# The multivariate gaussian distribution

February 23, 2018



## The multivariate gaussian distribution





## Covariance matrices

The covariance matrix $K_X$ of an $n$-dimensional random variable

$$X = (X_1, X_2, \ldots, X_n)^t$$

is the square $n \times n$ matrix defined by

$$K_X = \mathrm{E}\left((X - m_X)(X - m_X)^t\right)$$

$$= \begin{pmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{pmatrix}$$



## Covariance matrices

- $m_X$ is the expectation vector

$$m_X = \mathrm{E}(X) = (m_{X_1}, m_{X_2}, \ldots, m_{X_n})^t$$

- For $i \neq j$,

$$k_{ij} = \mathrm{E}\left((X_i - m_{X_i})(X_j - m_{X_j})\right) = \mathrm{Cov}(X_i, X_j)$$

- The diagonal entries of $K_X$ are

$$k_{ii} = \mathrm{E}\left((X_i - m_{X_i})^2\right) = \mathrm{Var}(X_i)$$

## Covariance matrices

The covariance matrix $K_X$ is:

- symmetric

$$k_{ij} = \text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i) = k_{ji}$$

- positive-semidefinite

  That is, for all $z = (z_1, z_2, \ldots, z_n)^t \in \mathbb{R}^n$,

$$z^t K_X z \geq 0$$



## Covariance matrices

Let us say that the random variables

$$X_1 - m_{X_1},\ X_2 - m_{X_2},\ \ldots,\ X_n - m_{X_n}$$

are linearly independent if

$$\sum_{i=0}^{n} z_i(X_i - m_{X_i}) = 0 \ \text{ with probability 1,}$$

then

$$z_1 = z_2 = \cdots = z_n = 0$$



## Covariance matrices

**Theorem**

*The random variables*

$$X_1 - m_{X_1},\ X_2 - m_{X_2},\ \ldots,\ X_n - m_{X_n}$$

*are linearly independent if and only if $K_X$ is positive-definite; that is, if and only if*

$$z^t K_X z > 0 \quad \text{for all } z \neq 0.$$



## Covariance matrices

Proof: Let

$$Y = z_1 X_1 + \cdots z_n X_n = z^t X$$

Notice that

$$m_Y = \sum_{i=1}^{n} z_i m_{X_i} = z^t m_X$$

Therefore

$$Y - m_Y = z^t (X - m_X)$$

## Covariance matrices

We have

$$
\begin{aligned}
z^t K_X z &= z^t \, \mathrm{E}\left((X - m_X)(X - m_X)^t\right) z \\
&= \mathrm{E}\left(z^t \, (X - m_X)(X - m_X)^t z\right) \\
&= \mathrm{E}\left((Y - m_Y)(Y - m_Y)^t\right) \\
&= \mathrm{E}\left((Y - m_Y)^2\right) = \sigma_Y^2 \geq 0
\end{aligned}
$$



## Covariance matrices

Moreover,

$z^t K_X z = 0$ for some $z \neq 0$

$\iff \sigma_Y^2 = 0$ for some $z \neq 0$

$\iff Y - m_Y = 0$ with probability 1, for some $z \neq 0$

$\iff \sum_{i=1}^n z_i (X_i - m_{X_i}) = 0$ with probability 1,

for some $z = (z_1, z_2, \ldots, z_n)^t \neq 0$

$\iff X_1 - m_{X_1}, X_2 - m_{X_2}, \ldots, X_n - m_{X_n}$ are linealy dependent



## Linear transformations

**Theorem**

*Let $X = (X_1, X_2, \ldots, X_n)^t$ be an $n$-dimensional random variable, let $A$ be an $m \times n$ real matrix, and let $Y = (Y_1, Y_2, \ldots, Y_m)^t$ be the $m$-dimensional random variable defined by*

$$Y = AX.$$

*Then*

$$m_Y = A \, m_X,$$

$$K_Y = A \, K_X \, A^t$$



## Linear transformations

Proof: We have

$$m_Y = \mathrm{E}(Y) = \mathrm{E}(AX) = A \, \mathrm{E}(X) = A m_X,$$

and

$$
\begin{aligned}
K_Y &= \mathrm{E}\left((Y - m_Y)(Y - m_Y)^t\right) \\
&= \mathrm{E}\left(A \, (X - m_X)(X - m_X)^t \, A^t\right) \\
&= A \, \mathrm{E}\left((X - m_X)(X - m_X)^t\right) A^t = A K_X A^t
\end{aligned}
$$

## Gaussian characteristic functions

Let $X_1$, $X_2$, ..., $X_n$ be independent gaussian random variables.

