KTH Matematik

Multivariate Gaussian Distribution

Auxiliary for SF2955

2011

Timo Koski
Institutionen för matematik
Kungliga tekniska högskolan (KTH) , Stockholm

Chapter 1

Gaussian Vectors

1.1 Multivariate Gaussian Distribution

Let us recall the following;

- X is a normal random variable, if

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2},$$

  where μ is real and $\sigma > 0$.

- Notation: $X \in N(\mu, \sigma^2)$.

- Properties: $X \in N(\mu, \sigma^2) \Rightarrow E(X) = \mu$, $\text{Var} = \sigma^2$.

- $X \in N(\mu, \sigma^2)$, then the moment generating function is

$$\psi_X(t) = E\left[e^{tX}\right] = e^{t\mu + \frac{1}{2}t^2\sigma^2}, \tag{1.1}$$

  and the characteristic function is

$$\varphi_X(t) = E\left[e^{itX}\right] = e^{it\mu - \frac{1}{2}t^2\sigma^2}. \tag{1.2}$$

- $X \in N(\mu, \sigma^2) \Rightarrow Y = aX + b \in N(a\mu + b, a^2\sigma^2)$.

- $X \in N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \in N(0, 1)$.

We shall next see that all of these properties are special cases of the corresponding properties of a multivariate normal/Gaussian random variable as defined below, which bears witness to the statement that the normal distribution is central in probability theory.

1.1.1 Notation for Vectors, Mean Vector, Covariance Matrix & Characteristic Functions

An $n \times 1$ random vector or a multivariate random variable is denoted by

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \left(X_1, X_2, \ldots, X_n\right)^{'},$$

where $'$ is the vector transpose. A vector in $\mathbf{R}^n$ is designated by

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \left(x_1, x_2, \ldots, x_n\right)^{'}.$$

We denote by $F_{\mathbf{X}}(\mathbf{x})$ the joint distribution function of $\mathbf{X}$, which means that

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbf{P}(\mathbf{X} \leq \mathbf{x}) = \mathbf{P}\left(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\right).$$

The following definitions are natural. We have the **mean vector**

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{pmatrix},$$

which is a $n \times 1$ column vector of means (=expected values) of the components of $\mathbf{X}$.
The **covariance matrix** is a square $n \times n$ -matrix

$$C_{\mathbf{X}} := E\left[\left(\mathbf{X} - \mu_{\mathbf{X}}\right)\left(\mathbf{X} - \mu_{\mathbf{X}}\right)^{'}\right],$$

where the entry at the position (i,j) is

$$c_{i,j} \stackrel{\text{def}}{=} C_{\mathbf{X}}(i,j) = E\left[\left(X_i - \mu_i\right)\left(X_j - \mu_j\right)\right],$$

that is the covariance of X_i and X_j. Every covariance matrix, now designated by $\mathbf{C}$, is by construction symmetric

$$\mathbf{C} = \mathbf{C}^{'} \tag{1.3}$$

and nonnegative definite, i.e, for all $\mathbf{x} \in \mathbf{R^n}$

$$\mathbf{x}^{'}\mathbf{C}\mathbf{x} \geq 0. \tag{1.4}$$

It is shown in linear algebra that nonnegative definiteness is equivalent to $\det \mathbf{C} \geq 0$. In terms of the entries $c_{i,j}$ of a covariance matrix $\mathbf{C} = \left(c_{i,j}\right)_{i=1,j=1}^{n,n,}$ there are the following necessary properties.

1. $c_{i,j} = c_{j,i}$ (symmetry).
2. $c_{i,i} = \mathrm{Var}(X_i) = \sigma_i^2 \geq 0$ (the elements in the main diagonal are the variances, and thus all elements in the main diagonal are nonnegative).
3. $c_{i,j}^2 \leq c_{i,i} \cdot c_{j,j}$.

Example 1.1.1 The covariance matrix of a bivariate random variable $\mathbf{X} = (X_1, X_2)^{'}$ is often written in the following form

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{1.5}$$

where $\sigma_1^2 = \mathrm{Var}(X_1)$, $\sigma_2^2 = \mathrm{Var}(X_2)$ and $\rho = \mathrm{Cov}(X,Y)/(\sigma_1\sigma_2)$ is the coefficient of correlation of X_1 and X_2. C is invertible ($\Rightarrow$ positive definite) if and only if $\rho^2 \neq 1$.

■

Linear transformations of random vectors are Borel functions $\mathbf{R}^n \mapsto \mathbf{R}^m$ of random vectors. The rules for finding the mean vector and the covariance matrix of a transformed vector are simple.

Proposition 1.1.2 $\mathbf{X}$ is a random vector with mean vector $\mu_{\mathbf{X}}$ and covariance matrix $C_{\mathbf{X}}$. B is a $m \times n$ matrix. If $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, then

$$E\mathbf{Y} = B\mu_{\mathbf{X}} + \mathbf{b} \tag{1.6}$$

$$C_{\mathbf{Y}} = BC_{\mathbf{X}}B^{'}. \tag{1.7}$$

Proof For simplicity of writing, take $\mathbf{b} = \mu = \mathbf{0}$. Then

$$C_{\mathbf{Y}} = E\mathbf{Y}\mathbf{Y}^{'} = EB\mathbf{X}(B\mathbf{X})^{'} =$$

$$= EB\mathbf{X}\mathbf{X}^{'}B^{'} = BE\left[\mathbf{X}\mathbf{X}^{'}\right]B^{'} = BC_{\mathbf{X}}B^{'}.$$

■

We have from [4, def. 4.2 on p. 77].

