The Multivariate Gaussian Distribution

Steffen Lauritzen, University of Oxford

BS2 Statistical Inference, Lecture 6, Hilary Term 2009

February 6, 2009
A $d$-dimensional random vector $X = (X_1, \ldots, X_d)$ is has a \textit{multivariate Gaussian distribution} or \textit{normal} distribution on $\mathcal{R}^d$ if there is a vector $\xi \in \mathcal{R}^d$ and a $d \times d$ matrix $\Sigma$ such that

$$\lambda^\top X \sim \mathcal{N}(\lambda^\top \xi, \lambda^\top \Sigma \lambda) \quad \text{for all } \lambda \in \mathcal{R}^d. \quad (1)$$

We then write $X \sim \mathcal{N}_d(\xi, \Sigma)$.

Taking $\lambda = e_i$ or $\lambda = e_i + e_j$ where $e_i$ is the unit vector with $i$-th coordinate 1 and the remaining equal to zero yields:

$$X_i \sim \mathcal{N}(\xi_i, \sigma_{ii}), \quad \text{Cov}(X_i, X_j) = \sigma_{ij}.$$

Hence $\xi$ is the \textit{mean vector} and $\Sigma$ the \textit{covariance matrix} of the distribution.
The definition (1) makes sense if and only if $\lambda^\top \Sigma \lambda \geq 0$, i.e. if $\Sigma$ is \textit{positive semidefinite}. Note that we have allowed distributions with variance zero.

The multivariate moment generating function of $X$ can be calculated using the relation (1) as

$$m_d(\lambda) = E\{e^{\lambda^\top X}\} = e^{\lambda^\top \xi + \lambda^\top \Sigma \lambda / 2}$$

where we have used that the univariate moment generating function for $\mathcal{N}(\mu, \sigma^2)$ is

$$m_1(t) = e^{t\mu + \sigma^2 t^2 / 2}$$

and let $t = 1$, $\mu = \lambda^\top \xi$, and $\sigma^2 = \lambda^\top \Sigma \lambda$.

In particular this means that \textit{a multivariate Gaussian distribution is determined by its mean vector and covariance matrix}. 
Assume $X^\top = (X_1, X_2, X_3)$ with $X_i$ independent and $X_i \sim \mathcal{N}(\xi_i, \sigma_i^2)$. Then

$$
\lambda^\top X = \lambda_1 X_1 + \lambda_2 X_2 + \lambda_3 X_3 \sim \mathcal{N}(\mu, \tau^2)
$$

with

$$
\mu = \lambda^\top \xi = \lambda_1 \xi_1 + \lambda_2 \xi_2 + \lambda_3 \xi_3, \quad \tau^2 = \lambda_1^2 \sigma_1^2 + \lambda_2^2 \sigma_2^2 + \lambda_3^2 \sigma_3^2.
$$

Hence $X \sim \mathcal{N}_3(\xi, \Sigma)$ with $\xi^\top = (\xi_1, \xi_2, \xi_3)$ and

$$
\Sigma = 
\begin{pmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{pmatrix}.
$$
Basic definitions

The multivariate Gaussian

Basic properties

Density of multivariate Gaussian

Simple example

Bivariate case

A counterexample

If Σ is positive definite, i.e. if $\lambda^\top \Sigma \lambda > 0$ for $\lambda \neq 0$, the distribution has density on $\mathbb{R}^d$

$$f(x | \xi, \Sigma) = (2\pi)^{-d/2} (\det K)^{1/2} e^{-(x-\xi)^\top K (x-\xi)/2}, \quad (2)$$

where $K = \Sigma^{-1}$ is the concentration matrix of the distribution. We then also say that Σ is regular.

If $X_1, \ldots, X_d$ are independent and $X_i \sim \mathcal{N}(\xi_i, \sigma_i^2)$ their joint density has the form $(2)$ with $\Sigma = \text{diag}(\sigma_i^2)$ and $K = \Sigma^{-1} = \text{diag}(1/\sigma_i^2)$.

Hence vectors of independent Gaussians are multivariate Gaussian.
In the bivariate case it is traditional to write

\[ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}, \]

with \( \rho \) being the correlation between \( X_1 \) and \( X_2 \). Then

\[ \det(\Sigma) = \sigma_1^2 \sigma_2^2 (1 - \rho^2) = \det(K)^{-1} \]

and

\[ K = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_1 \sigma_2 \rho \\ -\sigma_1 \sigma_2 \rho & \sigma_1^2 \end{pmatrix}. \]
Thus the density becomes

\[ f(x \mid \xi, \Sigma) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \times e^{- \frac{1}{2(1-\rho^2)} \left\{ \frac{(x_1-\xi_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1-\xi_1)(x_2-\xi_2)}{\sigma_1\sigma_2} + \frac{(x_2-\xi_2)^2}{\sigma_2^2} \right\}}. \]

The contours of this density are ellipses and the corresponding density is bell-shaped with maximum in \((\xi_1, \xi_2)\).
The marginal distributions of a vector $X$ can all be Gaussian without the joint being multivariate Gaussian:

For example, let $X_1 \sim \mathcal{N}(0, 1)$, and define $X_2$ as

$$X_2 = \begin{cases} X_1 & \text{if } |X_1| > c \\ -X_1 & \text{otherwise.} \end{cases}$$

Then, using the symmetry of the univariate Gaussian distribution, $X_2$ is also distributed as $\mathcal{N}(0, 1)$. 
However, the joint distribution is not Gaussian unless \( c = 0 \) since, for example, \( Y = X_1 + X_2 \) satisfies

\[
P(Y = 0) = P(X_2 = -X_1) = P(|X_1| \leq c) = \Phi(c) - \Phi(-c).
\]

Note that for \( c = 0 \), the correlation \( \rho \) between \( X_1 \) and \( X_2 \) is \(-1\) whereas for \( c = \infty \), \( \rho = 1 \).