Their joint characteristic function is

$$M_X(\omega_1, \omega_2, \dots, \omega_n) = M_{X_1}(\omega_1) M_{X_2}(\omega_2) \cdots M_{X_n}(\omega_n)$$

$$= \prod_{i=1}^{n} \exp\left( i\omega_i m_{X_i} - \frac{1}{2}\sigma_{X_i}^2 \omega_i^2 \right)$$

$$= \exp\left( \sum_{i=1}^{n} \left( i\omega_i m_{X_i} - \frac{1}{2}\sigma_{X_i}^2 \omega_i^2 \right) \right)$$

$$= \exp\left( i\omega^t m_X - \frac{1}{2}\,\omega^t K_X \omega \right)$$



## Gaussian characteristic functions

where

- $\omega = (\omega_1, \omega_2, \dots, \omega_n)^t$
- $m_X = (m_{X_1}, \dots, m_{X_n})$ is the expectation vector.
- $$K_X = \begin{pmatrix} \sigma_{X_1}^2 & . & . & . & 0 \\ & \sigma_{X_2}^2 & & & \\ & & . & & \\ & & & . & \\ 0 & . & . & . & \sigma_{X_n}^2 \end{pmatrix}$$

  is the covariance matrix.

  $K_X$ is diagonal because the random variables are independent and, hence, pairwise uncorrelated.



## Gaussian random vectors

**Definition**

If a random vector $X$ has characteristic function

$$M_X(\omega_1, \omega_2, \dots, \omega_n) = \exp\left( i\omega^t m - \frac{1}{2}\omega^t K \omega \right),$$

where $\omega^t = (\omega_1, \omega_2, \dots, \omega_n)$, $m$ is a column $n \times 1$ vector, and $K$ is a square positive-semidefinite $n \times n$ matrix, we say that $X$ is a $n$-dimensional gaussian random vector.

We also say that the $X_1$, $X_2$, ..., $X_n$ are jointly gaussian random variables.



## Marginal distributions

If $m = (m_j)$ and $K = (k_{ij})$ we have

$$M_{X_j}(\omega_j) = M_X(0, \dots, \omega_j, \dots, 0) = \exp\left( i\, m_j \omega_j - \frac{1}{2} k_{jj}\, \omega_j^2 \right)$$

- Each component $X_j$ is a 1-dimensional gaussian random variable with parameters

$$m_{X_j} = m_j$$

$$\sigma_j^2 = k_{jj}$$

## Marginal distributions

Moreover,

$$m_{11;X_rX_s} = \frac{1}{j^2}\left.\frac{\partial^2 M_X(\omega_1,\omega_2,\ldots,\omega_n)}{\partial\omega_r\,\partial\omega_s}\right|_{(0,0,\ldots,0)} = k_{rs} + m_r m_s$$

Therefore

$$\mathrm{Cov}(X_r, X_s) = k_{rs}$$

- $K$ is the covariance matrix of $X$.



## Eigenvalues of the covariance matrix

The matrix $K_X$ is symmetric. Thus, it can be transformed into a diagonal matrix by means of an orthogonal transformation.

- There exists an orthogonal matrix $C$ such that

$$CK_XC^t = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

- The numbers $\lambda_i \in \mathbb{R}$ are the eigenvalues of $K_X$.

Equivalently,

$$K_X = C^t DC$$



## Eigenvalues of the covariance matrix

Hence,

$$0 \le z^t K_X z = z^t C^t DCz = (Cz)^t\; D\;(Cz)$$
$$= y^t Dy = \sum_{i=1}^{n} y_i^2 \lambda_i,$$

where $y = Cz$. Therefore,

- All the eigenvalues are nonnegative:

$$\lambda_i \ge 0, \quad i = 1, 2, \ldots, n$$



## Linear independence of the components

- $K_X$ is a positive-definite matrix if and only if $\lambda_i > 0$ for all $i$. In that case, the random variables $X_i - m_{X_i}$, $1 \le i \le n$, are linearly independent.
- This condition is equivalent to $\det(K_X) \neq 0$. Only in this case there exists a density $f_X(x_1, x_2, \ldots, x_n)$.
- But gaussian random vectors are defined although $K_X$ is not necessarily invertible. (That is, $X_i - m_{X_i}$, $1 \le i \le n$, could be not all linearly independent.)

