Section 3.5. The Multivariate Normal Distribution

Note. We start with the bivariate normal distribution and then consider the multivariate distribution. Random vectors were introduced in our Section 2.6. Extension to Several Random Variables.

Definition. The random vector (X, Y) follows a *bivariate normal distribution* if its probability density function given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-q/2} \text{ for } x, y \in \mathbb{R},$$

where

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

and $-\infty < \mu_i < \infty$, $\sigma_i > 0$ for $i = 1, 2$, and ρ satisfies $\rho^2 < 1$.

Note 3.5.A. We'll show below that $f(x, y)$ is a probability density function with moment generating function

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

Note 3.5.B. By Note 2.1.D, the marginal moment generating function $M_X(t_1) = M_{(X,Y)}(t_1, 0)$, so we have $M_X(t_1) = \exp(t_1\mu_1 + t_1^2\sigma_1^2/2)$. By Note 3.4.B, this is the moment generating function of a random variable with distribution $N(\mu_1, \sigma_1^2)$. So X has a $N(\mu_1, \sigma_1^2)$ distribution, and similarly Y has a $N(\mu_1, \sigma_2^2)$ distribution. It is to be shown in Exercise 3.5.3 (with the help of Note 2.5.C) that

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

Now by Note 2.5.A, $\text{cov}(X,Y) = E[XY] - \mu_1\mu_2$ so that we must have $\text{cov}(X,Y) = \rho\sigma_1\sigma_2$ and hence ρ is the correlation between X and Y by Definition 2.5.2.

Lemma 3.5.A. Let random vector (X,Y) have the bivariate normal distribution. Then X and Y are independent if and only if they are uncorrelated (that is, $\rho = 0$).

Note. The bivariate normal probability density function $f(x,y)$ is "mound shaped" over $\mathbb{R}^2$ with a maximum at its mean $(\mu_1\mu_2)$ (as is to be shown in Exercise 3.5.4(a)). For a given $c > 0$, the points of equal probability are given by the set $\{(x,y) \mid f(x,y) = c\}$. These sets form ellipses in $\mathbb{R}^2$ as is to be shown in Exercise 3.5.4(b) (these ellipses are called *contours* of f). If X and Y are independent then these contours are circular. This case is illustrated in the figure below, which is from Andrew J. Baczkowski's webpage on Statistical Methods (accessed June 10, 2021).

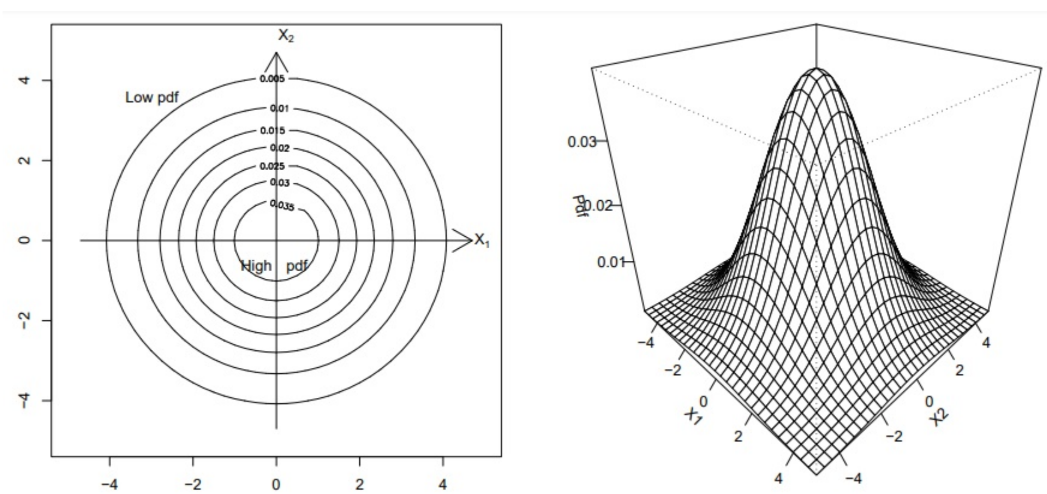

Figure. The contours (left) and the pdf for a bivariate normal distribution of independent variables.

