The Multivariate Gaussian Probability Distribution

Peter Ahrendt
IMM, Technical University of Denmark
mail : pa@imm.dtu.dk, web : www.imm.dtu.dk/∼pa

January 7, 2005

Contents

Chapter 1

Definition

The definition of a multivariate gaussian probability distribution can be stated in several equivalent ways. A random vector $\mathbf{X} = [\mathrm{X}_1 \mathrm{X}_2 \ldots \mathrm{X}_N]$ can be said to belong to a multivariate gaussian distribution if one of the following statements is true.

- Any linear combination $Y = a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \ldots + a_N\mathrm{X}_N,\ a_i \in \mathbb{R}$ is a (univariate) gaussian distribution.
- There exists a random vector $\mathbf{Z} = [\mathrm{Z}_1, \ldots, \mathrm{Z}_M]$ with components that are independent and standard normal distributed, a vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_N]$ and an N-by-M matrix $\mathbf{A}$ such that $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$.
- There exists a vector $\boldsymbol{\mu}$ and a symmetric, positive semi-definite matrix $\boldsymbol{\Gamma}$ such that the characteristic function of $\mathbf{X}$ can be written $\phi_{\mathbf{x}}(\mathbf{t}) \equiv \langle e^{it^T\mathbf{X}} \rangle = e^{i\boldsymbol{\mu}^T\mathbf{t} - \frac{1}{2}\mathbf{t}^T\boldsymbol{\Gamma}\mathbf{t}}$.

Under the assumption that the covariance matrix $\boldsymbol{\Sigma}$ is non-singular, the probability density function (pdf) can be written as :

$$
\begin{aligned}
N_{\mathbf{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \\
&= |2\pi\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)
\end{aligned}
\tag{1.1}
$$

Then $\boldsymbol{\mu}$ is the *mean value*, $\boldsymbol{\Sigma}$ is the *covariance matrix* and $|\cdot|$ denote the determinant. Note that it is possible to have multivariate gaussian distributions with singular covariance matrix and then the above expression cannot be used for the pdf. In the following, however, non-singular covariance matrices will be assumed.

In the limit of one dimension, the familiar expression of the univariate gaussian pdf is found.

$$N_x(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)\right) \tag{1.2}$$

Neither of them have a closed-form expression for the cumulative density function.

Symmetries

It is noted that in the one-dimensional case there is a symmetry in the pdf. $N_x(\mu,\sigma^2)$ which is centered on μ. This can be seen by looking at "contour lines", i.e. setting the exponent $-\frac{(x-\mu)^2}{2\sigma^2} = c$. It is seen that σ determines the width of the distribution.

In the multivariate case, it is similarly useful to look at $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = c$. This is a quadratic form and geometrically the contour curves (for fixed c) are hyperellipsoids. In 2D, this is normal ellipsoids with the form $(\frac{x-x_0}{a})^2 + (\frac{y-y_0}{b})^2 = r^2$, which gives symmetries along the principal axes. Similarly, the hyperellipsoids show symmetries along their principal axes.

Notation: If a random variable $\mathbf{X}$ has a gaussian distribution, it is written as $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The probability density function of this variable is then given by $N_{\mathbf{x}}(\boldsymbol{\mu},\boldsymbol{\Sigma})$.

Chapter 2

Functions of Gaussian Variables

Linear transformation and addition of variables

Let $\mathbf{A}, \mathbf{B} \in \mathbb{M}^{c \cdot d}$ and $\mathbf{c} \in \mathbb{R}^d$. Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{y}}, \boldsymbol{\Sigma}_{\mathrm{y}})$ be independent variables. Then

$$\mathbf{Z} = \mathbf{AX} + \mathbf{BY} + \mathbf{c} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu}_{\mathrm{x}} + \mathbf{B}\boldsymbol{\mu}_{\mathrm{y}} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}_{\mathrm{x}}\mathbf{A}^T + \mathbf{B}\boldsymbol{\Sigma}_{\mathrm{y}}\mathbf{B}^T) \tag{2.1}$$

