Appendix A

Some probability and statistics

A.1 Probabilities, random variables and their distribution

We summarize a few of the basic concepts of random variables, usually denoted by capital letters, X, Y, Z, etc, and their probability distributions, defined by the *cumulative distribution function* (CDF) $F_X(x) = \mathsf{P}(X \leq x)$, etc.

To a random experiment we define a sample space Ω, that contains all the outcomes that can occur in the experiment. A random variable, X, is a function defined on a sample space Ω. To each outcome $\omega \in \Omega$ it defines a real number $X(\omega)$, that represents the value of a numeric quantity that can be measured in the experiment. For X to be called a random variable, the probability $\mathsf{P}(X \leq x)$ has to be defined for all real x.

Distributions and moments

A probability distribution with cumulative distribution function $F_X(x)$ can be discrete or continuous with *probability function* $p_X(x)$ and *probability density function* $f_X(x)$, respectively, such that

$$F_X(x) = \mathsf{P}(X \leq x) = \begin{cases} \sum_{k \leq x} p_X(k), & \text{if } X \text{ takes only integer values}, \\ \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y. \end{cases}$$

A distribution can be of *mixed type*. The distribution function is then an integral plus discrete jumps; see Appendix B.

The *expectation* of a random variable X is defined as the center of gravity in the distribution,

$$\mathsf{E}[X] = m_X = \begin{cases} \sum_k k p_X(k), \\ \int_{x=-\infty}^{\infty} x f_X(x)\,\mathrm{d}x. \end{cases}$$

The *variance* is a simple measure of the spreading of the distribution and is defined as

$$\mathsf{V}[X] = \mathsf{E}[(X - m_X)^2] = \mathsf{E}[X^2] - m_X^2 = \begin{cases} \sum_k (k - m_X)^2 p_X(k), \\ \int_{x=-\infty}^{\infty} (x - m_X)^2 f_X(x)\,\mathrm{d}x. \end{cases}$$

Chebyshev's inequality states that, for all $\varepsilon > 0$,

$$\mathsf{P}(|X - m_X| > \varepsilon) \le \frac{\mathsf{E}[(X - m_X)^2]}{\varepsilon^2}.$$

In order to describe the statistical properties of a random function one needs the notion of *multivariate distributions*. If the result of a experiment is described by two different quantities, denoted by X and Y, e.g. length and height of randomly chosen individual in a population, or the value of a random function at two different time points, one has to deal with a two-dimensional random variable. This is described by its two-dimensional distribution function $F_{X,Y}(x,y) = P(X \le x, Y \le y)$, or the corresponding two-dimensional probability or density function,

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

Two random variables X,Y are *independent* if, for all x,y,

$$F_{X,Y}(x,y) = F_X(x) F_Y(y).$$

An important concept is the *covariance* between two random variables X and Y, defined as

$$\mathsf{C}[X,Y] = \mathsf{E}[(X - m_X)(Y - m_Y)] = \mathsf{E}[XY] - m_X m_Y.$$

The *correlation coefficient* is equal to the dimensionless, normalized covariance,[1]

$$\rho[X,Y] = \frac{\mathsf{C}[X,Y]}{\sqrt{\mathsf{V}[X]\mathsf{V}[Y]}}.$$

Two random variable with zero correlation, $\rho[X,Y] = 0$, are called *uncorrelated*. Note that if two random variables X and Y are independent, then they are also uncorrelated, but the reverse does not hold. It can happen that two uncorrelated variables are dependent.

---

[1] Speaking about the "correlation between two random quantities", one often means the degree of covariation between the two. However, one has to remember that the correlation only measures the degree of *linear covariation*. Another meaning of the term correlation is used in connection with two data series, $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$. Then sometimes the sum of products, $\sum_1^n x_k y_k$, can be called "correlation", and a device that produces this sum is called a "correlator".

