L E C T U R E 2

Example 1. A selection of four receipts from a university bookstore was obtained in order to investigate the nature of book sales. Each receipt provided, among other things, the number of books sold and the total amount of each sale. Let the first variable be total dollar sales and the second variable be number of books sold. Then we regard the corresponding numbers on the receipts as four measurements on two variables. Suppose the data, in tabular form, are

Variable 1 (dollar sales):	42	52	48	58
Variable 2 (number of books):	4	5	4	3

Using the notation just introduced, we have

$$x_{11} = 42, \quad x_{21} = 52, \quad x_{31} = 48, \quad x_{41} = 58$$

$$x_{12} = 4, \quad x_{22} = 5, \quad x_{32} = 4, \quad x_{42} = 3$$

and the data matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$$

with four rows and two columns.

Example 2.(The matrices $\bar{\mathbf{x}}$, $\mathbf{S}_n$ and $\mathbf{R}$ for bivariate data)
Consider the data introduced in Example 1. Each receipt yields a pair of measurements, total dollar sales, and number of books sold. Find the matrices $\bar{\mathbf{x}}$, $\mathbf{S}_n$ and $\mathbf{R}$. Since there are four receipts, we have a total of four measurements (observations) on each variable.

The sample means are

$$\bar{x}_1 = \frac{1}{4}\sum_{j=1}^{4} x_{j1} = \frac{1}{4}(42 + 52 + 48 + 58) = 50,$$

$$\bar{x}_2 = \frac{1}{4}\sum_{j=1}^{4} x_{j2} = \frac{1}{4}(4 + 5 + 4 + 3) = 4,$$

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} 50 \\ 4 \end{bmatrix}.$$

The sample variances and covariances are

$$s_{11} = \frac{1}{4}\sum_{j=1}^{4}(x_{j1} - \bar{x}_1)^2 = \frac{1}{4}((42-50)^2 + (52-50)^2 + (48-50)^2 + (58-50)^2) = 34,$$

$$s_{22} = \frac{1}{4}\sum_{j=1}^{4}(x_{j2} - \bar{x}_2)^2 = \frac{1}{4}((4-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2) = 0.5,$$

$$s_{12} = \frac{1}{4}\sum_{j=1}^{4}(x_{j1} - \bar{x}_1)(x_{j2} - \bar{x}_2) =$$

$$= \frac{1}{4}((42-50)(4-4) + (52-50)(5-4) + (48-50)(4-4) + (58-50)(3-4)) = -1.5,$$

$$s_{12} = s_{21}.$$

and

$$\mathbf{S}_n = \begin{bmatrix} 34 & -1.5 \\ -1.5 & 0.5 \end{bmatrix}.$$

The sample correlation is

$$r_{12} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} = \frac{-1.5}{\sqrt{34}\sqrt{0.5}} = -0.36,$$

$$r_{12} = r_{21},$$

so

$$\mathbf{R} = \begin{bmatrix} 1 & -0.36 \\ -0.36 & 1 \end{bmatrix}.$$

§4. DISTANCE

Although they may at first appear formidable, most multivariate techniques are based upon the simple concept of distance. Straight-line, or Euclidean, distance should be familiar. If we consider the point $P = (x_1, x_2)$ in the plane, the straight-line distance, $d(0, P)$, from P to the origin $O = (0, 0)$ is, according to the Pythagorean theorem,

$$d(O, P) = \sqrt{x_1^2 + x_2^2} \tag{1.7}$$

In general, if the point P has m coordinates so that $P = (x_1, x_2, \ldots, x_m)$, the straight-line distance from P to the origin $O = (0, 0, \ldots, 0)$ is

$$d(0, P) = \sqrt{x_1^2 + x_2^2 + \ldots + x_m^2}$$

All points $(x_1, x_2, \ldots, x_m)$ that lie a constant squared distance, such as c^2, from the origin satisfy the equation

$$d^2(0, P) = x_1^2 + x_2^2 + \ldots + x_m^2 = c^2 \tag{1.8}$$

Because this is the equation of a hypersphere (a circle if $m = 2$), points equidistant from the origin lie on a hypersphere.