Transform to standard normal variables

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \mu) \sim \mathcal{N}(\mathbf{0}, \mathbf{1}) \tag{2.2}$$

Note, that by $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ is actually meant a unique matrix, although in general matrices with fractional exponents are not. The matrix that is meant can be found from the diagonalisation into $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})(\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^T$ where $\boldsymbol{\Lambda}$ is the diagonal matrix with the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{U}$ is the matrix with the eigenvectors. Then $\boldsymbol{\Sigma}^{-\frac{1}{2}} = (\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}})^{-1} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{U}^{-1}$.

In the one-dimensional case, this corresponds to the transformation of $\mathrm{X} \sim \mathcal{N}(\mu, \sigma^2)$ into $\mathrm{Y} = \sigma^{-1}(\mathrm{X} - \mu) \sim \mathcal{N}(0, 1)$.

Addition

Let $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i \in 1, ..., N$ be independent variables. Then

$$\sum_i^N \mathbf{X}_i \sim \mathcal{N}(\sum_i^N \boldsymbol{\mu}_i, \sum_i^N \boldsymbol{\Sigma}_i) \tag{2.3}$$

Note: This is a direct implication of equation (2.1).

Quadratic

Let $\mathrm{X}_i \sim \mathcal{N}(0,1), i \in 1,...,N$ be independent variables. Then

$$\sum_i^N \mathrm{X}_i^2 \sim \chi_n^2 \tag{2.4}$$

Alternatively let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$. Then

$$\mathrm{Z} = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_n^2 \tag{2.5}$$

This is, however, the same thing since Z= $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \sum_i^N \tilde{\mathrm{X}_i}^2$, where $\tilde{\mathrm{X}}_i$ are the decorrelated components (see eqn. (2.2)).

Chapter 3

Characteristic function and Moments

The characteristic function of the univariate gaussian distribution is given by $\phi_{\mathsf{x}}(t) \equiv \langle e^{it\mathrm{X}} \rangle = e^{it\mu - \sigma^2 t^2/2}$. The generalization to multivariate gaussian distributions is

$$\phi_{\mathbf{x}}(\mathbf{t}) \equiv \langle e^{it^T\mathbf{X}} \rangle = e^{i\boldsymbol{\mu}^T\mathbf{t} - \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}} \tag{3.1}$$

The pdf $p(\mathbf{x})$ is related to the characteristic function.

$$p(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \phi_{\mathbf{x}}(\mathbf{t})\ e^{-i\mathbf{t}^T\mathbf{x}} \mathrm{d}\mathbf{t} \tag{3.2}$$

It is seen that the characteristic function is the inverse Fourier transform of the pdf.

Moments of a pdf are generally defined as :

$$\langle \mathrm{X}_1^{k_1} \mathrm{X}_2^{k_2} \cdots \mathrm{X}_N^{k_N} \rangle \equiv \int_{\mathbb{R}^d} \mathrm{x}_1^{k_1} \mathrm{x}_2^{k_2} \cdots \mathrm{x}_N^{k_N}\ p(\mathbf{x})\ \mathrm{d}\mathbf{x} \tag{3.3}$$

where $\langle \mathrm{X}_1^{k_1} \mathrm{X}_2^{k_2} \cdots \mathrm{X}_N^{k_N} \rangle$ is the *k'th order moment*, $\mathbf{k} = [k_1, k_2, \ldots, k_N]$ ($k_i \in \mathbb{N}$) and $k = k_1 + k_2 + \ldots + k_N$. A well-known example is the first order moment, called the mean value μ_i (of variable X_i) - or the mean $\boldsymbol{\mu} \equiv [\mu_1 \mu_2 \ldots \mu_N]$ of the whole random vector $\mathbf{X}$.

The *k'th order central moment* is defined as above, but with X_i replaced by $\mathrm{X}_i - \mu_i$ in equation (3.3). An example is the second order central moment, called the variance, which is given by $\langle (\mathrm{X}_i - \mu_i)^2 \rangle$.