Note 3.5.C. Let $Z_1, Z_2, \ldots, Z_n$ be independent and identically distributed ("iid") standard normal random variables. The probability density functions for the Z_i are $\frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2} z_i^2\right)$. Since the Z_i are independent, then the probability density function for $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ 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 $\mathbf{z} \in \mathbb{R}^n$ (notice that $\mathbf{z}'\mathbf{z}$ is the dot product of $\mathbf{z}$ with itself). The Z_i have mean 0, variance 1, and are independent (the pairwise correlation coefficients are all 0), so the mean vector and covariance matrix of $\mathbf{Z}$ are $E[\mathbf{Z}] = \mathbf{0}$ and $\text{Cov}[\mathbf{Z}] = \mathbf{I}_n$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The moment generating function of Z_i evaluated at t_i is $\exp(t_i^2/2)$ by Note 3.4.A. So by Note 2.6.B, the moment generating function of $\mathbf{Z}$ is

$$M_{\mathbf{Z}}(\mathbf{t}) = E[\exp(\mathbf{t}'\mathbf{Z})] = E\left[\prod_{i=1}^{n} \exp(t_i Z_i)\right] = \prod_{i=1}^{n} E[\exp(t_i Z_i)]$$
$$= \prod_{i=1}^{n} \exp(t_i^2/2) = \exp\left(\frac{1}{2} \sum_{i=1}^{n} t_i^2\right) = \exp\left(\frac{1}{2} \mathbf{t}'\mathbf{t}\right) \quad (3.5.7)$$

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

Definition. Let $Z_1, Z_2, \ldots, Z_n$ be independent and identical in distribution standard normal random variables. Then random vector $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)'$ has a *multivariate normal distribution* with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_n$. We denote this as $\mathbf{Z}$ has an $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution. The moment generating function of $\mathbf{Z}$ is given in (3.5.7) above.

Note. In the general case of a multivariate normal distribution, we consider the random variable $\mathbf{X} = \mathbf{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ where $\mathbf{Z}$ is as above, $\boldsymbol{\mu}$ gives the mean of $\mathbf{X}$, $\mathbf{\Sigma}$ corresponds to variance, and we will further explain the meaning of $\mathbf{\Sigma}^{1/2}$ below.

Note. Every real symmetric matrix $\mathbf{A}$ can be diagonalized as $\mathbf{A} = \mathbf{CDC}^{-1}$, where $\mathbf{D}$ is a diagonal matrix with eigenvectors of $\mathbf{A}$ as the diagonal entries and the columns of $\mathbf{C}$ consist of eigenvectors of $\mathbf{A}$ (with the eigenvalues in $\mathbf{D}$ and the eigenvectors in $\mathbf{C}$ are in corresponding positions); in addition, the eigenvectors can be chosen so that they form an orthonormal set and hence $\mathbf{C}$ is orthogonal. This claim is the Fundamental Theorem of Real Symmetric Matrices; see my online Linear Algebra (MATH 2010) notes on Section 6.3. Orthogonal Matrices (see Theorem 6.8). According to Hogg, McKean, Craig, this is called the *spectral decomposition* of $\mathbf{A}$. In addition a real symmetric matrix is positive semidefinite (or nonnegative definite) if and only if all of its eigenvalues are positive (see Theorem 3.8.14 in my online notes for Theory of Matrices [MATH 5090] on Section 3.8. Eigenanalysis; Canonical Factorizations).

Definition/Note 3.5.D. Let $\mathbf{\Sigma}$ be an $n \times n$ symmetric and positive semidefinite matrix. Then $\mathbf{\Sigma}$ has a spectral decomposition $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}$ where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$ are the eigenvalues of $\mathbf{\Sigma}$ and the columns of $\mathbf{\Gamma}'$, say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, are the corresponding (unit) eigenvectors. Then, by the Fundamental Theorem of real Symmetric Matrices described above, matrix $\mathbf{\Gamma}$ is orthogonal; that is, $\mathbf{\Gamma}^{-1} = \mathbf{\Gamma}'$ so that $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}_n$. Then the spectral decomposition of $\mathbf{\Sigma}$ is, as is to be shown in Exercise 3.5.19, $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i'$. Since the eigenvectors λ_i are nonnegative then we can define $\mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$. Then, since $\mathbf{\Gamma}$ is orthogonal and $\mathbf{\Gamma}\mathbf{\Gamma}' = \mathbf{I}_n$, we have

$$\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{\Gamma} = \mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{I}_n\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}$$

$$= \mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}(\mathbf{\Gamma}\mathbf{\Gamma}')\mathbf{\Lambda}^{1/2}\mathbf{\Gamma} = (\mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma})(\mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma}).$$

