# Multivariate Statistics (Week 2): Multivariate normal distribution

Erwan Koch

EPFL (Institute of Mathematics)

# Motivation

- Easy generalization of its univariate counterpart. Multivariate analysis almost parallel to the corresponding analysis based on univariate normality.
- Entirely defined by its mean vector and covariance matrix $\implies$ only $p(p+3)/2$ parameters in all.
- Zero correlation implies independence and pairwise independence implies independence.
- Multivariate normal distribution justified by the multivariate central limit theorem.
- Linear transformations of multivariate normal rvs are again multivariate normal.
- Marginal and conditional distributions are also multivariate normal distributions.
- Contours are ellipsoids and can be made into hyperspheres by a suitable change of coordinates $\implies$ geometric simplicity.
- ...

# Outline

# Definition

- Recall that a random variable $X$ follows the univariate normal distribution with mean $\mu$ and variance $\sigma^2$ if its density is written as

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right), \quad x \in \mathbb{R}.$$

  We write $X \sim N(\mu, \sigma^2)$.

- A plausible extension to the $p$-variate case is

$$f(\mathbf{x}) = \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^p,$$

  where $\Sigma$ is a positive definite matrix (sometimes denoted $\Sigma > 0$ in the following).

## Multivariate normal distribution

Let $\Sigma$ be a positive definite matrix. A random vector $\mathbf{X}$ is said to have the $p$-variate normal (or $p$-dimensional multinormal or multivariate normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ if its density is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^p. \tag{1}$$

We write $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$.

# Some consequences of the definition

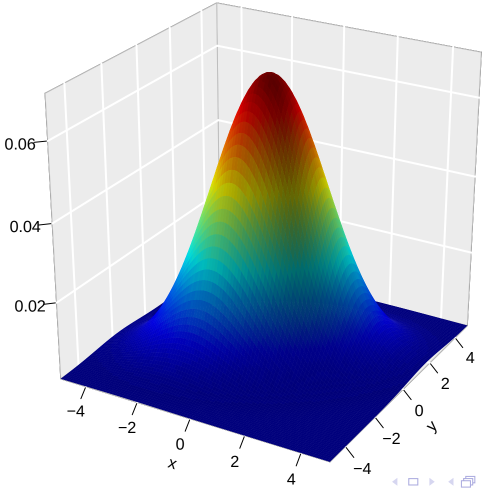
- The multivariate normal distribution (MND) is completely characterized by its mean vector $\boldsymbol{\mu}$ and its covariance matrix $\Sigma$.
- The sets of points with equal density are ellipsoids of the form

$$(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c,$$

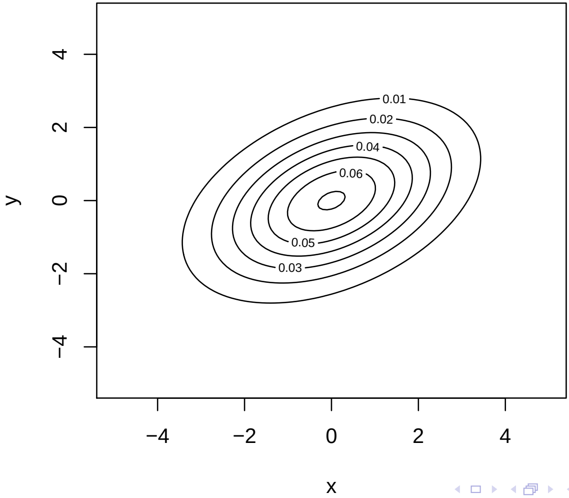
  where $c$ is a positive constant (contours of the distribution or "ellipsoids of equal concentration").
- Whenever a multivariate density $f(\mathbf{x})$ depends on $\mathbf{x}$ only through the quadratic form $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$, it is the density of a so-called elliptical distribution (see later).
- The components of $\mathbf{X}$ are mutually independent iff $\Sigma$ is diagonal, i.e., iff the components of $\mathbf{X}$ are uncorrelated.