Conditional distributions

If X and Y are two random variables with bivariate density function $f_{X,Y}(x,y)$, we can define the *conditional distribution* for X given $Y = y$, by the conditional density,

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

for every y where the *marginal density* $f_Y(y)$ is non-zero. The expectation in this distribution, the *conditional expectation*, is a function of y, and is denoted and defined as

$$\mathsf{E}[X \mid Y = y] = \int_x x\, f_{X|Y=y}(x)\,\mathrm{d}x = m(y).$$

The *conditional variance* is defined as

$$\mathsf{V}[X \mid Y = y] = \int_x (x - m(y))^2\, f_{X|Y=y}(x)\,\mathrm{d}x = \sigma_{X|Y}(y).$$

The unconditional expectation of X can be obtained from the *law of total probability*, and computed as

$$\mathsf{E}[X] = \int_y \left\{ \int_x x\, f_{X|Y=y}(x)\,\mathrm{d}x \right\} f_Y(y)\,\mathrm{d}y = \mathsf{E}[\mathsf{E}[X \mid Y]].$$

The unconditional variance of X is given by

$$\mathsf{V}[X] = \mathsf{E}[\mathsf{V}[X \mid Y]] + \mathsf{V}[\mathsf{E}[X \mid Y]].$$

A.2 Multidimensional normal distribution

A one-dimensional normal random variable X with expectation m and variance σ^2 has probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2 \right\},$$

and we write $X \sim N(m,\sigma^2)$. If $m = 0$ and $\sigma = 1$, the normal distribution is standardized. If $X \sim N(0,1)$, then $\sigma X + m \in N(m,\sigma^2)$, and if $X \sim N(m,\sigma^2)$ then $(X-m)/\sigma \sim N(0,1)$. We accept a constant random variable, $X \equiv m$, as a normal variable, $X \sim N(m,0)$.

Now, let $X_1, \dots, X_n$ have expectation $m_k = E[X_k]$ and covariances $\sigma_{jk} = \mathsf{C}[X_j, X_k]$, and define, (with $'$ for transpose),

$$\begin{aligned}\boldsymbol{\mu} &= (m_1, \dots, m_n)', \\ \boldsymbol{\Sigma} &= (\sigma_{jk}) = \text{the covariance matrix for } X_1, \dots, X_n.\end{aligned}$$

It is a characteristic property of the normal distributions that all linear combinations of a multivariate normal variable also has a normal distribution. To formulate the definition, write $\mathbf{a} = (a_1, \dots, a_n)'$ and $\mathbf{X} = (X_1, \dots, X_n)'$, with $\mathbf{a}'\mathbf{X} = a_1 X_1 + \dots + a_n X_n$. Then,

$$\begin{aligned}\mathsf{E}[\mathbf{a}'\mathbf{X}] &= \mathbf{a}'\boldsymbol{\mu} = \sum_{j=1}^{n} a_j m_j, \\ \mathsf{V}[\mathbf{a}'\mathbf{X}] &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = \sum_{j,k}^{n} a_j a_k \sigma_{jk}.\end{aligned} \tag{A.1}$$

Definition A.1. *The random variables $X_1, \dots, X_n$, are said to have an n-dimensional normal distribution is every linear combination $a_1 X_1 + \dots + a_n X_n$ has a normal distribution. From (A.1) we have that $\mathbf{X} = (X_1, \dots, X_n)'$ is n-dimensional normal, if and only if $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ for all $\mathbf{a} = (a_1, \dots, a_n)'$.*

Obviously, $X_k = 0 \cdot X_1 + \dots + 1 \cdot X_k + \dots + 0 \cdot X_n$, is normal, i.e., all marginal distributions in an n-dimensional normal distribution are one-dimensional normal. However, the reverse is not necessarily true; there are variables $X_1, \dots, X_n$, each of which is one-dimensional normal, but the vector $(X_1, \dots, X_n)'$ is not n-dimensional normal.

It is an important consequence of the definition that sums and differences of n-dimensional normal variables have a normal distribution.

If the covariance matrix $\boldsymbol{\Sigma}$ is non-singular, the n-dimensional normal distribution has a probability density function (with $\mathbf{x} = (x_1, \dots, x_n)'$)

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \tag{A.2}$$

The distribution is said to be *non-singular*. The density (A.2) is constant on every ellipsoid $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = C$ in $\mathbb{R}^n$.

Note: the density function of an n-dimensional normal distribution is uniquely determined by the expectations and covariances.

