Linear Models in Statistics I

Lecture notes

Andrea Kraus

Department of Mathematics and Statistics
Faculty of Science
Masaryk University
Brno, Czech Republic

Autumn 2018

1. Motivation
   - What we are doing and why
2. Univariate normal distribution
   - Definition
   - Properties
   - Related distributions
3. Multivariate normal distribution
   - Definition
   - Properties
   - Related distributions

Linear model

- $Y_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k} + \varepsilon_i$, $i \in \{1, \ldots, n\}$
  - Y_i: outcome, response, output, dependent variable
    - random variable, we observe a realization y_i
    - (odezva, závisle proměnná, regresand)
  - $x_{i,1}, \ldots, x_{i,k}$: covariates, predictors, explanatory variables, input, independent variables
    - given, known
    - (nezávisle proměnné, regresory)
  - $\beta_0, \ldots, \beta_k$: coefficients
    - unknown, fixed, we want to estimate
    - (regresní koeficienty)
  - ε_i: random error
    - random variable, unobserved
- $\varepsilon_i \overset{\text{iid}}{\sim} (0, \sigma^2)$, $i \in \{1, \ldots, n\}$
  - $\mathrm{E}\,\varepsilon_i = 0$: no systematic errors
  - $\mathrm{Var}\,\varepsilon_i = \sigma^2$: same precision
- we often assume that $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, $i \in \{1, \ldots, n\}$

Linear model in matrix form

- $Y_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k} + \varepsilon_i,\ i \in \{1, \ldots, n\}$
- matrix notation:

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n,1} & x_{n,2} & \ldots & x_{n,k} \end{pmatrix}}_{\mathbf{X}} \times \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_k \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

- linear model in matrix form: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$
  and often $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
  - $\mathbf{X}$: design matrix
    - (regresní matice, matice plánu)
- let $p = k + 1$
- then $\underbrace{\mathbf{Y}}_{n\times 1} = \underbrace{\mathbf{X}}_{n\times p} \underbrace{\boldsymbol{\beta}}_{p\times 1} + \underbrace{\boldsymbol{\varepsilon}}_{n\times 1}$
- we assume that $n > p$ (and often think about $n \to \infty$, p fixed)

Linear model for `fev` data

- question: association between the FEV [l] and Smoking,
  corrected for Age [years], Height [cm] and Gender
- data:

FEV	Age	Height	Gender	Smoking
1.708	9	144.8	Female	Non
1.724	8	171.5	Female	Non
1.720	7	138.4	Female	Non
1.558	9	134.6	Male	Non
...	...	...	...	...
3.727	15	172.7	Male	Current
2.853	18	152.4	Female	Non
2.795	16	160.0	Female	Current
3.211	15	168.9	Female	Non

- $\texttt{FEV}_i = \beta_0 + \beta_1 \times \texttt{Age}_i + \beta_2 \times \texttt{Height}_i + \beta_3 \times \texttt{Gender}_i + \beta_4 \times \texttt{Smoking}_i + \varepsilon_i$
- model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\begin{pmatrix} 1.708 \\ 1.724 \\ 1.720 \\ 1.558 \\ \dots \\ 3.727 \\ 2.853 \\ 2.795 \\ 3.211 \end{pmatrix} = \begin{pmatrix} 1 & 9 & 144.8 & 0 & 0 \\ 1 & 8 & 171.5 & 0 & 0 \\ 1 & 7 & 138.4 & 0 & 0 \\ 1 & 9 & 134.6 & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 15 & 172.7 & 1 & 1 \\ 1 & 18 & 152.4 & 0 & 0 \\ 1 & 16 & 160.0 & 0 & 1 \\ 1 & 15 & 168.9 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \dots \\ \varepsilon_{651} \\ \varepsilon_{652} \\ \varepsilon_{653} \\ \varepsilon_{654} \end{pmatrix}$$

Linear model for `bloodpress` data

- question: association between the mean arterial blood pressure and age [years], weight [kg], body surface area [m^2], stress, duration of hypertension [years], basal pulse [beats/min]
- data:

MAP	Age	Weight	BSA	DoH	Pulse	Stress
105	47	85.4	1.75	5.1	63	33
115	49	94.2	2.10	3.8	70	14
...	...	...	...	...	...	...
110	48	90.5	1.88	9.0	71	99
122	56	95.7	2.09	7.0	75	99

- $\texttt{MAP}_i = \beta_0 + \beta_1 \times \texttt{Age}_i + \beta_2 \times \texttt{Weight}_i + \beta_3 \times \texttt{BSA}_i + {} + \beta_4 \times \texttt{DoH}_i + \beta_5 \times \texttt{Pulse}_i + \beta_6 \times \texttt{Stress}_i + \varepsilon_i$
- model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

$$\begin{pmatrix} 105 \\ 115 \\ \dots \\ 110 \\ 122 \end{pmatrix} = \begin{pmatrix} 1 & 47 & 85.4 & 1.75 & 5.1 & 63 & 33 \\ 1 & 49 & 94.2 & 2.10 & 3.8 & 70 & 14 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 48 & 90.5 & 1.88 & 9.0 & 71 & 99 \\ 1 & 56 & 95.7 & 2.09 & 7.0 & 75 & 99 \end{pmatrix} \times \begin{pmatrix} \beta_0 \\ \dots \\ \beta_6 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_{19} \\ \varepsilon_{20} \end{pmatrix}$$