Definition 1.1.1

$$\phi_{\mathbf{X}}(\mathbf{s}) \stackrel{\text{def}}{=} E\left[e^{i\mathbf{s}^{'}\mathbf{X}}\right] = \int_{\mathbf{R}^n} e^{i\mathbf{s}^{'}\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}) \tag{1.8}$$

is the **characteristic function** of the random vector $\mathbf{X}$.

■

In (1.8) $\mathbf{s}^{'}\mathbf{x}$ is a scalar product in $\mathbf{R}^n$,

$$\mathbf{s}^{'}\mathbf{x} = \sum_{i=1}^{n} s_i x_i.$$

As $F_{\mathbf{X}}$ is a joint distribution function on $\mathbf{R}^n$ and $\int_{\mathbf{R}^n}$ is a notation for a multiple integral over $\mathbf{R}^n$, we know that

$$\int_{\mathbf{R}^n} dF_{\mathbf{X}}(\mathbf{x}) = 1,$$

which means that $\phi_{\mathbf{X}}(\mathbf{0}) = 1$, where $\mathbf{0}$ is a $n \times 1$ -vector of zeros.

Theorem 1.1.3 *[***Kac's theorem***]* $\mathbf{X} = (X_1, X_2, \cdots, X_n)^{'}$. The components $X_1, X_2, \cdots, X_n$ are independent if and only if

$$\phi_{\mathbf{X}}(\mathbf{s}) = E\left[e^{i\mathbf{s}^{'}\mathbf{X}}\right] = \prod_{i=1}^{n} \phi_{X_i}(s_i),$$

where $\phi_{X_i}(s_i)$ is the characteristic function for X_i.

Proof Assume that $\mathbf{X} = (X_1, X_2, \cdots, X_n)^{'}$ is a vector with independent X_i, $i = 1, \ldots, n$, that have, for convenience of writing, a joint density $f_{\mathbf{X}}(\mathbf{x})$ we have in (1.8)

$$\phi_{\mathbf{X}}(\mathbf{s}) = \int_{\mathbf{R}^n} e^{i\mathbf{s}^{'}\mathbf{x}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

$$= \int_{\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i(s_1 x_1 + \ldots + s_n x_n)} \prod_{i=1}^{n} f_{X_i}(x_i)\, dx_1 \cdots dx_n \tag{1.9}$$

$$= \int_{\infty}^{\infty} e^{is_1 x_1} f_{X_1}(x_1)\, dx_1 \cdots \int_{-\infty}^{\infty} e^{is_n x_n} f_{X_n}(x_n)\, dx_n = \phi_{X_1}(s_1) \cdots \phi_{X_n}(s_n),$$

where $\phi_{X_i}(s_i)$ is the characteristic function for X_i. The rest of the proof is omitted.

■

1.1.2 Multivariate Normal Distribution

Definition 1.1.2 $\mathbf{X}$ has a multivariate normal distribution with mean vector μ and covariance matrix $\mathbf{C}$, written as $\mathbf{X} \in N(\mu, \mathbf{C})$, if and only if the characteristic function is given as

$$\phi_{\mathbf{X}}(\mathbf{s}) = e^{i\mathbf{s}^{'}\mu - \frac{1}{2}\mathbf{s}^{'}\mathbf{C}\mathbf{s}}. \tag{1.10}$$

■

Theorem 1.1.4 $\mathbf{X}$ has a multivariate normal distribution $N(\mu, \mathbf{C})$ if and only of

$$\mathbf{a}^{'}\mathbf{X} = \sum_{i=1}^{n} a_i X_i \tag{1.11}$$

has a normal distribution for **all** vectors $\mathbf{a}^{'} = (a_1, a_2, \ldots, a_n)$.

Proof Assume that $\mathbf{a}^{'}\mathbf{X}$ has a multivariate normal distribution for all $\mathbf{a}$ and that μ and $\mathbf{C}$ are the mean vector and covariance matrix of $\mathbf{X}$, respectively. Here (1.6) and (1.7) with $B = \mathbf{a}^{'}$ give

$$Ea^{'}\mathbf{X} = \mathbf{a}^{'}\mu, \mathrm{Var}\left[\mathbf{a}^{'}\mathbf{X}\right] = \mathbf{a}^{'}\mathbf{C}\mathbf{a}.$$

Hence, if we set $Y = \mathbf{a}^{'}\mathbf{X}$, then by assumption $Y \in N\left(\mathbf{a}^{'}\mu, \mathbf{a}^{'}\mathbf{C}\mathbf{a}\right)$ and the characteristic function of Y is by (1.2)

$$\varphi_Y(t) = e^{it\mathbf{a}^{'}\mu - \frac{1}{2}t^2\mathbf{a}^{'}\mathbf{C}\mathbf{a}}.$$

The characteristic function of $\mathbf{X}$ is by definition

$$\varphi_{\mathbf{X}}(\mathbf{s}) = Ee^{i\mathbf{s}^{'}\mathbf{X}}.$$

Thus

$$\varphi_{\mathbf{X}}(\mathbf{a}) = Ee^{i\mathbf{a}^{'}\mathbf{X}} = \varphi_Y(1) = e^{i\mathbf{a}^{'}\mu - \frac{1}{2}\mathbf{a}^{'}\mathbf{C}\mathbf{a}}.$$