# Density plot in the bivariate case

Normal distribution with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$



# Contours of the distribution



# Outline

## Symmetric decomposition of matrices

Any positive semi-definite matrix $\Sigma$ can be written as $\Sigma = B^2$, where $B$ is a symmetric matrix. $B$ is usually denoted $\Sigma^{1/2}$ and is termed the square root matrix of $\Sigma$. If $\Sigma$ is positive definite, then so is $\Sigma^{1/2}$.

## Theorem

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma > 0$, and let $\mathbf{Y} = (Y_1, \ldots, Y_p)' = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, where $\Sigma^{-1/2}$ is the symmetric positive definite square root of $\Sigma^{-1}$. Then $Y_1, \ldots, Y_p \overset{iid}{\sim} N(0, 1)$, which can be written

$$\mathbf{Y} \sim \mathrm{N}_p(\mathbf{0}, I_p). \tag{2}$$

## Proof.

By Week 1, we deduce that the absolute value of the Jacobian of the transformation $u$ from **Y** to **X** is

$$|J| = |\det(\Sigma^{1/2})| = \det(\Sigma^{1/2}) = \sqrt{\det(\Sigma)}.$$

Now, $\mathbf{x} = \Sigma^{1/2}\mathbf{y} + \boldsymbol{\mu} = u(\mathbf{y})$ yields

$$(u(\mathbf{y}) - \boldsymbol{\mu})'\Sigma^{-1}(u(\mathbf{y}) - \boldsymbol{\mu}) = (\Sigma^{1/2}\mathbf{y})'\Sigma^{-1}(\Sigma^{1/2}\mathbf{y}) = \mathbf{y}'\Sigma^{1/2}\Sigma^{-1/2}\Sigma^{-1/2}\Sigma^{1/2}\mathbf{y} = \mathbf{y}'\mathbf{y}.$$

Thus, the theorem about the transformation of rvs yields

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(u(\mathbf{y}))|J| &= \frac{1}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right)\sqrt{\det(\Sigma)} \\ &= \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) \\ &= \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y_i^2}{2}\right). \end{aligned}$$

$\square$

## Corollary

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma > 0$. Then

$$\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu} \quad \text{and} \quad \mathrm{V}(\mathbf{X}) = \Sigma.$$

## Proof.

Introducing $\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, we have by (2) that $\mathrm{E}[\mathbf{Y}] = \mathbf{0}$ and $\mathrm{V}(\mathbf{Y}) = I_p$. Hence, since $\mathbf{X} = \Sigma^{1/2}\mathbf{Y} + \boldsymbol{\mu}$, we directly obtain (see Week 1) $\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu}$ and

$$\mathrm{V}(\mathbf{X}) = \Sigma^{1/2}\mathrm{V}(\mathbf{Y})(\Sigma^{1/2})' = \Sigma^{1/2} I_p (\Sigma^{1/2})' = \Sigma.$$

$\square$

## Theorem

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma > 0$. Then

$$U = (\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_p.$$

## Proof.

We introduce

$$\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu}),$$

which, by (2), follows $N_p(\mathbf{0}, I_p)$. Thus

$$U = (\mathbf{Y}'\Sigma^{1/2})\Sigma^{-1}(\Sigma^{1/2}\mathbf{Y}) = \mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{p} Y_i^2$$

where $Y_1, \ldots, Y_p \overset{iid}{\sim} N(0, 1)$. The result follows by definition of the chi-square distribution. $\square$

## Characteristic function

The cf of $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, with $\Sigma > 0$, is $\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left(i\mathbf{t}'\boldsymbol{\mu} - \mathbf{t}'\Sigma\mathbf{t}/2\right)$, $\mathbf{t} \in \mathbb{R}^p$.

## Proof.

Again, we use $\mathbf{Y} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, which gives $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Y}$. Hence, we have, for $\mathbf{t} \in \mathbb{R}^p$,

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[\exp(i\mathbf{t}'\mathbf{X})] = \mathbb{E}\left[\exp(i\mathbf{t}'\boldsymbol{\mu})\right] \mathbb{E}\left[\exp(i\mathbf{t}'\Sigma^{1/2}\mathbf{Y})\right].$$