## Uncorrelation and independence

**Theorem**

*If the random variables $X_1$, $X_2$, ..., $X_n$ are jointly gaussian and pairwise uncorrelated, then they are jointly independent.*

Proof:

$$\mathrm{Cov}(X_i, X_j) = 0 \Longrightarrow K_X = \mathrm{diag}(\sigma^2_{X_1}, \sigma^2_{X_2}, \dots, \sigma^2_{X_n})$$



## Uncorrelation and independence

Therefore

$$M_X(\omega_1, \omega_2, \dots, \omega_n) = \exp\left(i\omega^t m_X - \frac{1}{2}\omega^t K_X \omega\right)$$

$$= \exp\left(\sum_{k=1}^{n}\left(i\omega_k m_{X_k} - \frac{1}{2}\sigma^2_{X_k}\omega_k^2\right)\right)$$

$$= \prod_{k=1}^{n} \exp\left(i\omega_k m_{X_k} - \frac{1}{2}\sigma^2_{X_k}\omega_k^2\right)$$

$$= M_{X_1}(\omega_1) M_{X_2}(\omega_2) \cdots M_{X_n}(\omega_n)$$



## Linear combinations

**Theorem**

*Let $X$ be an $n$-dimensional gaussian random variable, let $A$ be an $m \times n$ real matrix, and let*

$$Y = AX.$$

*Then, $Y$ is an $m$-dimensional gaussian random variable with $m_Y = A\, m_X$ and $K_Y = AK_X A^t$.*

- If $m \leq n$, $A$ has full rang $m$, and $X$ has a probability density $f_X(x_1, \dots, x_n)$, then the random vector $Y$ also has a density $f_Y(y_1, \dots, y_m)$.



## Linear combinations

Proof: We need only to prove that $Y$ is gaussian.

$$M_Y(\omega_1, \omega_2, \dots, \omega_n) = \mathrm{E}\left(e^{i\omega^t\, Y}\right) = \mathrm{E}\left(e^{i\omega^t\, AX}\right) = M_X(\omega^t A)$$

$$= \exp\left(i\ (\omega^t A) m_X) - \frac{1}{2}(\omega^t A) K_X (\omega^t A)^t\right)$$

$$= \exp\left(i\ \omega^t (A m_X) - \frac{1}{2}\ \omega^t\ (A K_X A^t)\ \omega\right)$$

$$= \exp\left(i\ \omega^t m_Y - \frac{1}{2}\omega^t K_Y \omega\right)$$

## Linear combinations

**Theorem**

*The n-dimensional random variable $X = (X_1, \dots, X_n)^t$ is gaussian if and only if the 1-dimensional random variable*

$$Y = a_1 X_1 + \cdots + a_n X_n = a^t X$$

*is gaussian for all $a = (a_1, a_2, \dots a_n)^t \in \mathbb{R}^n$*



## Linear combinations

Proof: If $X$ is gaussian so is $Y$.

Reciprocally, suppose that $Y = \omega^t X$ is gaussian for all $\omega$.

Since

$$m_Y = \omega^t m_X, \qquad \sigma_Y^2 = \omega^t K_X \omega$$

we have

$$M_X(\omega) = \mathrm{E}\left(e^{i\omega^t X}\right) = \mathrm{E}\left(e^{iY}\right)$$

$$= M_Y(1) = \exp\left(i m_Y - \frac{1}{2}\sigma_Y^2\right)$$

$$= \exp\left(i\omega^t m_X - \frac{1}{2}\omega^t K_X \omega\right)$$



## The multivariate gaussian density

Let $X_1, X_2, \dots, X_n$ be independent gaussian random variables.

Then,

$$f_X(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_{X_i}}\, e^{-\frac{1}{2}\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2}$$

$$= \frac{1}{(2\pi)^{n/2}\sigma_{X_1}\sigma_{X_2}\cdots\sigma_{X_n}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\left((x_i - m_{X_i})/\sigma_{X_i}\right)^2}$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}}\, e^{-\frac{1}{2}(x - m_X)^t\, K_X^{-1}\,(x - m_X)}$$



## The multivariate gaussian density

Hence,

$$f_X(x_1, x_2, \dots, x_n) =$$

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}} \exp\left(-\frac{1}{2}(x - m_X)^t\, K_X^{-1}\,(x - m_X)\right)$$

where

$$x = (x_1, x_2, \dots, x_n)^t$$

## The multivariate gaussian density

Now, let us consider a linear transformation

$$Y = AX,$$

being $A$ a non-singular $n \times n$ matrix.