Thereby we have established that the characteristic function of $\mathbf{X}$ is

$$\varphi_{\mathbf{X}}(\mathbf{s}) = e^{i\mathbf{s}^{'}\mu - \frac{1}{2}\mathbf{s}^{'}\mathbf{C}\mathbf{s}}.$$

In view of definition 1.1.2 this shows that $\mathbf{X} \in N(\mu, \mathbf{C})$. The proof of the statement in the other direction is obvious. ■

Example 1.1.5 In this example we study a bivariate random variable $(X, Y)^{'}$ such that both X and Y have normal marginal distribution but there is a linear combination (in fact, $X + Y$), which does not have a normal distribution. Therefore $(X, Y)^{'}$ is not a bivariate normal random variable.
Let $X \in N\left(0, \sigma^2\right)$. Let $U \in \mathrm{Be}\left(\frac{1}{2}\right)$ and be independent of X. Define

$$Y = \begin{cases} X & \text{if } U = 0 \\ -X & \text{if } U = 1. \end{cases}$$

Let us find the distribution of Y. We compute the characteristic function by double expectation

$$\begin{aligned}\varphi_Y(t) &= E\left[e^{itY}\right] = E\left[E\left[e^{itY} \mid U\right]\right] \\ &= E\left[e^{itY} \mid U = 0\right] \cdot \frac{1}{2} + E\left[e^{itY} \mid U = 1\right] \cdot \frac{1}{2} \\ &= E\left[e^{itX} \mid U = 0\right] \cdot \frac{1}{2} + E\left[e^{-itX} \mid U = 1\right] \cdot \frac{1}{2}\end{aligned}$$

and since X and U are independent, the independent condition drops out, and $X \in N\left(0, \sigma^2\right)$,

$$= E\left[e^{itX}\right] \cdot \frac{1}{2} + E\left[e^{-itX}\right] \cdot \frac{1}{2} = \frac{1}{2} \cdot e^{-\frac{t^2\sigma^2}{2}} + \frac{1}{2} \cdot e^{-\frac{t^2\sigma^2}{2}} = e^{-\frac{t^2\sigma^2}{2}},$$

which by uniqueness of characteristic functions says that $Y \in N\left(0, \sigma^2\right)$. Hence both marginal distributions of the bivariate random variable (X, Y) are normal distributions. Yet, the sum

$$X + Y = \begin{cases} 2X & \text{if } U = 0 \\ 0 & \text{if } U = 1 \end{cases}$$

is *not* a normal random variable. Hence (X, Y) is according to theorem 1.1.4 not a bivariate Gaussian random variable. Clearly we have

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & (-1)^U \end{pmatrix} \begin{pmatrix} X \\ X \end{pmatrix}. \tag{1.12}$$

Hence we multiply $(X, X)^{'}$ once by a random matrix to get $(X, Y)^{'}$ and therefore should not expect $(X, Y)^{'}$ to have a joint Gaussian distribution. We take next a look at the details. If $U = 1$, then

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X \\ X \end{pmatrix} = A_1 \begin{pmatrix} X \\ X \end{pmatrix}$$

and if $U = 0$,

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ X \end{pmatrix} = A_0 \begin{pmatrix} X \\ X \end{pmatrix}.$$

The covariance matrix of $(X, X)^{'}$ is clearly

$$\mathbf{C}_X = \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

We set

$$\mathbf{C}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{C}_0 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

One can verify , c.f. (1.7), that $\sigma^2 \mathbf{C}_1 = A_1 \mathbf{C}_X A_1^{'}$ and $\sigma^2 \mathbf{C}_0 = A_0 \mathbf{C}_X A_0^{'}$. Hence $\sigma^2 \mathbf{C}_1$ is the covariance matrix of (X, Y), if $U = 1$, and $\sigma^2 \mathbf{C}_0$ is the covariance matrix of (X, Y), if $U = 0$.

It is clear by the above that the joint distribution $F_{X,Y}$ should actually be a *mixture* of two distributions $F_{X,Y}^{(1)}$ and $F_{X,Y}^{(0)}$ with mixture coefficients $\left(\frac{1}{2}, \frac{1}{2}\right)$,

$$F_{X,Y}(x, y) = \frac{1}{2} \cdot F_{X,Y}^{(1)}(x, y) + \frac{1}{2} \cdot F_{X,Y}^{(0)}(x, y).$$

We understand this as follows. We draw first a value u from $\mathrm{Be}\left(\frac{1}{2}\right)$, which points out one of the distributions, $F_{X,Y}^{(u)}$, and then draw a sample of (X, Y) from $F_{X,Y}^{(u)}$. We can explore these facts further.

■

Additional properties are:

1. **Theorem 1.1.6** If $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, and $\mathbf{X} \in N(\mu, \mathbf{C})$, then

$$\mathbf{Y} \in N\left(B\mu + \mathbf{b}, B\mathbf{C}B^{'}\right).$$

Proof We check the characteristic function of $\mathbf{Y}$; some linear algebra gives

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = E\left[e^{i\mathbf{s}^{'}\mathbf{Y}}\right] = E\left[e^{i\mathbf{s}^{'}(\mathbf{b}+B\mathbf{X})}\right] =$$

$$= e^{i\mathbf{s}^{'}\mathbf{b}} E\left[e^{i\mathbf{s}^{'}B\mathbf{X}}\right] = e^{i\mathbf{s}^{'}\mathbf{b}} E\left[e^{i\left(B^{'}\mathbf{s}\right)^{'}\mathbf{X}}\right]$$

or

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = e^{i\mathbf{s}^{'}\mathbf{b}} E\left[e^{i\left(B^{'}\mathbf{s}\right)^{'}\mathbf{X}}\right]. \tag{1.13}$$