Define the *square root* of (positive semidefinite) $\mathbf{\Sigma}$ as $\mathbf{\Sigma}^{1/2} = \mathbf{\Gamma}'\mathbf{\Sigma}^{1/2}\mathbf{\Gamma}$. Notice that $\mathbf{\Sigma}^{1/2}$ is symmetric since

$$(\mathbf{\Sigma}^{1/2})' = (\mathbf{\Gamma}'\mathbf{\Sigma}^{1/2}\mathbf{\Gamma})' = \mathbf{\Gamma}'(\mathbf{\Lambda}^{1/2})\mathbf{\Gamma}'' = \mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma} = \boldsymbol{\sigma}^{1/2}.$$

Also, $\mathbf{\Sigma}^{1/2}$ is positive definite since

$$\det(\mathbf{\Sigma}^{1/2} - \lambda\mathbf{I}_n) = \det(\mathbf{\Gamma}'\mathbf{\Lambda}^{1/2}\mathbf{\Gamma} - \lambda\mathbf{\Gamma}'\mathbf{\Gamma}) = \det(\mathbf{\Gamma}'(\mathbf{\Lambda}^{1/2} - \lambda\mathbf{I}_n)\mathbf{\Gamma})$$

$$= \det(\mathbf{\Gamma}')\det(\mathbf{\Lambda}^{1/2} - \lambda\mathbf{I}_n)\det(\mathbf{\Gamma}) = \det(\mathbf{\Gamma}^{-1})\det(\mathbf{\Lambda}^{1/2} - \lambda\mathbf{I}_n)\det(\mathbf{\Gamma})$$

$$= \det(\mathbf{\Gamma})^{-1}\det(\mathbf{\Lambda}^{1/2} - \lambda\mathbf{I}_n)\det(\mathbf{\Gamma}) = \det(\mathbf{\Lambda}^{1/2} - \lambda\mathbf{I}_n)$$

and hence the eigenvalues of $\mathbf{\Sigma}^{1/2}$ and $\mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$ are the same (and are nonnegative). If $\mathbf{\Sigma}$ is positive definite (so that the eigenvalues of $\mathbf{\Sigma}$ are positive) then we can define

$$(\mathbf{\Sigma})^{-1/2} = (\mathbf{\Sigma}^{1/2})^{-1} = \mathbf{\Gamma}'(\mathbf{\Lambda}^{1/2})^{-1}\mathbf{\Gamma},$$

as is to be shown in Exercise 3.5.13.

Note 3.5.E. For $\mathbf{Z}$ with a $N_n(\mathbf{0}, \mathbf{I}_n)$ distribution and $\boldsymbol{\Sigma}$ a positive semidefinite symmetric matrix and let $\boldsymbol{\mu}$be an $n \times 1$ vector of constants, define the random vector $\mathbf{X}$ as $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$. Since the mean of $\mathbf{Z}$ is $\mathbf{0}$, then the mean of $\mathbf{X}$ is $\boldsymbol{\mu}$. Since $\mathrm{Cov}(\mathbf{Z}) = \mathbf{I}_n$ then, by Theorem 2.6.3,

$$\mathrm{Cov}(\mathbf{X}) = \mathrm{Cov}(\boldsymbol{\Sigma}^{1/2}\mathbf{Z}+\boldsymbol{\mu}) = \mathrm{Cov}(\boldsymbol{\Sigma}^{1/2}\mathbf{Z}) = \boldsymbol{\Sigma}^{1/2}\mathrm{Cov}(\mathbf{Z})(\boldsymbol{\Sigma}^{1/2})' = \boldsymbol{\Sigma}^{1/2}\mathbf{I}_n\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}.$$