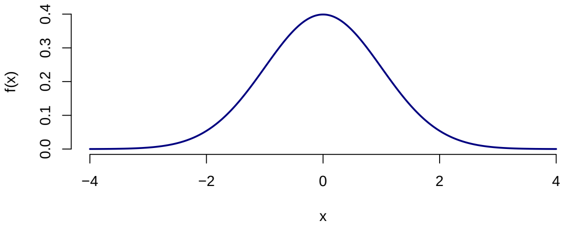
Normal distribution in a linear model

- model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- assumptions of the normal linear model:
  - $\mathbf{X}$ fixed and known
  - $\boldsymbol{\beta}$ fixed unknown
  - $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$
    $\Rightarrow \mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$
- estimators of $\boldsymbol{\beta}$ and σ^2
  - functions of $\mathbf{Y}$
- test statistics concerning $\boldsymbol{\beta}$ and σ^2
  - functions of $\mathbf{Y}$

$\Rightarrow$ to make inference in normal linear model, we need to study
  - multivariate normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - distributions of functions of $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Normal distribution $\mathrm{N}(\mu, \sigma^2)$

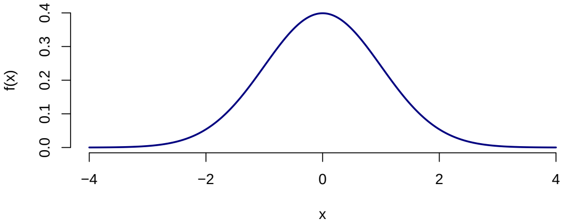
- let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$
  - density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$
  - for the standard normal distribution ($\mu = 0$, $\sigma^2 = 1$):



- if $\sigma^2 = 0$ then $X = \mu$ a.s.

Properties of $\mathrm{N}(\mu, \sigma^2)$: $\mu \in \mathbb{R}$, $\sigma^2 > 0$

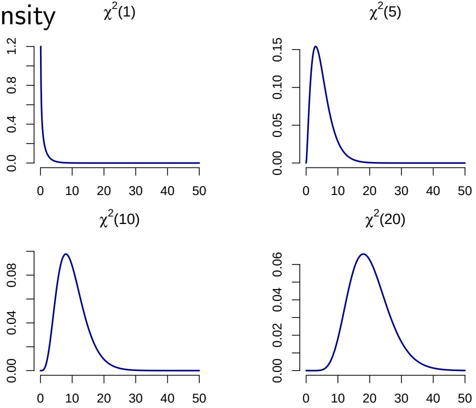
- Let $X \sim \mathrm{N}(\mu, \sigma^2)$. Then $\mathrm{E}\,X = \mu$ and $\mathrm{Var}\,X = \sigma^2$.
- Let $a, b \in \mathbb{R}$, $X \sim \mathrm{N}(\mu, \sigma^2)$. Then $aX + b \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.
  - Let $X \sim \mathrm{N}(\mu, \sigma^2)$ and $Z = \frac{1}{\sigma}(X - \mu)$. Then $Z \sim \mathrm{N}(0, 1)$.
  - If $X \sim \mathrm{N}(\mu, \sigma^2)$, then $X \overset{d}{=} \mu + \sigma Z$, where $Z \sim \mathrm{N}(0, 1)$.



- Let $a_i, b_i \in \mathbb{R}$, $X_i \overset{\text{ind.}}{\sim} \mathrm{N}(\mu_i, \sigma_i^2)$ for $i \in \{1, \ldots, n\}$.
  Then $\sum_{i=1}^n (a_i X_i + b_i) \sim \mathrm{N}\big(\sum_{i=1}^n (a_i\mu_i + b_i), \sum_{i=1} a_i^2\sigma_i^2\big)$.