Here

$$E\left[e^{i\left(B^{'}\mathbf{s}\right)^{'}\mathbf{X}}\right] = \varphi_{\mathbf{X}}\left(B^{'}\mathbf{s}\right).$$

Furthermore

$$\varphi_{\mathbf{X}}\left(B^{'}\mathbf{s}\right) = e^{i\left(B^{'}\mathbf{s}\right)^{'}\mu - \frac{1}{2}\left(B^{'}\mathbf{s}\right)^{'}\mathbf{C}\left(B^{'}\mathbf{s}\right)}.$$

Since

$$\left(B^{'}\mathbf{s}\right)^{'}\mu = \mathbf{s}^{'}B\mu, \quad \left(B^{'}\mathbf{s}\right)^{'}\mathbf{C}\left(B^{'}\mathbf{s}\right) = \mathbf{s}^{'}B\mathbf{C}B^{'}\mathbf{s},$$

we get

$$e^{i\left(B^{'}\mathbf{s}\right)^{'}\mu - \frac{1}{2}\left(B^{'}\mathbf{s}\right)^{'}\mathbf{C}\left(B^{'}\mathbf{s}\right)} = e^{i\mathbf{s}^{'}B\mu - \frac{1}{2}\mathbf{s}^{'}B\mathbf{C}B^{'}\mathbf{s}}.$$

Therefore

$$\varphi_{\mathbf{X}}\left(B^{'}\mathbf{s}\right) = e^{i\mathbf{s}^{'}B\mu - \frac{1}{2}\mathbf{s}^{'}B\mathbf{C}B^{'}\mathbf{s}} \tag{1.14}$$

and by (1.14) and (1.13) above we get

$$\varphi_{\mathbf{Y}}(\mathbf{s}) = e^{i\mathbf{s}^{'}\mathbf{b}}\varphi_{\mathbf{X}}\left(B^{'}\mathbf{s}\right) = e^{i\mathbf{s}^{'}\mathbf{b}}e^{i\mathbf{s}^{'}B\mu - \frac{1}{2}\mathbf{s}^{'}B\mathbf{C}B^{'}\mathbf{s}}$$

$$= e^{i\mathbf{s}^{'}(\mathbf{b}+B\mu) - \frac{1}{2}\mathbf{s}^{'}B\mathbf{C}B^{'}\mathbf{s}},$$

which by uniqueness of characteristic functions proves the claim as asserted. ∎

2. **Theorem 1.1.7** A Gaussian multivariate random variable has independent components if and only if the covariance matrix is diagonal.

Proof Let $\mathbf{\Lambda}$ be a diagonal covariance matrix with λ_is on the main diagonal, i.e.,

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & \ddots & \vdots & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Then

$$\varphi_{\mathbf{X}}(\mathbf{t}) = e^{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\mathbf{\Lambda}\mathbf{t}} =$$

$$= e^{i\sum_{i=1}^{n}\mu_i t_i - \frac{1}{2}\sum_{i=1}^{n}\lambda_i t_i^2}$$

$$= e^{i\mu_1 t_1 - \frac{1}{2}\lambda_1 t_1^2} e^{i\mu_2 t_2 - \frac{1}{2}\lambda_2 t_2^2} \cdots e^{i\mu_n t_n - \frac{1}{2}\lambda_n t_n^2}$$

is the product of the characteristic functions of $X_i \in N(\mu_i, \lambda_i)$, which are by theorem 1.1.3 seen to be independent. ■

3. **Theorem 1.1.8** If $\mathbf{C}$ is positive definite ($\Rightarrow \det \mathbf{C} > 0$), then it can be shown that there is a simultaneous density of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \mathbf{C}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})'\mathbf{C}^{-1}(\mathbf{x}-\mu_{\mathbf{X}})}. \qquad (1.15)$$

Proof It can be checked by a lengthy but straightforward computation that

$$e^{i\mathbf{s}'\mu - \frac{1}{2}\mathbf{s}'\mathbf{C}\mathbf{s}} = \int_{\mathbf{R}^n} e^{i\mathbf{s}'\mathbf{x}} \frac{1}{(2\pi)^{n/2}\sqrt{\det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\mathbf{C}^{-1}(\mathbf{x}-\mu)} d\mathbf{x}.$$

■

4. **Theorem 1.1.9** $(X_1, X_2)'$ is a bivariate Gaussian random variable. The conditional distribution for X_2 given $X_1 = x_1$ is

$$N\left(\mu_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right), \qquad (1.16)$$

where $\mu_2 = E(X_2)$, $\mu_1 = E(X_2)$, $\sigma_2 = \sqrt{\mathrm{Var}(X_2)}$, $\sigma_1 = \sqrt{\mathrm{Var}(X_1)}$ and $\rho = \mathrm{Cov}(X_1, X_2)/(\sigma_1 \cdot \sigma_2)$.

Proof is done by an explicit evaluation of (1.15) followed by an explicit evaluation of the pertinent conditional density and is deferred to Appendix 1.4. ■

Definition 1.1.3 $\mathbf{Z} \in N(\mathbf{0}, I)$ is a standard Gaussian vector, where I is $n \times n$ identity matrix.