The straight-line distance between two arbitrary points P and Q with coordinates $P = (x_1, x_2, \ldots, x_m)$ and $Q = (y_1, y_2, \ldots, y_m)$ is given by

$$d(P, Q) = \sqrt{(x_l - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_m - y_m)^2} \tag{1.9}$$

Straight-line, or Euclidean, distance is unsatisfactory for most statistical purposes. This is because each coordinate contributes equally to the calculation of Euclidean distance. When the coordinates represent measurements that are subject to random fluctuations of differing magnitudes, it is often desirable to weight coordinates subject to a great deal of variability less heavily than those that are not highly variable. This suggests a different measure of distance. Our purpose now is to develop a "statistical" distance that accounts for differences in variation and, in due course, the presence of correlation. Because our choice will depend upon the sample variances and covariances, at this point we use the term statistical distance to distinguish it from ordinary Euclidean distance. It is statistical distance that is fundamental to multivariate analysis.

To begin, we take as fixed the set of observations graphed as the m-dimensional scatter plot. From these, we will construct a measure of distance from the origin to a point $P = (x_1, x_2, \ldots, x_m)$. In our arguments, the coordinates $(x_1, x_2, \ldots, x_m)$ of P can vary to produce different locations for the point. The data that determine distance will, however, remain fixed.

To illustrate, suppose we have n pairs of measurements on two variables each having mean zero. Call the variables x_1 and x_2, and assume that the x_1 measurements vary independently of the x_2 measurements. In addition, assume that the variability in the x_1 measurements is larger than the variability in the x_2 measurements.

Let us assume that the inherent variability in the x_1 direction is greater than the variability in the x_2 direction. Consequently, large x_1 coordinates (in absolute value) are not as unexpected as large x_2 coordinates. It seems reasonable, then, to weight an x_2 coordinate more heavily than an x_1 coordinate of the same value when computing the "distance" to the origin. One way to proceed is to divide each coordinate by the sample standard deviation. Therefore, upon division by the standard deviations, we have the "standardized" coordinates $x_1^* = x_1/\sqrt{s_{11}}$ and $x_2^* = x_2/\sqrt{s_{22}}$. The standardized coordinates are now on an equal footing with one another. After taking the differences in variability into account, we determine distance using the standard Euclidean formula. Thus, a statistical distance of the point $P = (x_1, x_2)$ from the origin $\mathbf{O} = (0, 0)$ can be computed from its standardized coordinates $x_1^* = x_1/\sqrt{s_{11}}$ and $x_2^* = x_2/\sqrt{s_{22}}$ as

$$d(O, P) = \sqrt{(x_1^*)^2 + (x_2^*)2} = \sqrt{\left(x_1/\sqrt{s_{11}}\right)^2 + \left(x_2/\sqrt{s_{22}}\right)^2} = \sqrt{x_1^2/s_{11} + x_2^2/s_{22}} \tag{1.10}$$

Comparing (1.10) with (1.7), we see that the difference between the two expressions is due to the weights $k_1 = 1/s_{11}$ and $k_2 = 1/s_{22}$ attached to x_1^2 and x_2^2 in (1.10). Note that if the sample variances are the same, $k_1 = k_2$, and x_1^2 and x_2^2 will receive the same weight. In cases where the weights are the same, it is convenient to ignore the common divisor and use the usual Euclidean distance formula. In other words, if the variability in the x_1 direction is the same as the variability in the x_2 direction, and the x_1 values vary independently of the x_2 values, Euclidean distance is appropriate. Using (1.10), we see that all points which have coordinates (x_1, x_2) and are a constant squared distance c^2 from the origin must satisfy

$$x_1^2/s_{11} + x_2^2/s_{22} = c^2. \tag{1.11}$$

Equation (1.11) is the equation of an ellipse centered at the origin whose major and minor axes coincide with the coordinate axes. That is, the statistical distance in (1.10) has an ellipse as the locus of all points a constant distance from the origin.

Example 3. (Calculating a statistical distance) A set of paired measurements (x_1, x_2) on two variables yields $\overline{x_1} = \overline{x_2} = 0$, $s_{11} = 4$, and $s_{22} = 1$. Suppose the x_1 measurements are

unrelated to the x_2 measurements; that is, measurements within a pair vary independently of one another. Since the sample variances are unequal, we measure the square of the distance of an arbitrary point $P = (x_1, x_2)$ to the origin $O = (0, 0)$ by

$$d^2(O, P) = \frac{x_1^2}{4} + \frac{x_2^2}{1}.$$

Al points (x_1, x_2) that are a constant distance 1 from the origin satisfy the equation

$$\frac{x_1^2}{4} + \frac{x_2^2}{1} = 1.$$

The coordinates of some points a unit distance from the origin are presented in the following table:

Coordinates: (x_1, x_2)	Distance: $\frac{x_1^2}{4} + \frac{x_2^2}{1} = 1$
(0, 1)	$\frac{0^2}{4} + \frac{1^2}{1} = 1$
(0, -1)	$\frac{0^2}{4} + \frac{(-1)^2}{1} = 1$
(2, 0)	$\frac{2^2}{4} + \frac{0^2}{1} = 1$
$(1, \sqrt{3}/2)$	$\frac{1^2}{4} + \frac{(\sqrt{3}/2)^2}{1} = 1$

A plot of the equation $x_1^2/4 + x_2^2/1 = 1$ is an ellipse centered at (0, 0) whose major axis lies along the x_1 coordinate axis and whose minor axis lies along the x_2 coordinate axis. The half-lengths of these major and minor axes are $\sqrt{4} = 2$ and $\sqrt{1} = 1$, respectively. All points on the ellipse are regarded as being the same statistical distance from the origin – in this case, a distance of 1.

The expression in (1.11) can be generalized to accommodate the calculation of statistical distance from an arbitrary point $P = (x_1, x_2)$ to any fixed point $Q = (y_1, y_2)$. If we assume that the coordinate variables vary independently of one another, the distance from P to Q is given by

$$d(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}}} \tag{1.12}$$

The extension of this statistical distance to more than two dimensions is straightforward. Let the points P and Q have m coordinates such that $P = (x_1, x_2, \ldots, x_m)$ and $Q = (y_1, y_2, \ldots, y_m)$. Suppose Q is a fixed point (it may be the origin $0 = (0, 0, \ldots, 0)$)

and the coordinate variables vary independently of one another. Let $s_{11}, s_{22}, \ldots, s_{mm}$ be sample variances constructed from n measurements on $x_1, x_2, \ldots, x_m$, respectively. Then the statistical distance from P to Q is

$$d(P,Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \cdots + \frac{(x_m - y_m)^2}{s_{mm}}} \tag{1.13}$$

All points P that are a constant squared distance from Q lie on a hyperellipsoid centered at Q whose major and minor axes are parallel to the coordinate axes. We note the following:

1. The distance of P to the origin 0 is obtained by setting $y_1 = y_2 = \ldots = y_m = 0$ in (1.13).

2. If $s_{11} = s_{22} = \ldots = s_{mm}$ the Euclidean distance formula in (1.9) is appropriate.

The distance in (1.13) still does not include most of the important cases we shall encounter, because of the assumption of independent coordinates. In fact, the coordinates of the pairs (x_1, x_2) exhibit a tendency to be large or small together, and the sample correlation coefficient is positive. Moreover, the variability in the x_2 direction is larger than the variability in the x_1 direction.

What is a meaningful measure of distance when the variability in the x_1 direction is different from the variability in the x_2 direction and the variables x_1 and x_2 are correlated? Actually, we can use what we have already introduced, provided that we look at things in the right way. We see that if we rotate the original coordinate system through the angle (while keeping the scatter fixed and label the rotated axes $\widetilde{x}_1$ and $\widetilde{x}_2$, the scatter in terms of the new axes.

This suggests that we calculate the sample variances using the $\widetilde{x}_1$ and $\widetilde{x}_2$ coordinates and measure distance as in Equation (1.10). That is, with reference to the $\widetilde{x}_1$ and $\widetilde{x}_2$ axes, we define the distance from the point $p = (\widetilde{x}_1, \widetilde{x}_2)$ to the origin $0 = (0, 0)$ as

$$d(O,P) = \sqrt{\frac{\widetilde{x}_1^2}{\widetilde{s}_{11}} + \frac{\widetilde{x}_2^2}{\widetilde{s}_{22}}} \tag{1.14}$$

where $\widetilde{s}_{11}$ and $\widetilde{s}_{22}$ denote the sample variances computed with the $\widetilde{x}_1$ and $\widetilde{x}_2$ measurements.

The relation between the original coordinates (x_1, x_2) and the rotated coordinates $(\widetilde{x}_1, \widetilde{x}_2)$ is provided by

$$\widetilde{x}_1 = x_1 \cos\theta + x_2 \sin\theta$$

$$\widetilde{x}_2 = -x_1 \sin\theta + x_2 \cos\theta \tag{1.15}$$

Given the relations in (1.15), we can formally substitute for $\widetilde{x}_1$ and $\widetilde{x}_2$ in (1.14) and express the distance in terms of the original coordinates.