$\chi^2(n)$ distribution

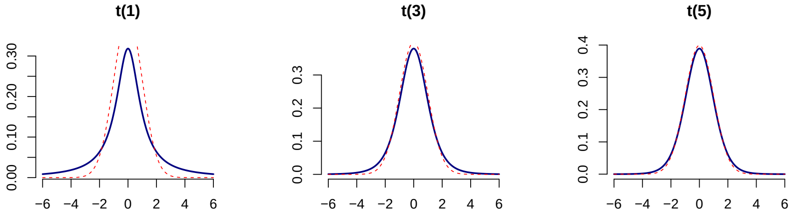
- let $Z \sim \mathrm{N}(0,1) \rightsquigarrow Z^2 \sim \chi^2(1)$
- let $Z_i \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$ for $i \in \{1,\dots,n\} \rightsquigarrow X = \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$
- density



- $\mathrm{E}\, X = n, \mathrm{Var}\, X = 2n$

Student's t-distribution

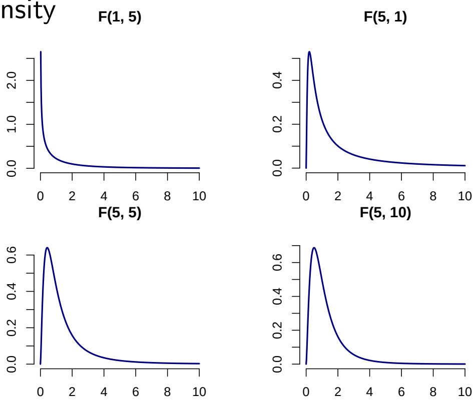
- let $Z \sim \mathrm{N}(0,1)$ and $X \sim \chi^2(n)$, $Z \perp\!\!\!\perp X$
  - $T = \frac{Z}{\sqrt{X/n}} \sim t(n)$
- density



- $\mathrm{E}\, T = 0$ for $n > 1$, $\mathrm{Var}\, T = n/(n-2)$ for $n > 2$

Fisher–Snedecor distribution

- let $X_1 \sim \chi^2(n_1)$ and $X_2 \sim \chi^2(n_2)$, $X_1 \perp\!\!\!\perp X_2$
  - $F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$
- density



- $\mathrm{E}\, F = n_2/(n_2 - 2)$ for $n_2 > 2$

Multivariate normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ positive semidefinite matrix

Definition

A random vector $\mathbf{X} : (\Omega, \mathcal{A}) \mapsto (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ has multivariate normal distribution $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $\mathbf{a}^\top \mathbf{X} \sim \mathrm{N}(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \boldsymbol{\Sigma}\, \mathbf{a})$ for every $\mathbf{a} \in \mathbb{R}^n$.

- if $\mathrm{rank}(\boldsymbol{\Sigma}) = n$ then $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is non-degenerate
  - has density

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

- if $\mathrm{rank}(\boldsymbol{\Sigma}) = r < n$ then $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is degenerate
  - a.s. "lives" in a subspace of $\mathbb{R}^n$ of dimension r
  - no density w.r.t. Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$

Properties of $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (MVN 1)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\mathrm{E}\,\mathbf{X} = \boldsymbol{\mu}$ and $\mathrm{Var}\,\mathbf{X} = \boldsymbol{\Sigma}$.

Theorem (MVN 2)

Let $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$. Then $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$.

Theorem (MVN 3)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\mathbf{A}$ be an $m \times n$ real matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{AX} + \mathbf{b} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.

Proof of MVN 1

- let $\mathbf{e}_i = (0, 0, \dots, 0, 1, 0, \dots, 0)^\top$ be a vector with 1 on the i^{th} position and 0 elsewhere
- by definition: $\mathbf{e}_i^\top \mathbf{X} \sim \mathrm{N}(\mathbf{e}_i^\top \boldsymbol{\mu}, \mathbf{e}_i^\top \boldsymbol{\Sigma}\, \mathbf{e}_i)$,
  which is $X_i \sim \mathrm{N}(\mu_i, \sigma_{i,i})$
  so $\mathsf{E}\, X_i = \mu_i$ and $\mathsf{Var}\, X_i = \sigma_{i,i}$
- let $\mathbf{e}_{i,j} = (0, 0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0)^\top$ be a vector with 1 on the i^{th} and the j^{th} positions and 0 elsewhere
- by definition: $\mathbf{e}_{i,j}^\top \mathbf{X} \sim \mathrm{N}(\mathbf{e}_{i,j}^\top \boldsymbol{\mu}, \mathbf{e}_{i,j}^\top \boldsymbol{\Sigma}\, \mathbf{e}_{i,j})$,
  which is $(X_i + X_j) \sim \mathrm{N}\{(\mu_i + \mu_j), (\sigma_{i,i} + \sigma_{i,j} + \sigma_{j,i} + \sigma_{j,j})\}$
  so $\mathsf{Var}(X_i + X_j) = \sigma_{i,i} + \sigma_{i,j} + \sigma_{j,i} + \sigma_{j,j}$
  but also $\mathsf{Var}(X_i + X_j) = \mathsf{Var}\, X_i + 2\, \mathsf{Cov}(X_i, X_j) + \mathsf{Var}\, X_j$ and $\sigma_{i,j} = \sigma_{j,i}$
  hence $\mathsf{Cov}(X_i, X_j) = \sigma_{i,j}$