Any moment (that exists) can be found from the characteristic function [8]:

$$\langle \mathrm{X}_1^{k_1} \mathrm{X}_2^{k_2} \cdots \mathrm{X}_N^{k_N} \rangle = (-j)^k \frac{\partial^k \phi_{\mathbf{x}}(\mathbf{t})}{\partial t_1^{k_1} \ldots \partial t_N^{k_N}} \bigg|_{\mathbf{t}=\mathbf{0}} \tag{3.4}$$

where $k = k_1 + k_2 + \ldots + k_N$.

<u>1. Order Moments</u>

$$\text{Mean } \boldsymbol{\mu} \equiv \langle \mathbf{X} \rangle \tag{3.5}$$

<u>2. Order Moments</u>

$$\text{Variance } c_{ii} \equiv \langle (\mathrm{X}_i - \mu_i)^2 \rangle = \langle \mathrm{X}_i^2 \rangle - \mu_i^2 \tag{3.6}$$

$$\text{Covariance } c_{ij} \equiv \langle (\mathrm{X}_i - \mu_i)(\mathrm{X}_j - \mu_j) \rangle \tag{3.7}$$

$$\text{Covariance matrix } \boldsymbol{\Sigma} \equiv \langle (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \rangle \equiv [c_{ij}] \tag{3.8}$$

<u>3. Order Moments</u>

Often the skewness is used.

$$\mathrm{Skew}(\mathbf{X}) \equiv \frac{\langle (\mathrm{X}_i - \langle \mathrm{X}_i \rangle)^3 \rangle}{\langle (\mathrm{X}_i - \langle \mathrm{X}_i \rangle)^2 \rangle^{\frac{3}{2}}} = \frac{\langle (\mathrm{X}_i - \mu_i)^3 \rangle}{\langle (\mathrm{X}_i - \mu_i)^2 \rangle^{\frac{3}{2}}} \tag{3.9}$$

All 3. order central moments are zero for gaussian distributions and thus also the skewness.

<u>4. Order Moments</u>

The kurtosis is (in newer literature) given as

$$\mathrm{Kurt}(\mathbf{X}) \equiv \frac{\langle (\mathrm{X}_i - \mu_i)^4 \rangle}{\langle (\mathrm{X}_i - \mu_i)^2 \rangle^2} - 3 \tag{3.10}$$

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

$$\langle (\mathrm{X}_i - \mu_i)(\mathrm{X}_j - \mu_j)(\mathrm{X}_k - \mu_k)(\mathrm{X}_l - \mu_l) \rangle = c_{ij}c_{kl} + c_{il}c_{jk} + c_{ik}c_{lj} \tag{3.11}$$

and

$$\mathrm{Kurt}(\mathbf{X}) = 0 \tag{3.12}$$

<u>N. Order Moments</u>

Any central moment of a gaussian distribution can (fairly easily) be calculated with the following method [3] (sometimes known as Wicks theorem).

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then

- Assume k is odd. Then the central k'th order moments are all zero.
- Assume k is even. Then the central k'th order moments are equal to $\sum(c_{ij}c_{kl}\ldots c_{xz})$. The sum is taken over all *different* permutations of the k indices, where it is noted that $c_{ij} = c_{ji}$. This gives $(k-1)!/(2^{k/2-1}(k/2-1)!)$ terms which each is the product of $k/2$ covariances.