After some straightforward algebraic manipulations, the distance from $P = (\widetilde{x}_1, \widetilde{x}_2)$ to the origin $0 = (0, 0)$ can be written in terms of the original coordinates x_1 and x_2 of P as

$$d(O, P) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}, \tag{1.16}$$

where the a's are numbers such that the distance is nonnegative for all possible values of x_1 and x_2. Here a_{11}, a_{12}, and a_{22} are determined by the angle θ, and s_{11}, s_{12}, and s_{22} calculated from the original data. The particular forms for a_{11}, a_{12}, and a_{22} are not important at this point. What is important is the appearance of the crossproduct term $2a_{12}x_1x_2$ necessitated by the nonzero correlation r_{12}.

Equation (1.16) can be compared with (1.10). The expression in (1.10) can be regarded as a special case of (1.16) with $a_{11} = 1/s_{11}$, $a_{22} = 1/s_{22}$, and $a_{12} = 0$.

In general, the statistical distance of the point $P = (x_1, x_2)$ from the fixed point $Q = (y_1, y_2)$ for situations in which the variables are correlated has the general form

$$d(P, Q) = \sqrt{a_{11}(x_1 - y_1)^2 + 2a_{12}(x_1 - y_1)(x_2 - y_2) + a_{22}(x_2 - y_2)^2} \tag{1.17}$$

and can always be computed once a_{11}, a_{12}, and a_{22} are known. In addition, the coordinates of all points $P = (x_1, x_2)$ that are a constant squared distance c^2 from Q satisfy

$$a_{11}(x_1 - y_l)^2 + 2a_{12}(x_1 - y_l)(x_2 - y_2) + a_{22}(x_2 - y_2)^2 = c^2. \tag{1.18}$$

By definition, this is the equation of an ellipse centered at Q. The major (long) and minor (short) axes are indicated. They are parallel to the $\widetilde{x}_1$ and $\widetilde{x}_2$ axes. For the choice of a_{11}, a_{12}, and a_{22}, the $\widetilde{x}_1$ and $\widetilde{x}_2$ axes are at an angle (with respect to the x_1 and x_2 axes.

The generalization of the distance formulas of (1.16) and (1.17) to m dimensions is straightforward. Let $P = (x_1, x_2, \ldots, x_m)$ be a point whose coordinates represent variables that are correlated and subject to inherent variability. Let $0 = (\ 0,\ 0,\ .\ .\ .\ ,\ 0)$ denote the origin, and let $Q = (y_1, y_2, \ldots, y_m)$ be a specified fixed point. Then the distances from P to 0 and from P to Q have the general forms

$$d(O,P) = \sqrt{a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{pp}x_p^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \ldots + 2a_{m-1,m}x_{m-1}x_m} \tag{1.19}$$

and

$$d(P,Q) = \sqrt{\sum_{i=1}^{m}[a_{ii}(x_i - y_i)^2] + 2\sum_{i<j,i,j=1}^{m} a_{ij}(x_i - y_i)(x_j - y_j)} \tag{1.20}$$

where the a's are numbers such that the distances are always nonnegative.

Th algebraic expressions for the squares of the distances in (1.18) and (1.19) are known as quadratic forms and, in particular, positive definite quadratic forms.

We note that the distances in (1.19) and (1.20) are completely determined by the coefficients (weights) $a_{ik},\ i = 1, 2, \ldots, m,\ k = 1, 2, \ldots, m$. These coefficients can be set out in the rectangular array

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{12} & a_{22} & \ldots & a_{2m} \\ \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ a_{1m} & a_{2m} & \ldots & a_{mm} \end{bmatrix} \tag{1.21}$$

where the a_{ik}'s with i /k are displayed twice, since they are multiplied by 2 in the distance formulas. Consequently, the entries in this array specify the distance functions. The a_{ik}'s cannot be arbitrary numbers; they must be such that the computed distance is nonnegative for every pair of points. Contours of constant distances computed from (1.19) and (1.20) are hyperellipsoids.