Proof of MVN 2

- we will verify that $\mathbf{Z}$ satisfies the definition of $\mathrm{N}(\mathbf{0}, \mathbf{I})$
  - let $\mathbf{a} \in \mathbb{R}^n$
  - then $\mathbf{a}^\top \mathbf{Z} = \sum_{i=1}^n a_i Z_i$
  - recall that the sum of independent normals is a normal:

$$\sum_{i=1}^n a_i Z_i \sim \mathrm{N}\left(\sum_{i=1}^n a_i \times 0, \sum_{i=1}^n a_i^2 \times 1\right) =$$
$$= \mathrm{N}\left(0, \sum_{i=1}^n a_i^2\right) = \mathrm{N}(\mathbf{a}^\top \mathbf{0}, \mathbf{a}^\top \mathbf{I}\, \mathbf{a})$$

Proof of MVN 3

- we have $\underbrace{\mathbf{X}}_{n\times 1}$, $\underbrace{\mathbf{A}}_{m\times n}$ and $\underbrace{\mathbf{b}}_{m\times 1}$
- so $\mathbf{Y} = \underbrace{\mathbf{AX} + \mathbf{b}}_{m\times 1}$
- we verify that the def. of $\mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$ holds for $\mathbf{Y}$:
  - let $\mathbf{a} \in \mathbb{R}^m$
  - $\mathbf{a}^\top\mathbf{Y} = \mathbf{a}^\top(\mathbf{AX} + \mathbf{b}) = \mathbf{a}^\top\mathbf{A}\,\mathbf{X} + \mathbf{a}^\top\mathbf{b}$
  - denote $\widetilde{\mathbf{a}} = \underbrace{\mathbf{A}^\top\mathbf{a}}_{n\times 1}$
  - $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so $\widetilde{\mathbf{a}}^\top\mathbf{X} \sim \mathrm{N}(\widetilde{\mathbf{a}}^\top\boldsymbol{\mu}, \widetilde{\mathbf{a}}^\top\boldsymbol{\Sigma}\,\widetilde{\mathbf{a}})$, which is $\mathbf{a}^\top\mathbf{AX} \sim \mathrm{N}(\mathbf{a}^\top\mathbf{A}\,\boldsymbol{\mu}, \mathbf{a}^\top\mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}^\top\mathbf{a})$
  - now, $\mathbf{a}^\top\mathbf{b}$ is a constant, so $\mathbf{a}^\top\mathbf{b} \sim \mathrm{N}(\mathbf{a}^\top\mathbf{b}, 0)$ and $\mathbf{a}^\top\mathbf{b}$ is independent of $\mathbf{a}^\top\mathbf{AX}$
  - recall that sum of univariate normals is normal, so $\mathbf{a}^\top\mathbf{AX} + \mathbf{a}^\top\mathbf{b} \sim \mathrm{N}(\mathbf{a}^\top\mathbf{A}\,\boldsymbol{\mu} + \mathbf{a}^\top\mathbf{b}, \mathbf{a}^\top\mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}^\top\mathbf{a} + 0)$

Non-degenerate $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ seen through $\mathrm{N}(\mathbf{0}, \mathbf{I})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive definite matrix
- $\text{rank}(\boldsymbol{\Sigma}) = n$
  - spectral decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$
  - $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \underbrace{\mathbf{U}\boldsymbol{\Lambda}^{1/2}}_{\widetilde{\boldsymbol{\Sigma}}}\boldsymbol{\Lambda}^{1/2}\mathbf{U}^\top = \widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}^\top$
- Let $\mathbf{Z} = \widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top(\mathbf{X} - \boldsymbol{\mu})$. Then $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ (n-dimensional).
  - MVN3: Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\mathbf{A}$ be an $m \times n$ real matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{AX} + \mathbf{b} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.
  - $\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} - \widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu} = \mathbf{0}$
  - $\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\Sigma}}^{-1})^\top = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}$
- If $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \widetilde{\boldsymbol{\Sigma}}\mathbf{Z}$, where $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ (n-dimensional).

Degenerate $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ seen through $\mathrm{N}(\mathbf{0}, \mathbf{I})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix
- suppose that $\mathrm{rank}(\boldsymbol{\Sigma}) = r < n$
  - spectral decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$, $\lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_n = 0$
  - $$\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \underbrace{\mathbf{U}_{n\times r}}_{(\mathbf{u}_{,1} \mid \mathbf{u}_{,2} \mid \ldots \mid \mathbf{u}_{,r})} \underbrace{\boldsymbol{\Lambda}_{r\times r}}_{\mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}} \mathbf{U}_{n\times r}^\top = \underbrace{\mathbf{U}_{n\times r}\boldsymbol{\Lambda}_{r\times r}^{1/2}}_{\widetilde{\boldsymbol{\Sigma}}} \boldsymbol{\Lambda}_{r\times r}^{1/2}\mathbf{U}_{n\times r}^\top = \widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}^\top$$
- Let $\mathbf{Z} = \widetilde{\boldsymbol{\Sigma}}^{+}(\mathbf{X} - \boldsymbol{\mu}) = \boldsymbol{\Lambda}_{r\times r}^{-1/2}\mathbf{U}_{n\times r}^\top(\mathbf{X} - \boldsymbol{\mu})$. Then $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ (r-dimensional).
- If $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + \widetilde{\boldsymbol{\Sigma}}\mathbf{Z}$, where $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ (r-dimensional).
  - $\mathbf{X}$ a.s. “lives” in a subspace of $\mathbb{R}^n$ of dimension r