Example A.1. Suppose X_1, X_2 have a two-dimensional normal distribution If

$$\det \boldsymbol{\Sigma} = \sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0,$$

then $\boldsymbol{\Sigma}$ is non-singular, and

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\det \boldsymbol{\Sigma}} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix}.$$

With $Q(x_1, x_2) = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$,

$$\begin{aligned} &Q(x_1, x_2) = \\ &= \tfrac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \left\{ (x_1 - m_1)^2 \sigma_{22} - 2(x_1 - m_1)(x_2 - m_2)\sigma_{12} + (x_2 - m_2)^2 \sigma_{11} \right\} = \\ &= \tfrac{1}{1-\rho^2} \left\{ \left(\tfrac{x_1 - m_1}{\sqrt{\sigma_{11}}} \right)^2 - 2\rho \left(\tfrac{x_1 - m_1}{\sqrt{\sigma_{11}}} \right) \left(\tfrac{x_2 - m_2}{\sqrt{\sigma_{22}}} \right) + \left(\tfrac{x_2 - m_2}{\sqrt{\sigma_{22}}} \right)^2 \right\}, \end{aligned}$$

where we also used the correlation coefficient $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$, and,

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi \sqrt{\sigma_{11}\sigma_{22}(1-\rho^2)}} \exp\left(-\frac{1}{2} Q(x_1, x_2) \right). \tag{A.3}$$

For variables with $m_1 = m_2 = 0$ and $\sigma_{11} = \sigma_{22} = \sigma^2$, the bivariate density is

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma^2 \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)} (x_1^2 - 2\rho x_1 x_2 + x_2^2) \right).$$

We see that, if $\rho = 0$, this is the density of two independent normal variables.

Figure A.1 shows the density function $f_{X_1, X_2}(x_1, x_2)$ and level curves for a bivariate normal distribution with expectation $\boldsymbol{\mu} = (0, 0)$, and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The correlation coefficient is $\rho = 0.5$. ▲

Remark A.1. *If the covariance matrix $\boldsymbol{\Sigma}$ is singular and non-invertible, then there exists at least one set of constants $a_1, \ldots, a_n$, not all equal to 0, such that $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 0$. From (A.1) it follows that $\mathsf{V}[\mathbf{a}'\mathbf{X}] = 0$, which means that $\mathbf{a}'\mathbf{X}$ is constant equal to $\mathbf{a}'\boldsymbol{\mu}$. The distribution of $\mathbf{X}$ is concentrated to a hyper plane $\mathbf{a}'\mathbf{x} = \text{constant}$ in $\mathbb{R}^n$. The distribution is said to be singular and it has no density function in $\mathbb{R}^n$.*

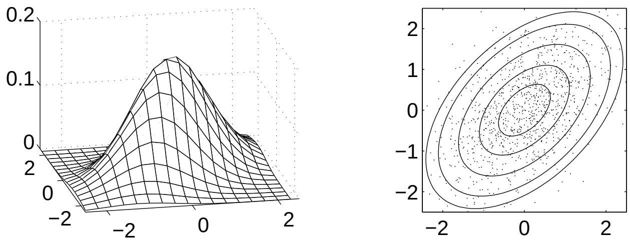

Figure A.1 *Two-dimensional normal density . Left: density function; Right: elliptic level curves at levels* $0.01, 0.02, 0.05, 0.1, 0.15$.

Remark A.2. *Formula (A.1) implies that every covariance matrix* $\mathbf{\Sigma}$ *is positive definite or positive semi-definite, i.e.,* $\sum_{j,k} a_j a_k \sigma_{jk} \geq 0$ *for all* $a_1, \ldots, a_n$*. Conversely, if* $\mathbf{\Sigma}$ *is a symmetric, positive definite matrix of size* $n \times n$*, i.e., if* $\sum_{j,k} a_j a_k \sigma_{jk} > 0$ *for all* $a_1, \ldots, a_n \neq 0, \ldots, 0$*, then (A.2) defines the density function for an n-dimensional normal distribution with expectation* m_k *and covariances* σ_{jk}*. Every symmetric, positive definite matrix is a covariance matrix for a non-singular distribution.*

Furthermore, for every symmetric, positive semi-definite matrix, $\mathbf{\Sigma}$*, i.e., such that*

$$\sum_{j,k} a_j a_k \sigma_{jk} \geq 0$$

for all $a_1, \ldots, a_n$ *with equality holding for some choice of* $a_1, \ldots, a_n \neq 0, \ldots, 0$*, there exists an n-dimensional normal distribution that has* $\mathbf{\Sigma}$ *as its covariance matrix.*

For n-dimensional normal variables, "uncorrelated" and "independent" are equivalent.