Denoting $\mathbf{u}' = \mathbf{t}'\Sigma^{1/2}$ and using the fact that $Y_1, \ldots, Y_p \overset{iid}{\sim} N(0,1)$,

$$\begin{aligned}\mathbb{E}\left[\exp(i\mathbf{t}'\Sigma^{1/2}\mathbf{Y})\right] = \mathbb{E}\left[\exp(i\mathbf{u}'\mathbf{Y})\right] = \mathbb{E}\left[\prod_{j=1}^{p} \exp(iu_j Y_j)\right] &= \prod_{i=1}^{p} \phi_{Y_i}(u_i) = \prod_{i=1}^{p} \exp\left(-u_i^2/2\right) \\ &= \exp(-\mathbf{u}'\mathbf{u}/2) = \exp(-\mathbf{t}'\Sigma\mathbf{t}/2),\end{aligned}$$

where the $u_i$ are the components of $\mathbf{u}$. □

Also true with more general definitions of the MVN, i.e., with $\Sigma$ not necessarily positive-definite but positive semi-definite (sometimes denoted $\Sigma \geq 0$ in the following); see below.

### Theorem: linear combination of components

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$. For any $\mathbf{a} \in \mathbb{R}^p \backslash \{\mathbf{0}\}$,

$$\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a}).$$

### Proof.

Let $\mathbf{a} \in \mathbb{R}^p \backslash \{\mathbf{0}\}$. The cf of $Y = \mathbf{a}'\mathbf{X}$ is

$$\phi_Y(t) = \exp(itY) = \exp(it\mathbf{a}'\mathbf{X}) = \phi_{\mathbf{X}}(t\mathbf{a}) = \exp(it\mathbf{a}'\boldsymbol{\mu} - t^2\mathbf{a}'\Sigma\mathbf{a}/2),$$

which is the cf of a normal random variable with mean $\mathbf{a}'\boldsymbol{\mu}$ and variance $\mathbf{a}'\Sigma\mathbf{a}$. □

Following the idea of the Cramér-Wold device, we can give a more general (and density-free) definition of the MND.

### Linear combinations-based definition

A $p$-dimensional rv $\mathbf{X}$ has the $p$-dimensional normal distribution iff $\mathbf{a}'\mathbf{X}$ is univariate normal for any $\mathbf{a} \in \mathbb{R}^p$.

# Comments on the linear combinations-based definition

- One can prove the existence of the MND defined in this way using the cf and univariate normality. The MND with mean $\boldsymbol{\mu}$ and covariance $\Sigma$ defined in this way is also denoted $N_p(\boldsymbol{\mu}, \Sigma)$.
- Only $\Sigma \geq 0$ is required. If $\Sigma > 0$, then the df of $\mathbf{X}$ is absolutely continuous with density given in (1).
- If $\Sigma$ is not positive-definite, we say that the MND is singular and one can define the singular density using the generalized inverse.
- Geometric interpretation.

## Theorem: linear transformation

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma \geq 0$, and $B \in \mathbb{R}^{k \times p}$, $\mathbf{b} \in \mathbb{R}^k$. We have

$$B\mathbf{X} + \mathbf{b} \sim N_k(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B'). \tag{3}$$

## Proof.

Let $\mathbf{c} \in \mathbb{R}^k$ and let us denote $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$. Then $\mathbf{c}'\mathbf{Y} = \mathbf{c}'B\mathbf{X} + \mathbf{c}'\mathbf{b} = \mathbf{a}'\mathbf{X} + \mathbf{c}'\mathbf{b}$, where $\mathbf{a} = B'\mathbf{c} \in \mathbb{R}^p$. We know that $\mathbf{a}'\mathbf{X}$ is univariate normal and thus $\mathbf{a}'\mathbf{X} + \mathbf{c}'\mathbf{b}$ as well. Thus, $\mathbf{Y}$ is multivariate normal. We have $\mathrm{E}(\mathbf{Y}) = B\mathrm{E}(\mathbf{X}) + \mathbf{b} = B\boldsymbol{\mu} + \mathbf{b}$ and $\mathrm{V}(\mathbf{Y}) = B\mathrm{V}(\mathbf{X})B' = B\Sigma B'$. The result follows. $\square$

## Corollary

Any subset of elements of a multinormal vector itself has a MND. In particular the individual elements each have univariate normal distributions.