An example is illustrative. The different 4. order central moments of $\mathbf{X}$ are found with the above method to give

$$\begin{aligned}
\langle(\mathrm{X}_i-\mu_i)^4\rangle &= 3c_{ii}^2 \\
\langle(\mathrm{X}_i-\mu_i)^3(\mathrm{X}_j-\mu_j)\rangle &= 3c_{ii}c_{ij} \\
\langle(\mathrm{X}_i-\mu_i)^2(\mathrm{X}_j-\mu_j)^2\rangle &= c_{ii}c_{jj}+2c_{ij}^2 \\
\langle(\mathrm{X}_i-\mu_i)^2(\mathrm{X}_j-\mu_j)(\mathrm{X}_k-\mu_k)\rangle &= c_{ii}c_{jk}+2c_{ij}c_{ik} \\
\langle(\mathrm{X}_i-\mu_i)(\mathrm{X}_j-\mu_j)(\mathrm{X}_k-\mu_k)(\mathrm{X}_l-\mu_l)\rangle &= c_{ij}c_{kl}+c_{il}c_{jk}+c_{ik}c_{lj}
\end{aligned} \tag{3.13}$$

The above results were found by seeing that the different permutations of the k=4 indices are (12)(34), (13)(24) and (14)(23). Other permutations are equivalents, such as for instance (32)(14) which is equivalent to (14)(23). When calculating e.g. $\langle(\mathrm{X}_i-\mu_i)^2(\mathrm{X}_j-\mu_j)(\mathrm{X}_k-\mu_k)\rangle$, the assignment $(1 \to i, 2 \to i, 3 \to j, 4 \to k)$ gives the terms $c_{ii}c_{jk}$, $c_{ij}c_{ik}$ and $c_{ij}c_{ik}$ in the sum.

<u>Calculations with moments</u>

Let $\mathbf{b} \in \mathbb{R}^c$, $A, B \in \mathbb{M}^{c\cdot d}$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors and $\mathbf{f}$ and $\mathbf{g}$ vector functions. Then

$$\langle A\mathbf{f}(\mathbf{X}) + B\mathbf{g}(\mathbf{X}) + \mathbf{b}\rangle = A\langle\mathbf{f}(\mathbf{X})\rangle + B\langle\mathbf{g}(\mathbf{X})\rangle + \mathbf{b} \tag{3.14}$$

$$\langle A\mathbf{X} + \mathbf{b}\rangle = A\langle\mathbf{X}\rangle + \mathbf{b} \tag{3.15}$$

$$\langle\langle\mathbf{Y}|\mathbf{X}\rangle\rangle \equiv E(E(\mathbf{Y}|\mathbf{X})) = \langle\mathbf{Y}\rangle \tag{3.16}$$

If X_i and X_j are independent then

$$\langle\mathrm{X}_i\mathrm{X}_j\rangle = \langle\mathrm{X}_i\rangle\langle\mathrm{X}_j\rangle \tag{3.17}$$

Chapter 4

Marginalization and Conditional Distribution

4.1 Marginalization

Marginalization is the operation of integrating out variables of the pdf of a random vector $\mathbf{X}$. Assume that $\mathbf{X}$ is split into two parts (since the ordering of the X_i is arbitrary, this corresponds to any division of the variables), $\mathbf{X} = [\mathbf{X}_{1:c}^T \mathbf{X}_{c+1:N}^T]^T = [\mathrm{X}_1 \mathrm{X}_2 \dots \mathrm{X}_c \mathrm{X}_{c+1} \dots \mathrm{X}_N]^T$. Let the pdf of $\mathbf{X}$ be $p(\mathbf{x}) = p(x_1, \dots, x_N)$, then :

$$p(x_1, \dots, x_c) = \int \cdots \int_{\mathbb{R}^{c+1:N}} p(x_1, \dots, x_N)\, \mathrm{d}x_{c+1} \dots x_N \tag{4.1}$$

The nice part about gaussian distributions is that *every marginal distribution of a gaussian distribution is itself a gaussian.* More specifically, let $\mathbf{X}$ be split into two parts as above and $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then :

$$p(x_1, \dots, x_c) = p(\mathbf{x}_{1:c}) = \mathrm{N}_{1:c}(\boldsymbol{\mu}_{1:c}, \boldsymbol{\Sigma}_{1:c}) \tag{4.2}$$

where $\boldsymbol{\mu}_{1:c} = [\mu_1, \mu_2, \dots, \mu_c]$ and

$$\boldsymbol{\Sigma}_{1:c} = \begin{pmatrix} c_{11} & c_{21} & \dots & c_{c1} \\ c_{12} & c_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ c_{c1} & \dots & \dots & c_{cc} \end{pmatrix}$$