- The linear system $y = Ax$ has a unique solution $x = A^{-1}y$.
- The jacobian determinant is

$$J(x_1, x_2, \ldots, x_n) = \det(A)$$



## The multivariate gaussian density

Moreover,

- $$m_Y = Am_X \Longrightarrow m_X = A^{-1}m_Y$$
- $$K_Y = AK_XA^t \Longrightarrow$$
$$K_Y^{-1} = (A^t)^{-1}K_X^{-1}A^{-1} = (A^{-1})^tK_X^{-1}A^{-1}$$
- $$\det(K_Y) = \det(K_X)\det(A)^2$$



## The multivariate gaussian density

In this way, we have

$$f_Y(y_1, y_2, \ldots, y_n) = \left.\frac{f_X(x_1, x_2, \ldots, x_n)}{|J(x_1, x_2, \ldots, x_n)|}\right|_{x=A^{-1}y}$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_X)}\,|\det(A)|}\cdot$$
$$\cdot \exp\left(-\frac{1}{2}(A^{-1}y - A^{-1}m_Y)^t\, K_X^{-1}\,(A^{-1}y - A^{-1}m_Y)\right)$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}}\cdot$$
$$\cdot \exp\left(-\frac{1}{2}\left(A^{-1}(y - m_Y)\right)^t\, K_X^{-1}\, A^{-1}(y - m_Y)\right)$$



## The multivariate gaussian density

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}}\cdot$$
$$\cdot \exp\left(-\frac{1}{2}(y - m_Y)^t\,(A^{-1})^tK_X^{-1}A^{-1}\,(y - m_Y)\right)$$

$$= \frac{1}{(2\pi)^{n/2}\sqrt{\det(K_Y)}}\ \exp\left(-\frac{1}{2}(y - m_Y)^t\, K_Y^{-1}\,(y - m_Y)\right)$$

- The above expression is analogous to the one we have in the case of independent random variables.

  But now, the covariance matrix $K_Y$ is not necessarily a diagonal matrix.

## The multivariate gaussian density

For instance, for $n = 2$ we obtain:

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\,\sigma_X\sigma_Y}\exp\left(-\frac{1}{2}\cdot\frac{1}{1-\rho^2}\cdot a(x,y)\right),$$

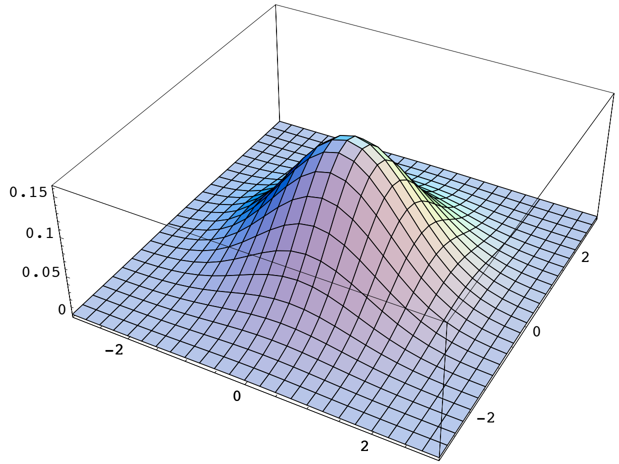
where

$$a(x,y) = \left(\frac{x-m_X}{\sigma_X}\right)^2 - 2\rho\,\frac{x-m_X}{\sigma_X}\cdot\frac{y-m_Y}{\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2$$



