathbf{X} + \mathbf{b})\right\}\right] \\
&= \exp\left\{\mathbf{t}'\mathbf{b}\right\} \mathbf{E}\left[\exp\left\{\left(A'\mathbf{t}\right)'\mathbf{X}\right\}\right] \\
&= \exp\left\{\mathbf{t}'\mathbf{b}\right\} \exp\left\{\left(A'\mathbf{t}\right)'\mu + (1/2)\left(A'\mathbf{t}\right)'\Sigma\left(A'\mathbf{t}\right)\right\} \\
&= \exp\left\{\mathbf{t}'(A\mu + \mathbf{b}) + \frac{1}{2}\mathbf{t}'A\Sigma A'\mathbf{t}\right\}
\end{aligned}$$

which is the mgf of a $N_m\left(A\mu + \mathbf{b}, A\Sigma A'\right)$ distribution. ■