Density of non-degenerate $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive definite matrix

Theorem (MVN 4)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *where* $\operatorname{rank}(\boldsymbol{\Sigma}) = n$. *Then* $\mathbf{X}$ *has density* $f(\mathbf{x})$ *w.r.t. Lebesgue measure on* $\mathcal{B}(\mathbb{R}^n)$ *and*

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

Proof of MVN 4

- recall the density of $Z \sim \mathrm{N}(0,1)$: $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\}$
- consider a random vector $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$: $Z_i \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$
- by independence, the joint density of $\mathbf{Z}$ is

$$f(\mathbf{z}) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z_i^2\right\} \right) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} z_i^2\right\}$$

- (recall that by MVN 2, $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and note that $f(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{I})}} \exp\left\{-\frac{1}{2} \mathbf{z}^\top \mathbf{I}^{-1} \mathbf{z}\right\}$)
- now if $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X} \overset{d}{=} \boldsymbol{\mu} + \widetilde{\boldsymbol{\Sigma}}\, \mathbf{Z}$, where $\widetilde{\boldsymbol{\Sigma}} = \mathbf{U}\, \boldsymbol{\Lambda}^{1/2}$ and $\boldsymbol{\Sigma} = \mathbf{U}\, \boldsymbol{\Lambda}\, \mathbf{U}^\top$,
- so the density of $\mathbf{X}$ can be derived from that of $\mathbf{Z}$

Proof of MVN 4 ctd. (extract from last year)

Mějme zobrazení $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$, kde $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \ldots, h_n(\mathbf{x}))$. To znamená, že $h_1, \ldots, h_n$ jsou funkce proměnných $x_1, \ldots, x_n$.

Předpokládejme, že existují parciální derivace $\frac{\partial h_i(x_1,\ldots,x_n)}{\partial x_j}$ $(i, j = 1, \ldots, n)$. Matice těchto parciálních derivací se nazývá **Jacobiho matice**.

Potom **Jacobiho determinant** (**jakobián**) je determinant Jacobiho matice

$$D_{\mathbf{h}}(\mathbf{x}) = \det \frac{\partial\, \mathbf{h}}{\partial\, \mathbf{x}'} = \det \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n} \end{pmatrix} = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_n}{\partial x_1} & \cdots & \frac{\partial h_n}{\partial x_n} \end{vmatrix}$$

Označme nyní $\mathbf{y} = \mathbf{h}(\mathbf{x})$, tj. $y_1 = h_1(\mathbf{x}), \ldots, y_n = h_n(\mathbf{x})$ a připomeňme definici regulárního zobrazení.

Definice 12.1. Říkáme, že zobrazení $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$ je **regulární** v množině $M \subseteq \mathbb{R}^n$, právě když

(1) M je **otevřená** množina,

(2) funkce $h_1, \ldots, h_n$ mají **spojité první parciální derivace** v M,

(3) pro $\forall\, \mathbf{x} \in M$ je **jakobián nenulový**, tj. $D_{\mathbf{h}}(\mathbf{x}) \neq 0$.

Připomeňme, že zobrazení $\mathbf{h}$ je **prosté** na M, jestliže pro $\mathbf{x}_1, \mathbf{x}_2 \in M$ takové, že $\mathbf{x}_1 \neq \mathbf{x}_2$, je $\mathbf{h}(\mathbf{x}_1) \neq \mathbf{h}(\mathbf{x}_2)$.

Věta 12.3. Věta o hustotě transformovaného náhodného vektoru. *Nechť náhodný vektor $\mathbf{X} = (X_1, \ldots, X_n)'$ má hustotu $f_{\mathbf{X}}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$. Nechť $\mathbf{h}$ je zobrazení $\mathbb{R}^n$ do $\mathbb{R}^n$, které je regulární a prosté na otevřené množině G, kterou zobrazuje na $\mathbf{h}(G)$ a pro niž platí*

$$\int_G f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = 1.$$

Nechť $\mathbf{h}^{-1}$ je inverzní zobrazení k $\mathbf{h}$. Potom náhodný vektor $\mathbf{Y} = \mathbf{h}(\mathbf{X})$ má hustotu $f_{\mathbf{Y}}(\mathbf{y})$ tvaru

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} f_{\mathbf{X}}\left(\mathbf{h}^{-1}(\mathbf{y})\right) |D_{\mathbf{h}^{-1}}(\mathbf{y})| & \textit{pro } \mathbf{y} \in \mathbf{h}(G), \\ 0 & \textit{jinak.} \end{cases} \qquad (3.12.4)$$

Proof of MVN 4 ctd.