Theorem A.1. *If the random variables* $X_1, \ldots, X_n$ *are n-dimensional normal and uncorrelated, then they are independent.*

Proof. We show the theorem only for non-singular variables with density function. It is true also for singular normal variables.

If $X_1, \ldots, X_n$ are uncorrelated, $\sigma_{jk} = 0$ for $j \neq k$, then $\mathbf{\Sigma}$, and also $\mathbf{\Sigma}^{-1}$ are diagonal matrices, i.e., (note that $\sigma_{jj} = \mathsf{V}[X_j]$),

$$\det \boldsymbol{\Sigma} = \prod_j \sigma_{jj}, \qquad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma_{11}^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{nn}^{-1} \end{pmatrix}.$$

This means that $(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \sum_j (x_j-\mu_j)^2/\sigma_{jj}$, and the density (A.2) is

$$\prod_j \frac{1}{\sqrt{2\pi\sigma_{jj}}} \exp\left\{-\frac{(x_j-m_j)^2}{2\sigma_{jj}}\right\}.$$

Hence, the joint density function for $X_1,\dots,X_n$ is a product of the marginal densities, which says that the variables are independent. ■

A.2.1 *Conditional normal distribution*

This section deals with partial observations in a multivariate normal distribution. It is a special property of this distribution, that conditioning on observed values of a subset of variables, leads to a conditional distribution for the unobserved variables that is also normal. Furthermore, the expectation in the conditional distribution is linear in the observations, and the covariance matrix does not depend on the observed values. This property is particularly useful in prediction of Gaussian time series, as formulated by the *Kalman filter*.

Conditioning in the bivariate normal distribution

Let X and Y have a bivariate normal distribution with expectations m_X and m_Y, variances σ_X^2 and σ_Y^2, respectively, and with correlation coefficient $\rho = C[X,Y]/(\sigma_X\sigma_Y)$. The simultaneous density function is given by (A.3).

The conditional density function for X given that $Y=y$ is

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{1}{\sigma_X\sqrt{1-\rho^2}\sqrt{2\pi}} \exp\left\{-\frac{(x-(m_X+\sigma_X\rho(y-m_Y)/\sigma_Y))^2}{2\sigma_X^2(1-\rho^2)}\right\}. \end{aligned}$$

Hence, the conditional distribution of X given $Y=y$ is normal with expectation and variance

$$m_{X|Y=y} = m_X + \sigma_X\rho(y-m_Y)/\sigma_Y, \qquad \sigma_{X|Y=y}^2 = \sigma_X^2(1-\rho^2).$$

Note: the conditional expectation depends linearly on the observed y-value, and the conditional variance is constant, independent of $Y=y$.

Conditioning in the multivariate normal distribution

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_m)'$ be two multivariate normal variables, of size n and m, respectively, such that $\mathbf{Z} = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)'$ is $(n+m)$-dimensional normal. Denote the expectations

$$\mathsf{E}[\mathbf{X}] = m_{\mathbf{X}}, \quad \mathsf{E}[\mathbf{Y}] = m_{\mathbf{Y}},$$

and partition the covariance matrix for $\mathbf{Z}$ (with $\mathbf{\Sigma}_{\mathbf{XY}} = \mathbf{\Sigma}'_{\mathbf{YX}}$),

$$\mathbf{\Sigma} = \mathrm{Cov}\left[\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}; \begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\right] = \begin{pmatrix}\mathbf{\Sigma}_{\mathbf{XX}} & \mathbf{\Sigma}_{\mathbf{XY}}\\ \mathbf{\Sigma}_{\mathbf{YX}} & \mathbf{\Sigma}_{\mathbf{YY}}\end{pmatrix}. \tag{A.4}$$

If the covariance matrix $\mathbf{\Sigma}$ is positive definite, the distribution of $(\mathbf{X}, \mathbf{Y})$ has the density function

$$f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{(m+n)/2}\sqrt{\det \mathbf{\Sigma}}}\, e^{-\frac{1}{2}(\mathbf{x}'-m'_{\mathbf{X}}, \mathbf{y}'-m'_{\mathbf{Y}})\mathbf{\Sigma}^{-1}(\mathbf{x}-m_{\mathbf{X}}, \mathbf{y}-m_{\mathbf{Y}})},$$

while the m-dimensional density of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}\sqrt{\det \mathbf{\Sigma}_{\mathbf{YY}}}}\, e^{-\frac{1}{2}(\mathbf{y}-m_{\mathbf{Y}})'\mathbf{\Sigma}_{\mathbf{YY}}^{-1}(\mathbf{y}-m_{\mathbf{Y}})}.$$