■

Let $\mathbf{X} \in N(\mu_{\mathbf{X}}, \mathbf{C})$. Then, if $\mathbf{C}$ is positive definite, we can factorize $\mathbf{C}$ as

$$\mathbf{C} = AA^{'},$$

for $n \times n$ matrix A, where A is lower triangular. Actually we can always decompose

$$\mathbf{C} = LDL',$$

where L is a unique $n \times n$ lower triangular, D is diagonal with positive elements on the main diagonal, and we write $A = L\sqrt{D}$. Then A^{-1} is lower triangular. Then

$$\mathbf{Z} = A^{-1}(\mathbf{X} - \mu_{\mathbf{X}})$$

is a standard Gaussian vector. In some applications, like, e.g., in time series analysis and signal processing, one refers to A^{-1} as a **whitening** matrix. It can be shown that A^{-1} is lower triangular, thus we have obtained $\mathbf{Z}$ by a **causal** operation, in the sense that Z_i is a function of $X_1, \ldots, X_i$. $\mathbf{Z}$ is known as the **innovations** of $\mathbf{X}$. Conversely, one goes from the innovations to $\mathbf{X}$ through another causal operation by $\mathbf{X} = A\mathbf{Z} + \mathbf{b}$, and then

$$\mathbf{X} = N\left(\mathbf{b}, AA^{'}\right).$$

Example 1.1.10 (Factorization of a 2×2 Covariance Matrix) Let

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N(\mu, \mathbf{C}).$$

Let Z_1 och Z_2 be independent $N(0,1)$. We consider the lower triangular matrix

$$\mathbf{B} = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}, \tag{1.17}$$

which clearly has an inverse, as soon as $\rho \neq \pm 1$. Moreover, one verifies that $\mathbf{C} = \mathbf{B} \cdot \mathbf{B}^{'}$, when we write $\mathbf{C}$ as in (1.5). Then we get

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mu + \mathbf{B}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \tag{1.18}$$

where, of course,

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \in N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

■

1.2 Partitioned Covariance Matrices

Assume that $\mathbf{X}$, $n \times 1$, is **partitioned** as

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)^{'},$$

where $\mathbf{X}_1$ is $p \times 1$ and $\mathbf{X}_2$ is $q \times 1$, $n = q + p$. Let the covariance matrix $\mathbf{C}$ be **partitioned** in the sense that

$$\mathbf{C} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{1.19}$$

where Σ_{11} is $p \times p$, Σ_{22} is $q \times q$ e.t.c.. The mean is partitioned correspondingly as

$$\mu := \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}. \tag{1.20}$$

Let $\mathbf{X} \in N_n(\mu, \mathbf{C})$, where N_n refers to a normal distribution in n variables, $\mathbf{C}$ and μ are partitioned as in (1.19)-(1.20). Then the marginal distribution of $\mathbf{X}_2$ is

$$\mathbf{X}_2 \in N_q(\mu_2, \Sigma_{22}),$$

if Σ_{22} is invertible. Let $\mathbf{X} \in N_n(\mu, \mathbf{C})$, where $\mathbf{C}$ and μ are partitioned as in (1.19)-(1.20). Assume that the inverse Σ_{22}^{-1} exists. Then the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is normal, or,

$$\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2 \in N_p\left(\underline{\mu}_{1|2}, \Sigma_{1|2}\right), \tag{1.21}$$

where

$$\underline{\mu}_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2) \tag{1.22}$$

and

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

By virtue of (1.21) and (1.22) the **best estimator in the mean square sense** and the **best *linear* estimator in the mean square sense** are one and the same random variable.

1.3 Appendix: Symmetric Matrices & Orthogonal Diagonalization & Gaussian Vectors

We quote some results from [1, chapter 7.2] or, from any textbook in linear algebra. An $n \times n$ matrix $\mathbf{A}$ is **orthogonally diagonalizable**, if there is an orthogonal matrix $\mathbf{P}$ (i.e., $\mathbf{P}^{'}\mathbf{P} = \mathbf{P}\mathbf{P}^{'} = \mathbf{I}$) such that

$$\mathbf{P}^{'}\mathbf{A}\mathbf{P} = \mathbf{\Lambda},$$

where $\mathbf{\Lambda}$ is a diagonal matrix. Then we have

Theorem 1.3.1 If $\mathbf{A}$ is an $n \times n$ matrix, then the following are equivalent:

(i) $\mathbf{A}$ is orthogonally diagonalizable.

(ii) $\mathbf{A}$ has an orthonormal set of eigenvectors.

(iii) $\mathbf{A}$ is symmetric.

■

Since covariance matrices are symmetric, we have by the theorem above that **all covariance matrices are orthogonally diagonalizable**.

Theorem 1.3.2 If $\mathbf{A}$ is a symmetric matrix, then

(i) Eigenvalues of $\mathbf{A}$ are all real numbers.

(ii) Eigenvectors from different eigenspaces are orthogonal.