- $\mathbf{h} : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\mathbf{h}(\mathbf{x}) = \boldsymbol{\mu} + \widetilde{\boldsymbol{\Sigma}}\,\mathbf{x}$
- then $\mathbf{h}^{-1}(\mathbf{y}) = \widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top(\mathbf{y} - \boldsymbol{\mu})$ and

$$\det\left\{D_{\mathbf{h}^{-1}}(\mathbf{y})\right\} = \det\left\{\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top\right\} = \det\left\{\mathbf{U}\,\boldsymbol{\Lambda}^{-1/2}\right\} =$$

$$= \sqrt{\det\{\mathbf{U}\,\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top\}} = \sqrt{\det\{\boldsymbol{\Sigma}^{-1}\}} = \frac{1}{\sqrt{\det\{\boldsymbol{\Sigma}\}}}$$

- so

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top\left(\widetilde{\boldsymbol{\Sigma}}^{-1}\right)^\top \widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\} =$$

$$= \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top\mathbf{U}\,\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top(\mathbf{x} - \boldsymbol{\mu})\right\} =$$

$$= \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

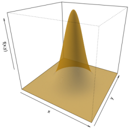
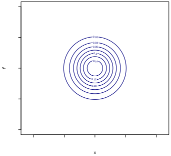
Density of non-degenerate $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

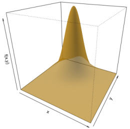
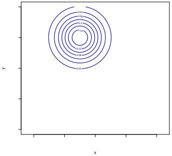
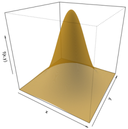
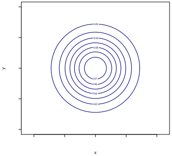
- $\boldsymbol{\Sigma}$: square symmetric positive definite matrix
  - spectral decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n > 0$
  - $\boldsymbol{\Sigma}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top$
- quadratic form $(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ can be written as
  - $(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top(\mathbf{x}-\boldsymbol{\mu}) = \{\mathbf{U}^\top(\mathbf{x}-\boldsymbol{\mu})\}^\top \boldsymbol{\Lambda}^{-1}\{\mathbf{U}^\top(\mathbf{x}-\boldsymbol{\mu})\}$
- level sets of $f(\mathbf{x})$, $l_c = \{\mathbf{x} \in \mathbb{R}^n;\ f(\mathbf{x}) = c\}$ for $c > 0$:
  - ellipsoids centred at $\boldsymbol{\mu}$
  - directions of principal axes: $\mathbf{u}_1, , \ldots, \mathbf{u}_n$,
  - lengths of principal semi-axes: $\sqrt{d\lambda_1}, \ldots, \sqrt{d\lambda_n}$

Non-degenerate bivariate normal distribution

- $$\mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$$

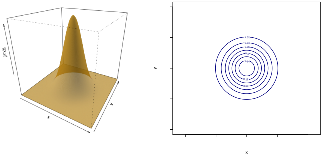



- $$\mathrm{N}\left(\begin{pmatrix}-1\\2\end{pmatrix},\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right) \qquad \mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}2 & 0\\0 & 2\end{pmatrix}\right)$$

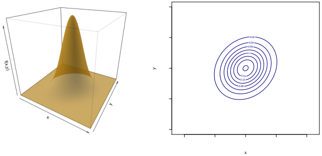





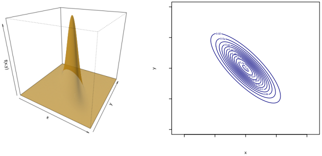
Non-degenerate bivariate normal distribution

- $$\mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}\right)$$



- $$\mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & 0.2\\0.2 & 1\end{pmatrix}\right) \qquad \mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & -0.8\\-0.8 & 1\end{pmatrix}\right)$$





Characteristic function (reminder)

Definition (Characteristic function of a random variable)

Let X be a random variable. The function $\psi_X : \mathbb{R} \mapsto \mathbb{C}$ defined by $\psi_X(t) = \mathrm{E}\exp\{\mathrm{i}\, t\, X\}$, $t \in \mathbb{R}$, is the characteristic function of X.

Definition (Characteristic function of a random vector)

Let $\mathbf{X}$ be an n-dimensional random vector. The function $\psi_{\mathbf{X}} : \mathbb{R}^n \mapsto \mathbb{C}$ defined by $\psi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}\exp\{\mathrm{i}\,\mathbf{t}^\top\mathbf{X}\}$, $\mathbf{t} \in \mathbb{R}^n$, is the characteristic function of $\mathbf{X}$.