In words, the *mean and covariance matrix of the marginal distribution is the same as the corresponding elements of the joint distribution.*

4.2 Conditional distribution

As in the previous, let $\mathbf{X} = [\mathbf{X}_{1:c}^T \mathbf{X}_{c+1:N}^T]^T = [\mathrm{X}_1 \mathrm{X}_2 \ldots \mathrm{X}_c \mathrm{X}_{c+1} \ldots \mathrm{X}_N]^T$ be a division of the variables into two parts. Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ and use the notation $\mathbf{X} = [\mathbf{X}_{1:c}^T \mathbf{X}_{c+1:N}^T]^T = [\mathbf{X}_{(1)}^T \mathbf{X}_{(2)}^T]^T$ and

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)} \\ \boldsymbol{\mu}_{(2)} \end{pmatrix}$$

and

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

It is found that the conditional distribution $p(\mathbf{x}_{(1)}|\mathbf{x}_{(2)})$ is in fact again a gaussian distribution and

$$\mathbf{X}_{(1)}|\mathbf{X}_{(2)} \sim \mathcal{N}(\boldsymbol{\mu}_{(1)} + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_{(2)}), \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{12}^T) \tag{4.3}$$

Chapter 5

Tips and Tricks

5.1 Products

Consider the product $N_{\mathbf{x}}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \cdot N_{\mathbf{x}}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$ and note that they both have $\mathbf{x}$ as their "random variable". Then

$$N_{\mathbf{x}}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a) \cdot N_{\mathbf{x}}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b) = z_c N_{\mathbf{x}}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{5.1}$$

where $\boldsymbol{\Sigma}_c = (\boldsymbol{\Sigma}_a^{-1} + \boldsymbol{\Sigma}_b^{-1})^{-1}$ and $\boldsymbol{\mu}_c = \boldsymbol{\Sigma}_c(\boldsymbol{\Sigma}_a^{-1}\boldsymbol{\mu}_a + \boldsymbol{\Sigma}_b^{-1}\boldsymbol{\mu}_b)$ and

$$\begin{aligned} z_c &= |2\pi\boldsymbol{\Sigma}_a\boldsymbol{\Sigma}_b\boldsymbol{\Sigma}_c^{-1}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^T \boldsymbol{\Sigma}_a^{-1}\boldsymbol{\Sigma}_c\boldsymbol{\Sigma}_b^{-1}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)\right) \\ &= |2\pi(\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_b)|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)^T (\boldsymbol{\Sigma}_a + \boldsymbol{\Sigma}_b)^{-1}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_b)\right) \end{aligned} \tag{5.2}$$

In words, the product of two gaussians is another gaussian (unnormalized). This can be generalised to a product of K gaussians with distributions $\mathbf{X}_k \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.

$$\prod_{k=1}^{K} N_{\mathbf{x}}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \tilde{z} \cdot N_{\mathbf{x}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}) \tag{5.3}$$

where $\tilde{\boldsymbol{\Sigma}} = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\right)^{-1}$ and $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}\left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right) = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\right)^{-1}\left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right)$ and

$$\tilde{z} = \frac{|2\pi\boldsymbol{\Sigma}_d|^{\frac{1}{2}}}{\prod_{k=1}^{K} |2\pi\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \prod_{i<j} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \mathbf{B}_{ij}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\right) \tag{5.4}$$

where

$$\mathbf{B}_{ij} = \mathbf{\Sigma}_i^{-1} \left(\sum_{k=1}^{K} \mathbf{\Sigma}_k^{-1} \right)^{-1} \mathbf{\Sigma}_j^{-1} \tag{5.5}$$