■

That is, **all eigenvalues of a covariance matrix are real**. Hence we have for any covariance matrix the **spectral decomposition**

$$\mathbf{C} = \sum_{i=1}^{n} \lambda_i e_i e_i^{'}, \tag{1.23}$$

where $\mathbf{C}e_i = \lambda_i e_i$. Since $\mathbf{C}$ is nonnegative definite, and its eigenvectors are orthonormal,

$$0 \leq e_i^{'} \mathbf{C} e_i = \lambda_i e_i^{'} e_i = \lambda_i,$$

and thus **the eigenvalues of a covariance matrix are nonnegative**.
Let now $\mathbf{P}$ be an orthogonal matrix such that

$$\mathbf{P}^{'} \mathbf{C}_{\mathbf{X}} \mathbf{P} = \mathbf{\Lambda},$$

and $\mathbf{X} \in N(\mathbf{0}, \mathbf{C}_{\mathbf{X}})$, i.e., $\mathbf{C}_{\mathbf{X}}$ is a covariance matrix and $\mathbf{\Lambda}$ is diagonal (with the eigenvalues of $\mathbf{C}_{\mathbf{X}}$ on the main diagonal). Then if $\mathbf{Y} = \mathbf{P}^{'}\mathbf{X}$, we have by theorem 1.1.6 that

$$\mathbf{Y} \in N(\mathbf{0}, \mathbf{\Lambda}).$$

In other words, $\mathbf{Y}$ is a Gaussian vector and has by theorem 1.1.7 independent components. This method of producing independent Gaussians has several important applications. One of these is the **principal component analysis**. In addition, the operation is invertible, as

$$\mathbf{X} = \mathbf{P}\mathbf{Y}$$

recreates $\mathbf{X} \in N(\mathbf{0}, \mathbf{C}_{\mathbf{X}})$ from $\mathbf{Y}$.

1.4 Appendix: Proof of (1.16)

Let $\mathbf{X} = (X_1, X_2)^{'} \in N(\mu_{\mathbf{X}}, C)$, $\mu_{\mathbf{X}} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and C in (1.5) with $\rho^2 \neq 1$. The inverse of C in (1.5) is

$$C^{-1} = \frac{1}{\sigma_1^2 \sigma_1^2 (1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}.$$

Then we get by straightforward evaluation in (1.15)

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{2\pi\sqrt{\det C}} e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{X}})^{'} C^{-1} (\mathbf{x}-\mu_{\mathbf{X}})} \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1,x_2)}, \end{aligned} \tag{1.24}$$

where

$$Q(x_1, x_2) =$$

$$\frac{1}{(1-\rho^2)} \cdot \left[\left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right].$$

Now we claim that

$$f_{X_2|X_1=x_1}(x_2) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-\frac{1}{2\tilde{\sigma}_2^2}(x_2-\tilde{\mu}_2(x_1))^2},$$

a density of a Gaussian random variable $X_2|X_1 = x_1$ with the (conditional) expectation $\tilde{\mu}_2(x_1)$ and the (conditional) variance $\tilde{\sigma}_2$

$$\tilde{\mu}_2(x_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1), \tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2}.$$

To prove these assertions about $f_{X_2|X_1=x_1}(x_2)$ we set

$$f_{X_1}(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2}, \tag{1.25}$$

and compute the ratio $\frac{f_{X_1,X_2}(x_1,x_2)}{f_X(x_1)}$. We get from the above by (1.24) and (1.25) that

$$\frac{f_{X_1,X_2}(x_1,x_2)}{f_X(x_1)} = \frac{\sigma_1\sqrt{2\pi}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}Q(x_1,x_2)+\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2},$$

which we organize, for clarity, by introducing the auxiliary function $H(x_1, x_2)$ by

$$-\frac{1}{2}H(x_1,x_2) \stackrel{\text{def}}{=} -\frac{1}{2}Q(x_1,x_2) + \frac{1}{2\sigma_1^2}(x_1-\mu_1)^2.$$

Here we have

$$H(x_1, x_2) =$$

$$\frac{1}{(1-\rho^2)} \cdot \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2} \right)^2 \right] - \left(\frac{x_1-\mu_1}{\sigma_1} \right)^2$$

$$= \frac{\rho^2}{(1-\rho^2)} \left(\frac{x_1-\mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2(1-\rho^2)} + \left(\frac{x_2-\mu_2}{\sigma_2^2(1-\rho^2)} \right)^2.$$

Evidently we have now shown

$$H(x_1, x_2) = \frac{\left(x_2 - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1) \right)^2}{\sigma_2^2(1-\rho^2)}.$$

Hence we have found that

$$\frac{f_{X_1,X_2}(x_1,x_2)}{f_X(x_1)} = \frac{1}{\sqrt{1-\rho^2}\sigma_2\sqrt{2\pi}} e^{-\frac{1}{2}\frac{\left(x_2-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1) \right)^2}{\sigma_2^2(1-\rho^2)}}.$$

This establishes the properties of bivariate normal random variables claimed in (1.16) above. ■

As an additional exercise on the use of (1.16) (and conditional expectation) we make the following check of correctness of our formulas.

Theorem 1.4.1 $\mathbf{X} = (X_1, X_2)^{'} \in N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, C \right) \Rightarrow \rho = \rho_{X_1,X_2}$.