- note that

$$\psi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}\exp\{\mathrm{i}\,\mathbf{t}^\top\mathbf{X}\} = \mathrm{E}\exp\{\mathrm{i} \times 1 \times \mathbf{t}^\top\mathbf{X}\} = \psi_{\mathbf{t}^\top\mathbf{X}}(1)$$

Properties of characteristic function (reminder)

Theorem (ChF 1)

Let $\mathbf{X}$ *be an n-dimensional random vector and* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *its subvectors such that* $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$. *Then* $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$ *iff* $\psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{X}_1}(\mathbf{t}_1) \times \psi_{\mathbf{X}_2}(\mathbf{t}_2)$ *for every* $\mathbf{t} = (\mathbf{t}_1^\top, \mathbf{t}_2^\top)^\top \in \mathbb{R}^n$.

- a proof can be found in *Petr Lachout: Teorie pravděpodobnosti (1998). Nakladatelství Univerzity Karlovy*

Theorem (ChF 2)

Let $X \sim \mathrm{N}(\mu, \sigma^2)$. *Then* $\psi_X(t) = \exp\left\{\mathrm{i}\, t\, \mu - \frac{1}{2}\sigma^2 t^2\right\}$.

Characteristic function of $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (MVN 5)

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

$$\psi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\,\mathbf{t}\right\}.$$

Proof of MVN 5

- need to compute

$$\psi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}\exp\{\mathrm{i}\,\mathbf{t}^\top \mathbf{X}\} = \psi_{\mathbf{t}^\top \mathbf{X}}(1)$$

- by definition of the multivariate normal distribution

$$\mathbf{t}^\top \mathbf{X} \sim \mathrm{N}(\mathbf{t}^\top \boldsymbol{\mu}, \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t})$$

- ChF 2: Let $X \sim \mathrm{N}(\mu, \sigma^2)$. Then $\psi_X(t) = \exp\left\{\mathrm{i}\,t\,\mu - \frac{1}{2}\sigma^2 t^2\right\}$.
- hence

$$\psi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma}\,\mathbf{t}\right\}$$

Subvectors of $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (MVN 6)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and let* $k \in \{1, \ldots, n\}$. *Then*

$$\begin{pmatrix} X_1 \\ X_2 \\ \ldots \\ X_k \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_k \end{pmatrix}, \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \ldots & \sigma_{1,k} \\ \sigma_{2,1} & \sigma_{2,2} & \ldots & \sigma_{2,k} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{k,1} & \sigma_{k,2} & \ldots & \sigma_{k,k} \end{pmatrix}\right).$$

- analogous statement is true for any sub-vector of $\mathbf{X}$
- converse is not true

Proof of MVN 6

- set $\mathbf{A} = (\mathbf{I}_{k\times k} \,|\, \mathbf{0}_{k\times(n-k)})$ and $\mathbf{b} = \mathbf{0}_{k\times 1}$
- then $(X_1, \ldots, X_k)^\top = \mathbf{AX} + \mathbf{b}$
- MVN3: Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\mathbf{A}$ be an $m \times n$ real matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{AX} + \mathbf{b} \sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$.
  - $\mathbf{A}\,\boldsymbol{\mu} + \mathbf{0} = (\mu_1, \ldots, \mu_k)^\top$
  - $$\mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}^\top = (\mathbf{I}_{k\times k} \,|\, \mathbf{0}_{k\times(n-k)}) \begin{pmatrix} \boldsymbol{\Sigma}_{1:k,1:k} & \boldsymbol{\Sigma}_{1:k,k+1:n} \\ \boldsymbol{\Sigma}_{k+1:n,1:k} & \boldsymbol{\Sigma}_{k+1:n,k+1:n} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{k\times k} \\ \mathbf{0}_{(n-k)\times k} \end{pmatrix} = \boldsymbol{\Sigma}_{1:k,1:k}$$

(In)dependence in $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (MVN 7)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and let* $k \in \{1, \dots, n-1\}$. *Denote* $\mathbf{X}_1 = (X_1, \dots, X_k)^\top$, $\mathbf{X}_2 = (X_{k+1}, \dots, X_n)^\top$ *and* $\mathbf{X}_1 \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{1,1})$, $\mathbf{X}_2 \sim \mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{2,2})$.
If

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{1,1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{2,2} \end{pmatrix}$$

then $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$.

- $\mathbf{AX} \perp\!\!\!\perp \mathbf{BX}$ iff $\mathbf{A\Sigma B}^\top = \mathbf{0}$