To find the conditional density of $\mathbf{X}$ given that $\mathbf{Y} = \mathbf{y}$,

$$f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} \mid \mathbf{y}) = \frac{f_{\mathbf{YX}}(\mathbf{y}, \mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})}, \tag{A.5}$$

we need the following matrix property.

Theorem A.2 ("Matrix inversions lemma")**.** *Let* $\mathbf{B}$ *be a* $p \times p$*-matrix (*$p = n+m$*):*

$$\mathbf{B} = \begin{pmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{pmatrix},$$

where the sub-matrices have dimension $n \times n$*,* $n \times m$*, etc. Suppose* $\mathbf{B}, \mathbf{B}_{11}, \mathbf{B}_{22}$ *are non-singular, and partition the inverse in the same way as* $\mathbf{B}$*,*

$$\mathbf{A} = \mathbf{B}^{-1} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{pmatrix}.$$

Then

$$\mathbf{A} = \begin{pmatrix}(\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21})^{-1} & -(\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21})^{-1}\mathbf{B}_{12}\mathbf{B}_{22}^{-1}\\ -(\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12})^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1} & (\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12})^{-1}\end{pmatrix}.$$

Proof. For the proof, see a matrix theory textbook, for example, [22]. ■

Theorem A.3 ("Conditional normal distribution")**.** *The conditional normal distribution for* $\mathbf{X}$*, given that* $\mathbf{Y}=\mathbf{y}$*, is n-dimensional normal with expectation and covariance matrix*

$$\mathsf{E}[\mathbf{X} \mid \mathbf{Y}=\mathbf{y}] \;=\; m_{\mathbf{X}|\mathbf{Y}=\mathbf{y}} = m_{\mathbf{X}} + \boldsymbol{\Sigma}_{\mathbf{XY}}\,\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}(\mathbf{y}-m_{\mathbf{Y}}), \quad \text{(A.6)}$$

$$\mathsf{C}[\mathbf{X} \mid \mathbf{Y}=\mathbf{y}] \;=\; \boldsymbol{\Sigma}_{\mathbf{XX}|\mathbf{Y}} = \boldsymbol{\Sigma}_{\mathbf{XX}} - \boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}}. \quad \text{(A.7)}$$

These formulas are easy to remember: the dimension of the submatrices, for example in the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{XX}|\mathbf{Y}}$*, are the only possible for the matrix multiplications in the right hand side to be meaningful.*

Proof. To simplify calculations, we start with $m_{\mathbf{X}} = m_{\mathbf{Y}} = \mathbf{0}$, and add the expectations afterwards. The conditional distribution of $\mathbf{X}$ given that $\mathbf{Y}=\mathbf{y}$ is, according to (A.5), the ratio between two multivariate normal densities, and hence it is of the form,

$$c\,\exp\left\{-\frac{1}{2}(\mathbf{x}',\mathbf{y}')\,\boldsymbol{\Sigma}^{-1}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} + \frac{1}{2}\mathbf{y}'\,\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\,\mathbf{y}\right\} = c\,\exp\{-\frac{1}{2}Q(\mathbf{x},\mathbf{y})\},$$

where c is a normalization constant, independent of $\mathbf{x}$ and $\mathbf{y}$. The matrix $\boldsymbol{\Sigma}$ can be partitioned as in (A.4), and if we use the matrix inversion lemma, we find that

$$\boldsymbol{\Sigma}^{-1} = \mathbf{A} = \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{pmatrix} \quad \text{(A.8)}$$

$$= \begin{pmatrix} (\boldsymbol{\Sigma}_{\mathbf{XX}}-\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}})^{-1} & -(\boldsymbol{\Sigma}_{\mathbf{XX}}-\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}})^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}}\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\\ -(\boldsymbol{\Sigma}_{\mathbf{YY}}-\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}})^{-1}\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1} & (\boldsymbol{\Sigma}_{\mathbf{YY}}-\boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XY}})^{-1}\end{pmatrix}.$$