### Theorem

1. Two $p$ and $q$-dimensional multinormal rvs $\mathbf{X}, \mathbf{Y}$ are independent iff $\mathrm{C}(\mathbf{X}, \mathbf{Y}) = 0_{p \times q}$, i.e., iff they are uncorrelated.
2. For two multinormal vectors, pairwise independence of their components implies complete independence.

### Proof.

The cf of $\mathbf{Z} = (\mathbf{X}', \mathbf{Y}')'$ factorizes as required only when the corresponding submatrix of the covariance matrix of $\mathbf{Z}$ is zero. This happens only when the vectors are uncorrelated. $\square$

### Corollary

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$ and $A \in \mathbb{R}^{k_1 \times p}$, $B \in \mathbb{R}^{k_2 \times p}$. Then $A\mathbf{X}$ and $B\mathbf{X}$ are independent iff $A \Sigma B' = 0_{k_1 \times k_2}$.

### Proof.

We know from the previous theorem that $A\mathbf{X}$ and $B\mathbf{X}$ are independent iff $\mathrm{C}(A\mathbf{X}, B\mathbf{X}) = 0_{k_1 \times k_2}$. Thus, the fact that

$$\mathrm{C}(A\mathbf{X}, B\mathbf{X}) = A\mathrm{C}(\mathbf{X}, \mathbf{X})B' = A\Sigma B'$$

yields the result. $\square$

## Partitioning and marginal distributions

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and write $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are $k$- and $(p-k)$-dimensional, respectively. Introducing in the same way $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$ and $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, we have

$$\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \mathbf{X}_2 \sim N_{p-k}(\boldsymbol{\mu}_2, \Sigma_{22}).$$

## Proof.

We have $\mathbf{X}_1 = B\mathbf{X}$, where $B = [I_k, 0_{k \times (p-k)}]$. Now, $B\boldsymbol{\mu} = \boldsymbol{\mu}_1$. Moreover, $B\Sigma = [\Sigma_{11}, \Sigma_{12}]$ and thus $B\Sigma B' = \Sigma_{11}$. Hence applying (3) we obtain the result for $\mathbf{X}_1$. The same reasoning with $B = [0_{(p-k) \times k}, I_{p-k}]$ yields the result for $\mathbf{X}_2$. □

## Convolution

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{Y} \sim N_p(\tilde{\boldsymbol{\mu}}, \tilde{\Sigma})$. If they are independent, then

$$\mathbf{X} + \mathbf{Y} \sim N_p(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma + \tilde{\Sigma}).$$

## Proof.

It is straightforward using cfs. □

# Outline

## Alternative definition

A rv $\mathbf{X} = (X_1, \dots, X_p)'$ has a multivariate normal distribution iff

$$\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + A\mathbf{Z}, \tag{4}$$

where $\mathbf{Z} = (Z_1, \dots, Z_k)'$ with $Z_1, \dots, Z_k \overset{iid}{\sim} N(0,1)$, $A \in \mathbb{R}^{p \times k}$ and $\boldsymbol{\mu} \in \mathbb{R}^p$.

- From above, we know that if $\mathbf{X}$ is defined by (4), then

$$\mathbf{X} \sim N_p(\boldsymbol{\mu} + A\mathrm{E}(\mathbf{Z}), A\mathrm{V}(\mathbf{Z})A'), \quad \text{i.e.,} \quad \mathbf{X} \sim N_p(\boldsymbol{\mu}, AA').$$

- In the non-singular case where $\mathrm{rank}(A) = p \leq k$, $\Sigma = AA'$ has full rank $p$ and is therefore invertible (non-singular) and positive definite $\implies$ the df of $\mathbf{X}$ is absolutely continuous with density given in (1).

For $\Sigma > 0$, assume that $AA' = \Sigma$ and that $\mathbf{X}$ is defined as in (4). Then $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$.
$\implies$ previous definition provides a simulation recipe for multinormal vectors.

**Cholesky factorization of matrices**

Any symmetric, positive-definite matrix $\Sigma$ can be written as $\Sigma = AA'$ for a lower triangular matrix $A$ with positive diagonal elements. $A$ is known as the Cholesky factor.