5.2 Gaussian Integrals

A nice thing about the fact that products of gaussian functions are again a gaussian function, is that it makes gaussian integrals easier to calculate since $\int N_{\mathbf{x}}(\tilde{\boldsymbol{\mu}}, \tilde{\mathbf{\Sigma}}) \mathrm{d}\mathbf{x} = 1$. Using this with the equations (5.1) and (5.3) of the previous section gives the following.

$$\begin{aligned} \int_{\mathbb{R}^d} N_{\mathbf{x}}(\boldsymbol{\mu}_a, \mathbf{\Sigma}_a) \cdot N_{\mathbf{x}}(\boldsymbol{\mu}_b, \mathbf{\Sigma}_b) \ \mathrm{d}\mathbf{x} &= \int_{\mathbb{R}^d} z_c N_{\mathbf{x}}(\boldsymbol{\mu}_c, \mathbf{\Sigma}_c) \ \mathrm{d}\mathbf{x} \\ &= z_c \end{aligned} \tag{5.6}$$

Similarly,

$$\int_{\mathbb{R}^d} \prod_{k=1}^{K} N_{\mathbf{x}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k) \ \mathrm{d}\mathbf{x} = \tilde{z} \tag{5.7}$$

Equation (5.3) can also be used to calculate integrals such as $\int |\mathbf{x}|^q \left(\prod_k N_{\mathbf{x}}(\boldsymbol{\mu}_k, \mathbf{\Sigma}_k)\right) \mathrm{d}\mathbf{x}$ or similar by using the same technique as above.

5.3 Useful integrals

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $\mathbf{a} \in \mathcal{R}^d$ an arbitrary vector. Then

$$\langle e^{\mathbf{a}^{\mathbf{T}}\mathbf{x}} \rangle \equiv \int N_{\mathbf{x}}(\boldsymbol{\mu}, \mathbf{\Sigma}) e^{\mathbf{a}^{\mathbf{T}}\mathbf{x}} \mathrm{d}\mathbf{x} = e^{\mathbf{a}^T \boldsymbol{\mu} + \frac{1}{2}\mathbf{a}^T \mathbf{\Sigma} \mathbf{a}} \tag{5.8}$$

From this expression, it is possible to find integrals such as $\int \exp(\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{a}^T \mathbf{x}) \ \mathrm{d}\mathbf{x}$. Another useful integral is

$$\langle e^{\mathbf{x}^T \mathbf{A} \mathbf{x}} \rangle \equiv \int N_{\mathbf{x}}(\boldsymbol{\mu}, \mathbf{\Sigma}) e^{\mathbf{x}^T \mathbf{A} \mathbf{x}} \mathrm{d}\mathbf{x} = |\mathbf{I} - 2\mathbf{\Sigma}\mathbf{A}|^{-\frac{1}{2}} e^{-\frac{1}{2}\boldsymbol{\mu}^T (\mathbf{\Sigma} - 2\mathbf{A}^{-1})^{-1} \boldsymbol{\mu}} \tag{5.9}$$

where $\mathbf{A} \in \mathbb{M}^{d \cdot d}$ is a non-singular matrix.

Bibliography

[1] T.W. Anderson, *An introduction to multivariate statistical analysis.* Wiley, 1984.

[2] C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[3] K. Triantafyllopoulos, "On the central moments of the multidimensional Gaussian distribution," *The Mathematical Scientist*, vol. 28, pp. 125–128, 2003.

[4] S. Roweis, "Gaussian Identities," http://www.cs.toronto.edu/ roweis/notes.html.

[5] J. Larsen, "Gaussian Integrals," *Tech. Rep.*, http://isp.imm.dtu.dk/staff-/jlarsen/pubs/frame.htm

[6] www.Wikipedia.org.

[7] www.MathWorld.wolfram.com

[8] P. Kidmose, "Blind Separation of Heavy Tail Signals," *Ph.d. Thesis*, http://isp.imm.dtu.dk/staff/kidmose/pk_publications.html