The moment generating function of $\mathbf{X}$ is

$$\begin{aligned}
M_{\mathbf{X}}(\mathbf{t}) &= E[\exp(\mathbf{t}'\mathbf{X})] = E[\exp(\mathbf{t}'\boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \mathbf{t}'\boldsymbol{\mu})] \\
&= E[\exp(\mathbf{t}'\boldsymbol{\Sigma}^{1/2}\mathbf{Z})\exp(\mathbf{t}'\boldsymbol{\mu})] = \exp(\mathbf{t}'\boldsymbol{\mu})E[\exp((\boldsymbol{\Sigma}^{1/2}\mathbf{t})'\mathbf{Z})] \\
&= \exp(\mathbf{t}'\boldsymbol{\mu})E[\exp((\boldsymbol{\Sigma}^{1/2}\mathbf{t})'\mathbf{Z})] \text{ since } \boldsymbol{\Sigma}^{1/2} \text{ is symmetric} \\
&= \exp(\mathbf{t}'\boldsymbol{\mu}) = M_{\mathbf{Z}}((\boldsymbol{\Sigma}^{1/2}\mathbf{t})\mathbf{Z}) \\
&= \exp(\mathbf{t}'\boldsymbol{\mu})\exp\left(\frac{1}{2}(\boldsymbol{\Sigma}^{1/2}\mathbf{t})'(\boldsymbol{\Sigma}^{1/2}\mathbf{t})\right) \text{ by (3.5.7)} \\
&= \exp(\mathbf{t}'\boldsymbol{\mu})\exp\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{t}\right) \\
&= \exp(\mathbf{t}'\boldsymbol{\mu})\exp\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right) \\
&= \exp\left(\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right).
\end{aligned}$$

Definition 3.5.1. An n-dimensional random vector $\mathbf{X}$ has a *multivariate normal distribution* if its moment generating function is

$$M_{\mathbf{X}}(\mathbf{t}) = \exp(\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) \text{ for all } \mathbf{t} \in \mathbb{R}^n,$$

where $\boldsymbol{\Sigma}$ is a symmetric, positive semi-definite matrix and $\boldsymbol{\mu} \in \mathbb{R}^n$. We say that $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

Note 3.5.F. If $\mathbf{\Sigma}$ is positive definite, then $\mathbf{\Sigma}^{-1/2}$ is defined as given above (and as given in Exercise 3.5.13). Since we $\mathbf{X} = \mathbf{\Sigma}^{1/2}(\mathbf{X} - \boldsymbol{\mu})$, then in this case we have $\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$. In Exercise 3.5.A it is to be shown that the probability density function of $\mathbf{X}$ is

$$f_{\mathbf{F}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(\frac{-1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \text{ for } \mathbf{x} \in \mathbb{R}^n. \qquad (3.5.16)$$

Note. If we set $f_{\mathbf{X}}(\mathbf{x})$ or (3.5.16) equal to a positive constant to find the contours of $f_{\mathbf{X}}(\mathbf{x})$. This corresponds to the condition $(\mathbf{X}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) = c^2$. This implies

$$(x_1 - \mu_1, x_2 - \mu_2, \ldots, x_n - \mu_n) \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_n \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{pmatrix}$$

$$= (x_1 - \mu_1, x_2 - \mu_2, \ldots, x_n - \mu_n) \begin{pmatrix} (x_1 - \mu_1)/\lambda_1 \\ (x_2 - \mu_2)/\lambda_2 \\ \vdots \\ (x_n - \mu_n)/\lambda_n \end{pmatrix}$$

$$= \frac{(x_1 - \mu_1)^2}{\lambda_1} + \frac{(x_2 - \mu_2)^2}{\lambda_2} + \cdots + \frac{(x_n - \mu_n)^2}{\lambda_n} = c^2$$

(notice that $(\mathbf{X} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is in fact positive) or

$$= \frac{(x_1 - \mu_1)^2}{c^2\lambda_1} + \frac{(x_2 - \mu_2)^2}{c^2\lambda_2} + \cdots + \frac{(x_n - \mu_n)^2}{c^2\lambda_n} = 1.$$

So the contours of $f_{\mathbf{X}}(\mathbf{x})$ are ellipsoids in n-dimensional space with center $(\mu_1, \mu_2, \ldots, \mu_n)$. In the event that $\mathbf{\Sigma}^{1/2} = \mathbf{I}_n$ and $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{X} = \mathbf{Z}$) then we see that the contours are spheres of radius $c > 0$ centered at the origin.