Proof We compute by double expectation

$$E\left[(X_1-\mu_1)(X_2-\mu_2)\right] = E(E([(X_1-\mu_1)(X_2-\mu_2)] \mid X_1)$$

and by taking out what is known,

$$= E((X_1-\mu_1)E\left[X_2-\mu_2\right] \mid X_1)) = E(X_1-\mu_1)\left[E(X_2|X_1)-\mu_2\right]$$

and by (1.16)

$$= E((X_1-\mu_1)\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1-\mu_1) - \mu_2\right]$$

$$= \rho\frac{\sigma_2}{\sigma_1}E(X_1-\mu_1)((X_1-\mu_1))$$

$$= \rho\frac{\sigma_2}{\sigma_1}E(X_1-\mu_1)^2 = \rho\frac{\sigma_2}{\sigma_1}\sigma_1^2 = \rho\sigma_2\sigma_1.$$

In other words, we have established that

$$\rho = \frac{E\left[(X_1-\mu_1)(X_2-\mu_2)\right]}{\sigma_2\sigma_1},$$

which says that ρ is the coefficient of correlation of $(X_1, X_2)^{'}$. ■

1.5 Exercises

1.5.1 The Rice Method

1. $X \in N(0, \sigma^2)$. Show that

$$E\left[\cos(X)\right] = e^{\frac{\sigma^2}{2}}.$$

1.5.2 Bivariate Gaussian Variables

1. Let $(X_1, X_2)^{'} \in N(\mu, \mathbf{C})$, where

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$\mathbf{C} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

   (a) Set $Y = X_1 - X_2$. Show that $Y \in N(0, 2-2\rho)$.

   (b) Show that

$$\mathbf{P}\left(|Y| \leq \varepsilon\right) \to 1,$$

   if $\rho \uparrow 1$.

4. $(X_1, X_2)^{'} \in N(\mathbf{0}, \mathbf{C})$, where $\mathbf{0} = (0,0)^{'}$.

   (a) Show that

$$\text{Cov}\left(X_1^2, X_2^2\right) = 2\left(\text{Cov}\left(X_1, X_2\right)\right)^2 \tag{1.26}$$

   (b) Find the mean vector and the covariance matrix of $\left(X_1^2, X_2^2\right)^{'}$.

7. In the mathematical theory of communication one introduces the **mutual information** $I(X, Y)$ between two continuous random variables X and Y by

$$I(X,Y) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x) f_Y(y)} dxdy, \tag{1.27}$$

where $f_{X,Y}(x,y)$ is the joint density of (X,Y), $f_X(x)$ and $f_Y(y)$ are the marginal densities of X and Y, respectively. $I(X,Y)$ is in fact a measure of dependence between random variables, and is theoretically speaking superior to correlation, as we measure with $I(X,Y)$ more than the mere degree of linear dependence between X and Y.

Assume now that $(X,Y) \in N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}\right)$. Check that

$$I(X,Y) = -\frac{1}{2}\log\left(1-\rho^2\right). \tag{1.28}$$

Aid: The following steps solution are in a sense instructive, as they rely on the explicit conditional distribution of $Y \mid X = x$, and provide an interesting decomposition of $I(X,Y)$ as an intermediate step. Someone may prefer other ways. Use

$$\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} = \frac{f_{Y|X=x}(y)}{f_Y(y)},$$

and then

$$I(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log f_{Y|X=x}(y)dxdy$$

$$-\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\log f_Y(y)dxdy.$$

Then one inserts in the first term on the right hand side

$$f_{X,Y}(x,y) = f_{Y|X=x}(y)\cdot f_X(x).$$

Observe that the conditional distribution of $Y \mid X = x$ is here

$$N\left(\rho x, \sigma^2(1-\rho^2)\right),$$

and take into account the marginal distributions of X and Y.

Interpret the result in (1.28) by considering $\rho = 0$, $\rho = \pm 1$. Note also that $I(X,Y) \geq 0$.

8. (From [5]) The matrix

$$\mathbf{Q} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \tag{1.29}$$

is known as the *rotation matrix*. Let

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$$

and let

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \mathbf{Q}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

and $\sigma_2^2 \geq \sigma_1^2$.

(i) Find $\mathrm{Cov}(Y_1, Y_2)$ and show that Y_1 and Y_2 are independent for all θ if and only if $\sigma_2^2 = \sigma_1^2$.

(ii) Supppose $\sigma_2^2 > \sigma_1^2$. For which values of θ are Y_1 and Y_2 are independent ?

9. (From [5]) Let
$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \in N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}\right).$$
Set
$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mathbf{Q}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$
where $\mathbf{Q}$ is the rotation matrix (1.29) with $\theta = \frac{\pi}{4}$. Show that
$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$
Hence we see that by rotating two independent Gaussian variables with variances $1+\rho$ and $1-\rho$, $\rho \neq 0$, with 45 degrees, we get a bivariate Gaussian vector, where covariance of the two variables is equal to ρ.

1.5.3 Covariance Matrices & The Four Product Rule

1. $\mathbf{C}$ is a positive definite covariance matrix. Show that $\mathbf{C}^{-1}$ is a covariance matrix.

2. $\mathbf{C}_1$ and $\mathbf{C}_2$ are two $n \times n$ covariance matrices. Show that

   (a) $\mathbf{C}_1 + \mathbf{C}_2$ is a covariance matrix.

   (b) $\mathbf{C}_1 \cdot \mathbf{C}_2$ is a covariance matrix.