## The multivariate gaussian density

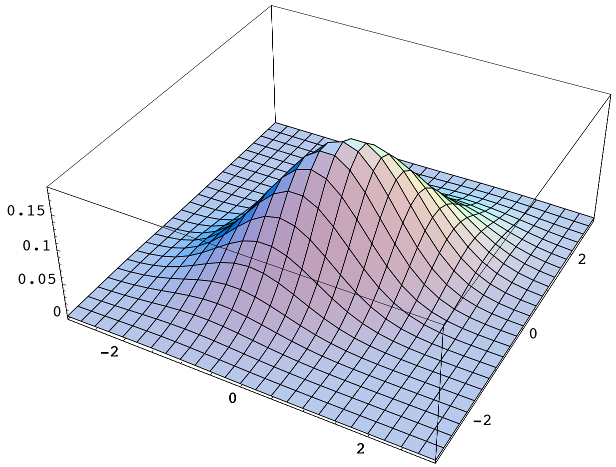
$\sigma_X = \sigma_Y \qquad \rho = 0$






## Multidimensional gaussian density

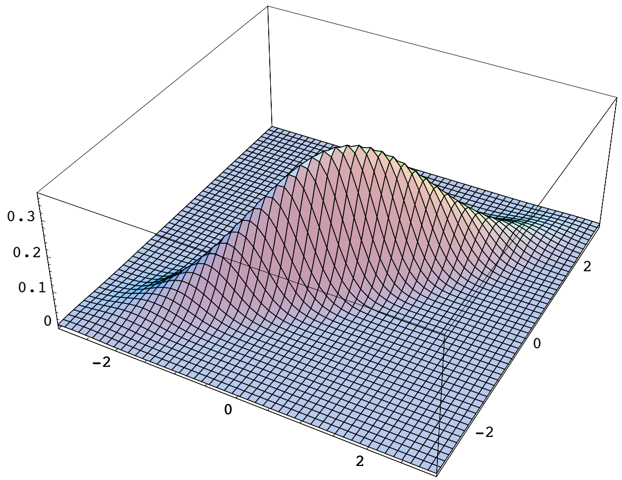
$\sigma_X = \sigma_Y \qquad \rho = 0.5$





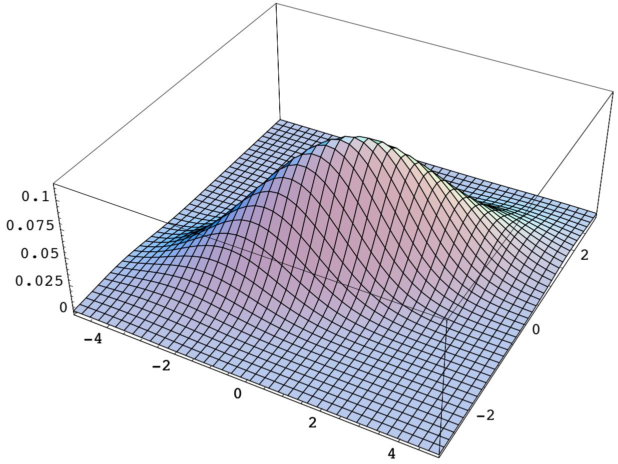

## The multivariate gaussian density

$\sigma_X = \sigma_Y \qquad \rho = 0.9$



## The multivariate gaussian density

$\sigma_X = 2\sigma_Y \qquad \rho = 0.7$






## Conditional densities

Let $X$, $Y$ be jointly gaussian. Then,

$$f_{Y|X}(y|X=x) = \frac{f_{XY}(x,y)}{f_X(x)}$$
$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\,\sigma_Y} \exp\left(-\frac{1}{2}\left(\frac{y - m_{Y|X}}{\sigma_{Y|X}}\right)^2\right)$$

- $m_{Y|X}$ is the expected value of $Y$ given $X$:

$$m_{Y|X} = \mathrm{E}(Y|X=x) = \rho\,\frac{\sigma_Y}{\sigma_X}(x - m_X) + m_Y$$

- $\sigma^2_{Y|X} = (1-\rho^2)\,\sigma^2_Y$.



## Additional results

Let $X \sim \mathrm{N}(m, K)$ denote that $X$ is a multidimensional gaussian random variable with expectation vector $m$ and covariance matrix $K$.

**Theorem**

*Let $X \sim \mathrm{N}(m, K)$ with $\det(K) > 0$. Then the random variable*

$$(X-m)^t K^{-1}(X-m)$$

*follows a $\chi^2(n)$-distribution, where $n$ is de dimension of $X$.*



## Additional results

For instance, for $n = 2$, the random variable

$$\frac{1}{1-\rho^2}\left(\left(\frac{X_1 - m_1}{\sigma_1}\right)^2 - 2\rho\,\frac{X_1 - m_1}{\sigma_1}\cdot\frac{X_2 - m_2}{\sigma_2} + \left(\frac{X_2 - m_2}{\sigma_2}\right)^2\right)$$

is $\chi^2(2)$-distributed.

## Additional results

Proof:

Let

$$CKC^t = D = \mathrm{diag}(\lambda_1, \dots, \lambda_n), \quad \lambda_i > 0,$$
$$D^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}),$$
$$K^{1/2} = C^t D^{1/2} C$$

Remember that $CC^t = C^t C = I$.

Then, $D^{1/2}D^{1/2} = D$, the matrix $K^{1/2}$ is symmetric, and

$$K^{1/2}K^{1/2} = (C^t D^{1/2} C)(C^t D^{1/2} C)$$
$$= C^t D^{1/2}(CC^t)D^{1/2}C = C^t(D^{1/2}D^{1/2})C = C^t DC = K$$



## Additional results

Let $K^{-1/2} = (K^{1/2})^{-1}$. The matrix $K^{-1/2}$ is also symmetric and $K^{-1/2}K^{-1/2} = K^{-1}$.