**Sampling from $N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$**

1. Perform a Cholesky decomposition of $\Sigma$ to obtain the Cholesky factor $A$.
2. Generate $Z_1, \ldots, Z_p \overset{iid}{\sim} N(0, 1)$.
3. Return $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Z}$, where $\mathbf{Z} = (Z_1, \ldots, Z_p)'$.

# Outline

# Preliminary result

## Lemma

Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\Sigma > 0$, and write $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are $k$- and $(p-k)$-dimensional, respectively. Moreover, let $\mathbf{X}_{2.1} = \mathbf{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1$. Then

$$\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \mathbf{X}_{2.1} \sim N_{p-k}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1}),$$

where

$$\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1 \quad \text{and} \quad \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Furthermore, $\mathbf{X}_1$ and $\mathbf{X}_{2.1}$ are independent.

## Proof.

We have $\mathbf{X}_1 = A\mathbf{X}$, where $A = [I_k, 0_{k\times(p-k)}]$, and $\mathbf{X}_{2.1} = B\mathbf{X}$, where $B = [-\Sigma_{21}\Sigma_{11}^{-1}, I_{p-k}]$. Equation (3) yields the first result about normality. Check it in detail! Now, it is easily checked that $A\Sigma B' = 0$, which yields the independence. $\square$

# Conditional distribution

## Theorem

Let $\mathbf{X}$ as before with $\Sigma > 0$. Then

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim N_{p-k}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22.1}).$$

## Proof.

As $\mathbf{X}_{2.1}$ is independent of $\mathbf{X}_1$, the distribution of $\mathbf{X}_{2.1}|\mathbf{X}_1$ is the same as the distribution of $\mathbf{X}_{2.1}$. Now, $\mathbf{X}_2 = \mathbf{X}_{2.1} + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1$, and the latter term is constant when $\mathbf{X}_1$ is given. Thus the conditional distribution of $\mathbf{X}_2|\mathbf{X}_1$ is multivariate normal. Moreover, its conditional mean is

$$\begin{aligned} \mathrm{E}[\mathbf{X}_2|\mathbf{X}_1] = \mathrm{E}[\mathbf{X}_{2.1}|\mathbf{X}_1] + \mathrm{E}[\Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1|\mathbf{X}_1] &= \boldsymbol{\mu}_{2.1} + \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1 \\ &= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \boldsymbol{\mu}_1), \end{aligned}$$

and

$$\mathrm{V}[\mathbf{X}_2|\mathbf{X}_1] = \mathrm{V}[\mathbf{X}_{2.1}|\mathbf{X}_1] + \mathrm{V}[\Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1|\mathbf{X}_1] = \mathrm{V}(\mathbf{X}_{2.1}) = \Sigma_{22.1}.$$

□

# Bivariate case

## Example: bivariate case

Let $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

We have

$$\Sigma_{21}\Sigma_{11}^{-1} = \rho\sigma_1\sigma_2/\sigma_1^2 = \rho\sigma_2/\sigma_1$$

and

$$\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = \sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2/\sigma_1^2 = \sigma_2^2(1-\rho^2).$$

Thus,

$$X_2|X_1 = x_1 \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1 - \mu_1), \sigma_2^2(1-\rho^2)\right).$$

- Mean is strictly increasing wrt $x_1$ if $\rho > 0$ and strictly decreasing if $\rho < 0$.
- The larger $|\rho|$, the smaller the variance, i.e., the more information $X_1$ gives about $X_2$.
- If, e.g., $\rho > 0$ and $\sigma_2 = \sigma_1$, mean of $X_2$ does not increase relative to $\mu_2$ as much as $x_1$ increases relative to $\mu_1$ $\implies$ "regression".