Theorem 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ is positive definite. Then the random variable $Y = (\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}(\mathbf{X}-\boldsymbol{\mu})$ has a $\chi^2(n)$ distribution. (Notice that Y is a single random variable and not a random vector.)

Note. The next theorem shows that a linear transformation of a multivariate normal random vector has, itself, a multivariate normal distribution.

Theorem 3.5.2. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{Y} = \mathbf{A}\mathbf{X}+\mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution.

Note. For $\mathbf{X}$ with a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, let $\mathbf{X}_1$ be any subvector of $\mathbf{X}$ of dimension, say, $m < n$. We can rearrange the components of $\mathbf{X}$ (and correspondingly, of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) so that we have $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ where $\mathbf{X}_2$ is of dimension $p = n - m$. This then leads to the partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ (where $\boldsymbol{\Sigma}_{11}$ is $m \times m$, $\boldsymbol{\Sigma}_{22}$ is $p \times p$, etc.). These decompositions can be done without loss of generality since the order of he random variables X_i in $\mathbf{X}$ is not relevant. For details on partitioned matrices, see my online notes for Theorem of Matrices (MATH 5090) on Section 3.2. Multiplication of Matrices and Multiplication of Vectors and Matrices.

Corollary 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are m dimensional and $\boldsymbol{\Sigma}_{\mathbf{11}}$ is $m \times m$. Then $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution.

Note. We define marginal probability density functions in the setting of two variables in Section 2.1. Distributions of Two Random Variables (see Note 2.1.C). Similarly, we see that Corollary 3.5.1 implies that any marginal distribution of normal random vector $\mathbf{X}$ is itself normal. Also, the mean and covariance of the marginal distribution are those associated with the partial vector $\mathbf{X}_1$.

Note. In Section 2.5. The Correlation Coefficient we see that if two random variables X and Y are independent, then $\text{Cov}(X, Y) = 0$ (see Theorem 3.5.2). In the setting of two random variables, the converse does not hold (see Example 2.5.3). The next result shows in the multivariate normal setting, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. Notice this implies in the two random variable setting when X and Y are both normally distributed and $\text{Cov}(X, Y) = 0$, then X and Y are independent.

Theorem 3.5.3. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Note. In Corollary 3.5.1 we saw that a marginal distribution of a multivariate normal random vector is itself normal. The next result shows that this also holds for conditional distributions.

Theorem 3.5.4. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Assume that $\boldsymbol{\Sigma}$ is positive definite. Then the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is

$$N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Example 3.5.A. Continuation of the Bivariate Normal Distribution.
Suppose $\mathbf{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$ and has a $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution where $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$. By Definition 2.5.2, the correlation coefficient between X and Y is $\rho = \dfrac{\Sigma_{12}}{\sigma_1\sigma_2} = \dfrac{\Sigma_{21}}{\sigma_1\sigma_2}$ or $\Sigma_{12} = \Sigma_{21} = \rho\sigma_1\sigma_2$. So

$$\det(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}| = \begin{vmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{vmatrix} = \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1 - \rho^2).$$

Recall that $-1 \leq \rho \leq 1$. If $\rho = \pm 1$ then $Y = \pm X$ (respectively) and we then effectively have a single variable. So we now assume $\rho^2 \neq 1$. We also have that X and Y are not constant (a constant random variable is not normally distributed) and so $\sigma_1 \neq 0 \neq \sigma_2$. So $\det(\mathbf{\Sigma}) \neq 0$ and hence $\mathbf{\Sigma}$ is 2×2, then

$$\mathbf{\Sigma}^{-1} = \frac{1}{\det(\mathbf{\Sigma})} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

(see my online notes for Linear Algebra [MATH 2010] on Section 4.3. Computation of Determinants and Cramer's Rule; see the example based on exercise 4.3.18). By equation (3.5.16) in Note 3.5.F, with $n = 2$ we have that the probability density function of $\mathbf{X}$ is