The need to consider statistical rather than Euclidean distance is illustrated heuristically, which depicts a cluster of points whose center of gravity (sample mean) is indicated by the point Q. Consider the Euclidean distances from the point Q to the point P and the origin 0. The Euclidean distance from Q to P is larger than the Euclidean distance from

Q to 0. However, P appears to be more like the points in the cluster than does the origin. If we take into account the variability of the points in the cluster and measure distance by the statistical distance in (1.16), then Q will be closer to P than to 0. This result seems reasonable given the nature of the scatter.

Other measures of distance can be advanced. At times, it is useful to consider distances that are not related to circles or ellipses. Any distance measure $d(P,Q)$ between two points P and Q is valid provided that it satisfies the following properties, where R is any other intermediate point:

$$d(P,Q) = d(Q,P)$$

$$d(P,Q) > 0 \quad \text{if} \quad P \neq Q$$

$$d(P,Q) = 0 \quad \text{if} \quad P = Q$$

$$d(P,Q) \leq d(P,R) + d(R,Q) \quad \text{(triangle inequality)}$$

FINAL COMMENTS.

We have attempted to motivate the study of multivariate analysis and to provide you with some rudimentary, but important, methods for organizing, summarizing, and displaying data. In addition, a general concept of distance has been introduced that will be used repeatedly in later chapters.

§5. BIVARIATE NORMAL DISTRIBUTION

A very important two–dimensional probability law which is a generalization of the one–dimensional normal probability law is called *Bivariate normal distribution*.

The random variables $\eta_1(\omega)$ and $\eta_2(\omega)$ are said to have a Bivariate normal distribution with parameters $(a_1, a_2, \sigma_1, \sigma_2, \rho)$ if their joint density function is given by

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-a_1}{\sigma_1}\right)^2 - 2\rho\cdot\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-a_2}{\sigma_2}\right)^2\right]\right\}. \tag{1.22}$$

We see that Bivariate normal distribution is determined by 5 parameters. These are a_1, a_2, σ_1, σ_2 and ρ such that $a_1 \in (-\infty, +\infty)$, $a_2 \in (-\infty, +\infty)$, $\sigma_1 > 0$, $\sigma_2 > 0$ and $\rho \in (-1, +1)$.

We know that the marginal density function of η_1 is defined by the following formula

$$f_{\eta_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{(x_1 - a_1)^2}{2\sigma_1^2}\right).$$

Similarly, for random variable η_2, we obtain:

$$f_{\eta_2}(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left(-\frac{(x_2 - a_2)^2}{2\sigma_2^2}\right).$$

Therefore, $\eta_1(\omega)$ and $\eta_2(\omega)$ are both normal random variables with respective parameters $\mathcal{N}(a_1, \sigma_1)$ and $\mathcal{N}(a_2, \sigma_2)$.

Thus, the marginal distributions of $\eta_1(\omega)$ and $\eta_2(\omega)$ are both normal, even though the joint distributions for $\rho = 0$, $\rho = 0.3$, $\rho = 0.6$ and $\rho = 0.9$ are quite different from each other. It is not difficult to verify that

$$\mathrm{Cov}\,(\eta_1, \eta_2) = E\,[(\eta_1 - a_1) \cdot (\eta_2 - a_2)] = \rho\,\sigma_1\sigma_2,$$

i. e. covariance between $\eta_1(\omega)$ and $\eta_2(\omega)$ is equal to $\rho\,\sigma_1\,\sigma_2$ and therefore

$$\rho(\eta_1, \eta_2) = \rho.$$

Thus, we have proved that

η_1 and η_2 are independent if and only if the correlation is equal to 0.

Indeed, substituting $\rho = 0$ in (1.21) we obtain

$$f(x_1, x_2) = f_{\eta_1}(x_1) \cdot f_{\eta_2}(x_2).$$

Definition 1. If $\eta_1(\omega)$ and $\eta_2(\omega)$ have a joint density function $f(x_1, x_2)$, then the conditional density function of $\eta_1(\omega)$, given that $\eta_2(\omega) = x_2$ is defined for all values of x_2 such that $f_{\eta_2}(x_2) \neq 0$, by

$$f_{\eta_1/\eta_2}(x_1|x_2) = \frac{f(x_1, x_2)}{f_{\eta_2}(x_2)}.$$

To motivate this definition, multiply the left–hand side by dx_1 and the right–hand side by $(dx_1 \cdot dx_2)/(dx_2)$ to obtain

$$f_{\eta_1/\eta_2}(x_1|x_2)dx_1 = \frac{f(x_1, x_2)\, dx_1\, dx_2}{f_{\eta_2}(x_2)\, dx_2} \approx \frac{P\{\omega\colon\ x_1 \le \eta_1(\omega) \le x_1 + dx_1, x_2 \le \eta_2(\omega) \le x_2 + dx_2\}}{P\{\omega\colon\ x_2 \le \eta_2(\omega) \le x_2 + dx_2\}} =$$
$$= P\{\omega\colon\ x_1 \le \eta_1(\omega) \le x_1 + dx_1 | x_2 \le \eta_2(\omega) \le x_2 + dx_2\}.$$