Aid: The symmetry of $\mathbf{C}_2 \cdot \mathbf{C}_1$ is immediate. The difficulty is to show that $\mathbf{C}_1 \cdot \mathbf{C}_2$ is nonnegative definite. We need a piece of linear algebra here, c.f. appendix 1.3. Any symmetric and nonnegative definite matrix can written using the **spectal decomposition**, see (1.23),
$$\mathbf{C} = \sum_{i=1}^{n} \lambda_i e_i e_i^{'},$$
where e_i is a real (i.e., has no complex numbers as elements) $n \times 1$ eigenvector, i.e., $\mathbf{C}e_i = \lambda_i e_i$ and $\lambda_i \geq 0$. The set $\{e_i\}_{i=1}^{n}$ is a complete orthonormal basis in $\mathbf{R}^n$, which amongst other things implies that every $\mathbf{x} \in \mathbf{R}^n$ can be written as
$$\mathbf{x} = \sum_{i=1}^{n} (\mathbf{x}^{'} e_i) e_i,$$
where the number $\mathbf{x}^{'} e_i$ is the coordinate of $\mathbf{x}$ w.r.t. the basis vector e_i. In addition, orthonormality is recalled as the property
$$e_j^{'} e_i = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \tag{1.30}$$

We make initially the *simplifying assumption* that $\mathbf{C}_1$ and $\mathbf{C}_2$ have the same eigenvectors, so that $\mathbf{C}_1 e_i = \lambda_i e_i$, $\mathbf{C}_2 e_i = \mu_i e_i$. Then we can diagonalize the quadratic form $\mathbf{x}^{'}\mathbf{C}_2\mathbf{C}_1\mathbf{x}$ as follows.

$$\mathbf{C}_1\mathbf{x} = \sum_{i=1}^{n} (\mathbf{x}^{'} e_i)\mathbf{C}_1 e_i = \sum_{i=1}^{n} \lambda_i (\mathbf{x}^{'} e_i) e_i$$

$$= \sum_{i=1}^{n} \lambda_i (\mathbf{x}^{'} e_i) e_i. \tag{1.31}$$

Also, since $\mathbf{C}_2$ is symmetric

$$\mathbf{x}^{'}\mathbf{C}_2 = (\mathbf{C}_2\mathbf{x})^{'} = \left(\sum_{j=1}^{n} (\mathbf{x}^{'} e_j)\mathbf{C}_2 e_j \right)^{'}$$

or

$$\mathbf{x}^{'}\mathbf{C}_2 = \sum_{j=1}^{n} \mu_j (\mathbf{x}^{'} e_j) e_j^{'}. \tag{1.32}$$

Then for any $\mathbf{x} \in \mathbf{R}^n$ we get from (1.31) and (1.32) that

$$\mathbf{x}^{'}\mathbf{C}_2\mathbf{C}_1\mathbf{x} = \sum_{j=1}^{n}\sum_{i=1}^{n} \mu_j \lambda_i (\mathbf{x}^{'} e_j)(\mathbf{x}^{'} e_i) e_j^{'} e_i$$

and because of (1.30)

$$= \sum_{i=1}^{n} \mu_j \lambda_i \left(\mathbf{x}^{'} e_i\right)^2 .$$

But since $\mu_j \geq 0$ and $\lambda_i \geq 0$, we see that

$$\sum_{i=1}^{n} \mu_j \lambda_i \left(\mathbf{x}^{'} e_i\right)^2 \geq 0,$$

or, for any $\mathbf{x} \in \mathbf{R}^n$,

$$\mathbf{x}^{'}\mathbf{C}_2\mathbf{C}_1\mathbf{x} \geq 0.$$

One may use the preceding approach to handle the general case, see, e.g., [2, p.8]. The remaining work is left for the interested reader.

(c) $\mathbf{C}$ is a covariance matrix. Show that $e^{\mathbf{C}}$ is a covariance matrix.
Aid: Use a limiting procedure based on that for any square matrix A

$$e^{A} \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{1}{k!} A^n .$$

(see, e.g., [2, p.9]). Do not forget to prove symmetry.

2. **Four product rule** Let $(X_1, X_2, X_3, X_4)^{'} \in N(\mathbf{0}, \mathbf{C})$. Show that

$$E[X_1X_2X_3X_4] =$$

$$E[X_1X_2] \cdot E[X_3X_4] + E[X_1X_3] \cdot E[X_2X_4] + E[X_1X_4] \cdot E[X_2X_3] \quad (1.33)$$

The result is a special case of **Isserli's theorem**.
Aid : Take the characteristic function of $(X_1, X_2, X_3, X_4)^{'}$. Then use

$$E[X_1X_2X_3X_4] = \frac{\partial^4}{\partial s_1 \partial s_2 \partial s_3 \partial s_4} \phi_{(X_1,X_2,X_3,X_4)}(\underline{s})\,|_{\underline{s}=\mathbf{0}}.$$

As an additional aid one may say that this requires a lot of handwork. Note also that we have

$$\frac{\partial^k}{\partial s_i^k} \phi_{\underline{\mathbf{X}}}(\underline{s})\,|_{\underline{s}=\mathbf{0}} = i^k E\left[X_i^k\right], \quad i = 1, 2, \ldots, n. \quad (1.34)$$

Bibliography

[1] H. Anton & C. Rorres: *Elementary Linear Algebra with Supplemental Applications*. John Wiley & Sons (Asia) Pte Ltd, 2011.

[2] A.V. Balakrishnan: *Introduction to Random Processes in Engineering*. John Wiley & Sons, Inc., New York, 1995.

[3] D.P. Bertsekas & J.N. Tsitsiklis: *Introduction to Probability*. Athena Scietific, Belmont, Massachusetts, 2002.

[4] A. Gut: *An Intermediate Course in Probability. 2nd Edition*. Springer Verlag, Berlin 2009.

[5] R.D. Yates & D.J. Goodman: *Probability and Stochastic Processes. A Friendly Introduction for Electrical and Computer Engineers. Second Edition*. John Wiley & Sons, Inc., New York, 2005.