It follows that there is a value of \( c \) so that \( X_1 \) and \( X_2 \) are uncorrelated, and still not jointly Gaussian.
Adding two independent Gaussians yields a Gaussian:

If $X_1 \sim \mathcal{N}_d(\xi_1, \Sigma_1)$ and $X_2 \sim \mathcal{N}_d(\xi_2, \Sigma_2)$ and $X_1 \perp \perp X_2$

$$X_1 + X_2 \sim \mathcal{N}_d(\xi_1 + \xi_2, \Sigma_1 + \Sigma_2).$$

To see this, just note that

$$\lambda^\top (X_1 + X_2) = \lambda^\top X_1 + \lambda^\top X_2$$

and use the univariate addition property.
Linear transformations preserve multivariate normality:
If $A$ is an $r \times d$ matrix, $b \in \mathcal{R}^r$ and $X \sim \mathcal{N}_d( \xi, \Sigma )$, then

$$Y = AX + b \sim \mathcal{N}_r( A\xi + b, A\Sigma A^\top ).$$

Again, just write

$$\gamma^\top Y = \gamma^\top (AX + b) = (A^\top \gamma)^\top X + \gamma^\top b$$

and use the corresponding univariate result.
Partition $X$ into into $X_1$ and $X_2$, where $X_1 \in \mathcal{R}^r$ and $X_2 \in \mathcal{R}^s$ with $r + s = d$.

Partition mean vector, concentration and covariance matrix accordingly as

$$
\xi = \begin{pmatrix}
\xi_1 \\
\xi_2 
\end{pmatrix}, \quad
K = \begin{pmatrix}
K_{11} & K_{12} \\
K_{21} & K_{22} 
\end{pmatrix}, \quad
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22} 
\end{pmatrix}
$$

so that $\Sigma_{11}$ is $r \times r$ and so on. Then, if $X \sim \mathcal{N}_d(\xi, \Sigma)$

$$
X_2 \sim \mathcal{N}_s(\xi_2, \Sigma_{22}).
$$

This follows simply from the previous fact using the matrix

$$
A = (0_{sr} \ I_s).
$$

where $0_{sr}$ is an $s \times r$ matrix of zeros and $I_s$ is the $s \times s$ identity matrix.
If $\Sigma_{22}$ is regular, it further holds that

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}_r(\xi_{1\mid2}, \Sigma_{1\mid2}),$$

where

$$\xi_{1\mid2} = \xi_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \xi_2) \quad \text{and} \quad \Sigma_{1\mid2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}.$$

In particular, $\Sigma_{12} = 0$ if and only if $X_1$ and $X_2$ are independent.
To see this, we simply calculate the conditional density.

\[
f(x_1 \mid x_2) \propto f_{\xi, \Sigma}(x_1, x_2)
\]

\[
\propto \exp \left\{ -(x_1 - \xi_1)^\top K_{11}(x_1 - \xi_1)/2 - (x_1 - \xi_1)^\top K_{12}(x_2 - \xi_2) \right\}.
\]

The linear term involving \(x_1\) has coefficient equal to

\[
K_{11}\xi_1 - K_{12}(x_2 - \xi_2) = K_{11} \left\{ \xi_1 - K_{11}^{-1}K_{12}(x_2 - \xi_2) \right\}.
\]

Using the matrix identities

\[
K_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}
\]  \hspace{1cm} (3)

and

\[
K_{11}^{-1}K_{12} = -\Sigma_{12}\Sigma_{22}^{-1},
\]  \hspace{1cm} (4)
we find

\[ f(x_1 | x_2) \propto \exp \left\{ -(x_1 - \xi_{1|2})^\top K_{11} (x_1 - \xi_{1|2})/2 \right\} \]

and the result follows.

From the identities (3) and (4) it follows in particular that then the conditional expectation and concentrations also can be calculated as

\[ \xi_{1|2} = \xi_1 - K_{11}^{-1} K_{12} (x_2 - \xi_2) \quad \text{and} \quad K_{1|2} = K_{11}. \]

Note that the *marginal covariance is simply expressed in terms of \( \Sigma \) where as the *conditional concentration is simply expressed in terms of \( K \).*
Consider $\mathcal{N}_3(0, \Sigma)$ with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$ 

The concentration matrix is

$$K = \Sigma^{-1} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}. $$

Steffen Lauritzen, University of Oxford

The Multivariate Gaussian Distribution
The marginal distribution of \((X_2, X_3)\) has covariance and concentration matrix

\[
\Sigma_{23} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad (\Sigma_{23})^{-1} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.
\]

The conditional distribution of \((X_1, X_2)\) given \(X_3\) has concentration and covariance matrix

\[
K_{12} = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}, \quad \Sigma_{12|3} = (K_{12})^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}.
\]

Similarly, \(\mathbf{V}(X_1 \mid X_2, X_3) = 1/k_{11} = 1/3\), etc.