Now consider

$$Y = K^{-1/2}(X - m)$$

Then,

$$\mathrm{E}(Y) = \mathrm{E}\left(K^{-1/2}(X - m)\right) = K^{-1/2}\mathrm{E}((X - m)) = 0$$

$$K_Y = \mathrm{E}\left(YY^t\right) = \mathrm{E}\left(K^{-1/2}(X - m)(X - m)^t (K^{-1/2})^t\right)$$
$$= \mathrm{E}\left(K^{-1/2}(X - m)(X - m)^t K^{-1/2}\right)$$
$$= K^{-1/2}\mathrm{E}\left((X - m)(X - m)^t\right) K^{-1/2}$$
$$= K^{-1/2}KK^{-1/2} = (K^{-1/2}K^{1/2})(K^{1/2}K^{-1/2}) = I$$



## Additional results

That is, $Y \sim \mathrm{N}(0, I)$. Then,

$$(X - m)^t K^{-1}(X - m)$$
$$= (X - m)^t K^{-1/2}K^{-1/2}(X - m) = Y^t Y = \sum_{i=1}^{n} Y_i^2$$

is a (1-dimensional) random variable following a $\chi^2(n)$-distribution because the random variables $Y_i$, $1 \leq i \leq n$, are independent and $\mathrm{N}(0, 1)$.



## Additional results

**Theorem**

*Let $X \sim \mathrm{N}(m, K)$ and set $Y = CX$, where $C$ is an ortogonal matrix such that*

$$CKC^t = D = \mathrm{diag}(\lambda_1, \dots, \lambda_n).$$

*Then, $Y \sim \mathrm{N}(Cm, D)$.*

*In particular, the components of $Y$ are independent and*

$$\mathrm{Var}(Y_k) = \lambda_k, \quad k = 1, \dots, n$$

Remark: It may occur that some eigenvalue is equal to 0, in which case the corresponding component is degenerate.

## Additional results

**Theorem**

*Let $X \sim \mathrm{N}(m, \sigma^2 I)$, where $\sigma^2 > 0$. Let $C$ be an arbitrary ortogonal matrix, and set $Y = CX$. Then, $Y \sim \mathrm{N}(Cm, \sigma^2 I)$.*

*In particular, the components of $Y$ are independent gaussian random variables with the same variance $\sigma^2$.*

Proof:

$$K_Y = CK_X C^t = C(\sigma^2 I)C^t = \sigma^2 CC^t = \sigma^2 I$$



## Additional results

**Theorem**

*Let $X \sim \mathrm{N}(m, K)$, and suppose that $K$ can be partitioned (possibly after reordering the components) as follows:*

$$K = \begin{pmatrix} K_1 & & & & \\ & K_2 & & 0 & \\ & & \cdots & & \\ & 0 & & \cdots & \\ & & & & K_p \end{pmatrix}.$$

*Then, $X$ can be partitioned into vectors $X^{(1)}$, $X^{(2)}$, ..., $X^{(p)}$, where $K_i$ is the covariance matrix of $X^{(i)}$, $i = 1, 2, \ldots, p$, and in such a way that the random vectors $X^{(1)}$, $X^{(2)}$, ..., $X^{(p)}$ are independent.*



## Example

Let $X = (X_1, X_2, X_3)^t$ be a gaussian vector with $m = (0, 0, 0)^t$ and

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

Then $X_1$ and $(X_2, X_3)$ are independent.



## Example

Indeed,

$$M_X(\omega_1, \omega_2, \omega_3) = \exp\left(i\omega^t m - \frac{1}{2}\omega^t K\omega\right)$$

$$= \exp\left(-\frac{1}{2}(\omega_1, \omega_2, \omega_3)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 4 & 9 \end{pmatrix}\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}\right)$$

$$= \exp\left(-\frac{1}{2}\omega_1^2\right) \cdot \exp\left(-\frac{1}{2}(\omega_2, \omega_3)\begin{pmatrix} 2 & 4 \\ 4 & 9 \end{pmatrix}\begin{pmatrix} \omega_2 \\ \omega_3 \end{pmatrix}\right)$$

$$= M_{X_1}(\omega_1) \cdot M_{X_2 X_3}(\omega_2, \omega_3)$$