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{(2\pi)^{(2)/2}(\sigma_1^2\sigma_2^2(1-\rho^2))^{1/2}} \exp\left(\frac{-1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2}[x-\mu_1,\ \ y-\mu_2]\frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\right. \\
&\qquad \left. \times \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x-\mu_1 \\ x-\mu_2 \end{bmatrix}\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right. \\
&\qquad \left. \times[(x-\mu_1)\sigma_2^2 + (y-\mu_2)(-\rho\sigma_1\sigma_2),\ \ (x-\mu_1)(-\rho\sigma_1\sigma_2) + (y-\mu_2)\sigma_1^2] \begin{bmatrix} x-\mu_1 \\ x-\mu_2 \end{bmatrix}\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2\sigma_1^2\sigma_2^2(1-\rho^2)}((x-\mu_1)^2\sigma_2^2\right. \\
&\qquad \left. -(y-\mu_2)\rho\sigma_1\sigma_2(x-\mu_1) - (x-\mu_1)\rho\sigma_1\sigma_2(y-\mu_2) + (y-\mu_2)2\sigma_1^2\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}
\end{aligned}$$

where

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

That is, the multivariate normal distribution reduces to the bivariate normal distribution when $n = 2$ (compare to the first definition in this section).

Example 3.5.A Continued. As above we take $\mathbf{X} = (Y, X)'$ to have a $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (with the entries of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ permuted as appropriate to correspond to the entries of $\mathbf{X}$) we have by Theorem 3.5.4 that conditional distribution of Y given $X = x$ is $N(\mu_1 + \rho\sigma_2/\sigma_1(x - \mu_1), \sigma_2^2(1 - \rho^2))$ since $\boldsymbol{\Sigma}_{12} = \rho\sigma_1\sigma_2$ (as shown above) and $\boldsymbol{\Sigma}_{22}^{-1} = 1/\sigma_1^2$ so that $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} = \rho\sigma_1/\sigma_2$, and $\boldsymbol{\Sigma}_{11} = \sigma_2^2$ so that

$$\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \sigma_2^2 - (\rho\sigma_1\sigma_2)(1/\sigma_1^2)(\rho\sigma_1\sigma_2) = \sigma_2^2(1 - \rho^2).$$

We now have that the conditional mean of Y given $X = x$ is $E(Y \mid x) = \mu_2 + (\rho\sigma_2/\sigma_1)(x - \mu_1)$; notice that this is a linear function of x (unless $\rho = 0$). Notice that the variance is $\sigma_2^2(1-\rho^2)$ and is independent of x. It follows that "most" of the probability distribution of X and Y lies "close" to the line $y = \mu_2+(\rho\sigma_2/\sigma_1)(x-\mu_1)$ (provided the standard deviation of Y, $\sigma_2\sqrt{1 - \rho 2}$, is "small"). We can interchange X and Y to get that the conditional distribution of X given $Y = y$ is $N_2(\mu_1 + (\rho\sigma_1/\sigma_2)(y - \mu_2), \sigma_1^2(1 - \rho^2))$, and that similar conclusions as above can be drawn.

Exercise 3.5.8. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 20$, $\mu_2 = 40$, $\sigma_1^2 = 9$, $\sigma_2^2 = 4$, and $\rho = 0.6$. Find the shortest interval for which 0.90 is the conditional probability that Y is in the interval, given that $x = 22$.

Note. Let the random vectors $\mathbf{X}$ have the multivariate distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where where $\boldsymbol{\Sigma}$ is positive definite (and symmetric since $\mathrm{cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$). Then by Note 3.5.D, $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of $\boldsymbol{\Sigma}$ and the columns of $\boldsymbol{\Gamma}'$ are the unit eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ corresponding to the eigenvalues. Recall that $\boldsymbol{\Gamma}$ is an orthogonal matrix, so that $\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}^{-1}$. So $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'$. Define random vector $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$. By Theorem 3.5.2, $\mathbf{Y}$ has a $N_n(\mathbf{0}, \boldsymbol{\Lambda})$ distribution. Since $\boldsymbol{\Lambda}$ is a diagonal matrix then $\mathrm{cov}(Y_i, Y_j) = 0$ and it follows from Theorem 3.5.3 that $Y_1, Y_2, \ldots, Y_n$ are independent random variables. Also, for $i = 1, 2, \ldots, n$ random variables Y_i has a $N(0, \lambda_i)$ distribution.