Proof of MVN 7

- write $\mathbf{t} = (\mathbf{t}_1^\top, \mathbf{t}_2^\top)^\top$, $\mathbf{t}_1 \in \mathbb{R}^k$, $\mathbf{t}_2 \in \mathbb{R}^{(n-k)}$
- recall that
  - MVN 5: Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\psi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{\mathrm{i}\,\mathbf{t}^\top\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\,\mathbf{t}\right\}$.
- and compute

$$\begin{aligned}\psi_{\mathbf{X}}(\mathbf{t}) &= \exp\left\{\mathrm{i}\mathbf{t}^\top\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t}\right\}\\ &= \exp\left\{\mathrm{i}\mathbf{t}_1^\top\boldsymbol{\mu}_1 + \mathrm{i}\mathbf{t}_2^\top\boldsymbol{\mu}_2 - \frac{1}{2}\mathbf{t}_1^\top\boldsymbol{\Sigma}_{1,1}\mathbf{t}_1 - \frac{1}{2}\mathbf{t}_2^\top\boldsymbol{\Sigma}_{2,2}\mathbf{t}_2\right\}\\ &= \psi_{\mathbf{X}_1}(\mathbf{t}_1)\psi_{\mathbf{X}_2}(\mathbf{t}_2)\end{aligned}$$

- this implies independence of $\mathbf{X}_1$, $\mathbf{X}_2$ by
  - ChF 1: Let $\mathbf{X}$ be an n-dimensional random vector and $\mathbf{X}_1$ and $\mathbf{X}_2$ its subvectors such that $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$. Then $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_2$ iff $\psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{X}_1}(\mathbf{t}_1) \times \psi_{\mathbf{X}_2}(\mathbf{t}_2)$ for every $\mathbf{t} = (\mathbf{t}_1^\top, \mathbf{t}_2^\top)^\top \in \mathbb{R}^n$.

Proof of the corollary

- the corollary follows from the multivariate normality of $((\mathbf{AX})^\top, (\mathbf{BX})^\top)^\top$ with $\mathrm{Cov}(\mathbf{AX}, \mathbf{BX}) = \mathbf{A\Sigma B}^\top$:
  - by MVN 3

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{X} = \begin{pmatrix} \mathbf{A\,X} \\ \mathbf{B\,X} \end{pmatrix} \mathbf{X} \sim \mathrm{N}\left(\dots, \begin{pmatrix} \mathbf{A\Sigma\,A}^\top & \mathbf{A\Sigma\,B}^\top \\ \mathbf{A\Sigma\,B}^\top & \mathbf{B\Sigma\,B}^\top \end{pmatrix}\right)$$

- "$\Rightarrow$" independence implies zero covariance
- "$\Leftarrow$" follows from MVN7

Quadratic forms

- Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (QF 1)

Let $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$*. Then* $\mathbf{Z}^\top \mathbf{Z} \sim \chi^2(n)$*.*

Theorem (QF 2)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *where* $\mathrm{rank}(\boldsymbol{\Sigma}) = n$*. Then* $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(n)$*.*

Theorem (QF 3)

Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *where* $\mathrm{rank}(\boldsymbol{\Sigma}) = r < n$*. Then* $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{+} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(r)$*.*

Proof of QF 1

- obviously $\mathbf{Z}^\top \mathbf{Z} = \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$

Proof of QF 2

- let $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$, $\boldsymbol{\Sigma}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^\top$
- define $\mathbf{Z} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^\top(\mathbf{X} - \boldsymbol{\mu})$, we see that $\mathbf{Z} \sim \mathrm{N}_n(\mathbf{0}, \mathbf{I})$
- therefore,

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\top \mathbf{Z} \sim \chi^2(n)$$

Proof of QF 3

- let $\boldsymbol{\Sigma} = \mathbf{U}_{n\times r}\boldsymbol{\Lambda}_{r\times r}\mathbf{U}_{n\times r}^\top$, $\boldsymbol{\Sigma}^{+} = \mathbf{U}_{n\times r}\boldsymbol{\Lambda}_{r\times r}^{-1}\mathbf{U}_{n\times r}^\top$
- define $\mathbf{Z} = \boldsymbol{\Lambda}_{r\times r}^{-1/2}\mathbf{U}_{n\times r}^\top(\mathbf{X} - \boldsymbol{\mu})$, we see that $\mathbf{Z} \sim \mathrm{N}_r(\mathbf{0}, \mathbf{I})$
- therefore,

$$(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{+}(\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}^\top \mathbf{Z} \sim \chi^2(r)$$

Quadratic forms

- Let $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\Sigma}$ is an $n \times n$ symmetric positive semidefinite matrix

Theorem (QF 4)

Let $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$ and let $\mathbf{P}$ be an $n \times n$ orthogonal projection matrix of rank r. Then $\mathbf{Z}^\top \mathbf{P} \mathbf{Z} \sim \chi^2(r)$.

Proof of QF 4

- $\mathbf{P}$ can be written as $\mathbf{P} = \mathbf{U}_{n\times r}\mathbf{U}_{n\times r}^\top$
- then $\mathbf{U}_{n\times r}^\top \mathbf{Z} \sim \mathrm{N}_r(\mathbf{0}, \mathbf{I})$
- therefore,

$$\mathbf{Z}^\top \mathbf{P} \mathbf{Z} = (\mathbf{U}_{n\times r}^\top \mathbf{Z})^\top (\mathbf{U}_{n\times r}^\top \mathbf{Z}) \sim \chi^2(r)$$