# Comments on the general case

- The conditional mean depends only linearly on the variates held fixed.
- In general, if there are positive covariances between $\mathbf{X}_1$ and $\mathbf{X}_2$, then a realization of $\mathbf{X}_1$ greater than $\boldsymbol{\mu}_1$ (componentwise) will result in a positive adjustment of the conditional mean (componentwise).
- Knowing that $\mathbf{X}_1 = \mathbf{x}_1$ alters the covariance matrix, but the new covariance matrix does not depend on $\mathbf{x}_1$.
- The matrix $\beta = \Sigma_{21}\Sigma_{11}^{-1}$ is called the matrix of regression coefficients of $\mathbf{X}_2$ on $\mathbf{x}_1$, and $\boldsymbol{\mu}_2 + \beta(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ is termed the regression function.
- The rv $\mathbf{X}_2 - \boldsymbol{\mu}_2 - \beta(\mathbf{X}_1 - \boldsymbol{\mu}_1)$ is called the vector of residuals of $\mathbf{X}_2$ from its regression on $\mathbf{X}_1$. It is independent of $\mathbf{X}_1$ and its covariance matrix is $\Sigma_{22.1}$.

# Partial variance, covariance and correlation

Let $\sigma_{ij.1,\ldots,k}$ be the $i,j$-th element of $\Sigma_{22.1}$, $i,j = 1,\ldots,p-k$.

## Definition: Partial variance and covariance

- For $i,j = 1,\ldots,p-k$, $\sigma_{ij.1,\ldots,k}$ is called the partial covariance between the $i$-th and $j$-th components of $\mathbf{X}_2$.
- For $i = 1,\ldots,p-k$, $\sigma_{ii.1,\ldots,k}$ is the partial variance of the $i$-th component of $\mathbf{X}_2$.

## Definition: Partial correlation

For $i,j = 1,\ldots,p-k$, the partial correlation between the $i$-th and $j$-th components of $\mathbf{X}_2$ holding $\mathbf{X}_1$ (i.e., $(X_1,\ldots,X_k)'$) fixed is defined by

$$\rho_{ij.1,\ldots,k} = \frac{\sigma_{ij.1,\ldots,k}}{\sqrt{\sigma_{ii.1,\ldots,k}}\sqrt{\sigma_{jj.1,\ldots,k}}}.$$

- These are the counterparts of classical variance, covariance and correlation, but when holding $\mathbf{X}_1$ fixed, i.e, eliminating the effect of $\mathbf{X}_1$.
- Note that partial variances, covariances and correlations do not depend on the realization $\mathbf{x}_1$ of $\mathbf{X}_1$.

# Multiple correlation coefficient

For $i = 1, \ldots, p-k$, let $X_i$ and $\mu_i$ denote the $i$-th components of $\mathbf{X}_2$ and $\boldsymbol{\mu}_2$, respectively. Moreover, let $\boldsymbol{\beta}_i'$ be the $i$-th row of $\beta$.

One can easily show that $\mu_i + \boldsymbol{\beta}_i'(\mathbf{X}_1 - \boldsymbol{\mu}_1)$ is the best linear predictor of $X_i$ in the sense that for all functions of $\mathbf{X}_1$ of the form $\mathbf{a}'\mathbf{X}_1 + c$, $\mathbf{a} \in \mathbb{R}^k$ and $c \in \mathbb{R}$, the mean squared error of the above is minimum. Related to:

## Theorem

For every vector $\boldsymbol{\alpha} \in \mathbb{R}^k$,

$$\mathrm{Corr}(X_i, \boldsymbol{\beta}_i'\mathbf{X}_1) \geq \mathrm{Corr}(X_i, \boldsymbol{\alpha}'\mathbf{X}_1).$$

## Definition: multiple correlation coefficient

The maximum correlation between $X_i$ and the linear combination $\boldsymbol{\alpha}'\mathbf{X}_1$, $\boldsymbol{\alpha} \in \mathbb{R}^k$, is called the multiple correlation between $X_i$ and $\mathbf{X}_1$.

$\Longrightarrow$ Measure of association between one variable and a set of others.

# Generalization

Let $\mathbf{X}$ be a $p$-dimensional rv (not necessarily multivariate normal) and write $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$.

- The fact that $\beta_i'\mathbf{X}_1$ is the best linear predictor and maximizes the correlation with linear functions of $\mathbf{X}_1$ depends only on the covariance structure $\Longrightarrow$ expression of the multiple correlation coefficient unchanged.
- We can still define the regression of $\mathbf{X}_2$ on $\mathbf{X}_1$ by

$$\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{X}_1 - \mu_1),$$

  and the residuals can be defined as before.
- Partial covariances and correlations can be defined as the covariances and correlations of residuals.