Definition. Let random vector $\mathbf{X}$ have the multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. With $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}$ as the "spectral decomposition" (or orthogonal diagonalization) of $\boldsymbol{\Sigma}$, the random vector $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ is the vector of *principal components* of $\mathbf{X}$.

Definition. The *total variation* ("TV") of (any) random vector is the sum of the variances of its components.

Note. Consider random vector $\mathbf{X}$ with the multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ described above. Then

$$\begin{aligned} \mathrm{TV}(\mathbf{X}) &= \sum_{i=1}^{n} \sigma_i^2 \text{ by definition} \\ &= \mathrm{tr}(\boldsymbol{\Sigma}) \text{ since the trace of a matrix is the sum of the diagonal entries} \end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}(\mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}) \text{ since } \mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma} \\
&= \operatorname{tr}(\mathbf{\Gamma}'(\mathbf{\Lambda}\mathbf{\Gamma})) = \operatorname{tr}((\mathbf{\Lambda}\mathbf{\Gamma})\mathbf{\Gamma}') \text{ since } \operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA}) \text{ for appropriate} \\
&\quad \text{sized matrices (this follows from the definition of trace and} \\
&\quad \text{matrix product)} \\
&= \operatorname{tr}(\mathbf{\Lambda}\mathbf{I}_n) \text{ since } \mathbf{\Gamma} \text{ is an orthogonal matrix, so } \mathbf{\Gamma}^{-1} = \mathbf{\Gamma}' \\
&= \operatorname{tr}(\mathbf{\Lambda}) = \sum_{i=1}^{n} \lambda_i \\
&= \operatorname{TV}(\mathbf{Y} \text{ since } \mathbf{Y} \text{ has a } N_n(\mathbf{0}, \mathbf{\Lambda}) \text{ distribution.}
\end{aligned}
$$

Therefore $\mathbf{X}$ and $\mathbf{Y} = \mathbf{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ have the same total variance.

Lemma 3.5.B. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathbf{Y} = \mathbf{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\mathbf{\Gamma}$ is an orthogonal positive definite matrix. Then for any $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\operatorname{Var}(\mathbf{a}'\mathbf{X}) \le \operatorname{Var}(Y_1)$. That is, Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ where $\|\mathbf{a}\| = \|\mathbf{a}'\| = 1$.

Definition. For random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathbf{Y} = \mathbf{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\mathbf{\Gamma}$ is an orthogonal positive definite matrix, random variable Y_1 (the first component of $\mathbf{Y}$ as described in Lemma 3.5.A) is the *first principal component* of $\mathbf{X}$.

Note. The next result is a generalization of Lemma 3.5.A. The proof is similar to that of the lemma and is to be given in Exercise 3.5.20.

Theorem 3.5.5. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\boldsymbol{\Gamma}$ is an orthogonal positive definite matrix. For $j = 1, 2, \ldots, n$ we have $\mathrm{Var}(\mathbf{a}'\mathbf{X}) \le \lambda_j = \mathrm{Var}(Y_j)$ for all vectors $\mathbf{a}$ such that $\mathbf{a} \perp \mathbf{v}_i$ for $i = 1, 2, \ldots, j-1$ and $\|\mathbf{a}\| = 1$.

Definition. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\boldsymbol{\Gamma}$ is an orthogonal positive definite matrix. The components $Y_1, Y_2, \ldots, Y_n$ are the *second*, *third*, through the *nth principal components* of $\mathbf{X}$, respectively.

Note. "Principal Component Analysis" ("PCA") is an area of study in multivariate statistical analysis. It involves reducing the number of dimensions of observational data, or the weighting of components of a random vector in producing a standardized linear combination ("SLC") of the components. See my online notes (in preparation) for Applied Multivariate Statistical Analysis (STAT 5730); see "Chapter 11. Principal Components Analysis."

Revised: 7/21/2021