We also see that

$$\begin{aligned} Q(\mathbf{x},\mathbf{y}) \;&=\; \mathbf{x}'\mathbf{A}_{11}\mathbf{x} + \mathbf{x}'\mathbf{A}_{12}\mathbf{y} + \mathbf{y}'\mathbf{A}_{21}\mathbf{x} + \mathbf{y}'\mathbf{A}_{22}\mathbf{y} - \mathbf{y}'\boldsymbol{\Sigma}_{\mathbf{YY}}^{-1}\mathbf{y}\\ &=\; (\mathbf{x}'-\mathbf{y}'\mathbf{C}')\mathbf{A}_{11}(\mathbf{x}-\mathbf{C}\mathbf{y}) + \widetilde{Q}(\mathbf{y})\\ &=\; \mathbf{x}'\mathbf{A}_{11}\mathbf{x} - \mathbf{x}'\mathbf{A}_{11}\mathbf{C}\mathbf{y} - \mathbf{y}'\mathbf{C}'\mathbf{A}_{11}\mathbf{x} + \mathbf{y}'\mathbf{C}'\mathbf{A}_{11}\mathbf{C}\mathbf{y} + \widetilde{Q}(\mathbf{y}), \end{aligned}$$

for some matrix $\mathbf{C}$ and quadratic form $\widetilde{Q}(\mathbf{y})$ in $\mathbf{y}$.

Here $\mathbf{A}_{11} = \mathbf{\Sigma}_{\mathbf{XX}|\mathbf{Y}}^{-1}$, according to (A.7) and (A.8), while we can find $\mathbf{C}$ by solving

$$-\mathbf{A}_{11}\mathbf{C} = \mathbf{A}_{12}, \quad \text{i.e.} \quad \mathbf{C} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{\Sigma}_{\mathbf{XY}}\mathbf{\Sigma}_{\mathbf{YY}}^{-1},$$

according to (A.8). This is precisely the matrix in (A.6).

If we reinstall the deleted $m_{\mathbf{X}}$ and $m_{\mathbf{Y}}$, we get the conditional density for $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ to be of the form

$$c\ \exp\{-\frac{1}{2}Q(\mathbf{x},\mathbf{y})\} = c\ \exp\{-\frac{1}{2}(\mathbf{x}' - m'_{\mathbf{X}|\mathbf{Y}=\mathbf{y}})\mathbf{\Sigma}_{\mathbf{XX}|\mathbf{Y}}^{-1}(\mathbf{x} - m_{\mathbf{X}|\mathbf{Y}=\mathbf{y}})\},$$

which is the normal density we were looking for. ■

A.2.2 Complex normal variables

In most of the book, we have assumed all random variables to be real valued. In many applications, and also in the mathematical background, it is advantageous to consider complex variables, simply defined as $Z = X + iY$, where X and Y have a bivariate distribution. The mean value of a complex random variable is simply

$$\mathsf{E}[Z] = \mathsf{E}[\Re Z] + i\mathsf{E}[\Im Z],$$

while the variance and covariances are defined with complex conjugate on the second variable,

$$\begin{aligned} \mathsf{C}[Z_1, Z_2] &= \mathsf{E}[Z_1\overline{Z_2}] - m_{Z_1}\overline{m_{Z_2}}, \\ \mathsf{V}[Z] &= \mathsf{C}[Z,Z] = \mathsf{E}[|Z|^2] - |m_Z|^2. \end{aligned}$$

Note, that for a complex $Z = X + iY$, with $\mathsf{V}[X] = V[Y] = \sigma^2$,

$$\begin{aligned} &\mathsf{C}[Z,Z] = \mathsf{V}[X] + \mathsf{V}[Y] = 2\sigma^2, \\ &\mathsf{C}[Z,\overline{Z}] = \mathsf{V}[X] - \mathsf{V}[Y] + 2i\mathsf{C}[X,Y] = 2i\mathsf{C}[X,Y]. \end{aligned}$$

Hence, if the real and imaginary parts are uncorrelated with the same variance, then the complex variable Z is uncorrelated with its own complex conjugate, $\overline{Z}$. Often, one uses the term *orthogonal*, instead of uncorrelated for complex variables.