In other words, for small values of dx_1 and dx_2, $f_{\eta_1/\eta_2}(x_1|x_2)$ represents the conditional probability that $\eta_1(\omega)$ is between x_1 and $x_1 + dx_1$, given that $\eta_2(\omega)$ is between x_2 and $x_2 + dx_2$.

One can show that the conditional density of η_1, given that $\eta_2 = x_2$, is the normal density with parameters

$$\mathcal{N}\left(a_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - a_2), \sigma_1^2(1-\rho^2)\right).$$

Similarly, the conditional density of η_2, given that $\eta_1 = x_1$, is the normal density with parameters

$$\mathcal{N}\left(a_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - a_1), \sigma_2^2(1-\rho^2)\right).$$

We can consider also nd m-dimensional normal distribution.

Definition 2. The random variables $\eta_1(\omega)$, $\eta_2(\omega)$, ... $\eta_m(\omega)$ are said to have m-variate normal distribution with parameters $(a_1, a_2, ..., a_m, \Sigma)$ if their joint density function is given by

$$\begin{aligned} f(x_1, x_2, ..., x_m) &= \frac{1}{(2\pi)^{m/2}\sqrt{|\Sigma|}} \times \\ &\times \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{i=1}^{m} c_{ij}(x_i - a_i)(x_j - a_j)\right\}, \end{aligned} \tag{1.23}$$

where $(a_1, a_2, ..., a_m)$ is the expectation of the vector $(\eta_1(\omega), \eta_2(\omega), ...\eta_m(\omega))$, $\Sigma = ||\mathrm{Cov}\,(\eta_i, \eta_j)||$ is the covariance matrix of $(\eta_1(\omega), \eta_2(\omega), ...\eta_m(\omega))$, $|\Sigma|$ is the determinant of covariance matrix and $C = ||c_{ij}|| = \Sigma^{-1}$ is the inverse matrix to Σ.

Below, for brevity, we will denote the m-variate normal vector as follows:

$$N_m(\mu, \Sigma),$$

where $\mu = (a_1, a_2, ..., a_m)$ is the mean vector.

Example 4. Let (η_1, η_2, η_3) be $N_3(\mu, \Sigma)$, where $\mu = (5, 3, 7)$ and

$$\Sigma = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{pmatrix}$$

Find

a) $P(\eta_1 > 6)$;

b) $P(5\eta_2 + 4\eta_3 > 70)$;

c) $P(4\eta_1 - 3\eta_2 + 5\eta_3 < 80)$.

Solution.

a) Since η_1 is $N(5, 4)$ then

$$P(\eta_1 > 6) = 1 - P(\eta_1 \leq 6) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085.$$

b) Since $5\eta_2 + 4\eta_3$ is $N(43, 324)$ we get

$$P(5\eta_2 + 4\eta_3 > 70) = 1 - P(5\eta_2 + 4\eta_3 \leq 70) = 1 - \Phi\left(\frac{70 - 43}{18}\right) = 1 - \Phi(1.5) = 0.0668.$$

c) Since $4\eta_1 - 3\eta_2 + 5\eta_3$ is $N(46, 289)$ we obtain

$$P(4\eta_1 - 3\eta_2 + 5\eta_3 < 80) = \Phi\left(\frac{80 - 46}{17}\right) = \Phi(2) = 0.9772.$$

Here we use the following formulae

$$\mathrm{Var}\left(\sum_{i=1}^{n} \eta_i\right) = \sum_{i=1}^{n} \mathrm{Var}(\eta_i) + 2\sum_{i<j} \mathrm{Cov}\,(\eta_i, \eta_j)$$

and

$$\mathrm{Var}\,(a\eta) = a^2\,\mathrm{Var}(\eta).$$