Charles University in Prague

Faculty of Mathematics and Physics

BACHELOR THESIS

First and last name of the author

Title of the thesis

Name of the department or institute

Supervisor of the bachelor thesis: RNDr. Name Surname, Ph.D.

Study programme: Mathematics

Specialization: Specialization

Prague 2013

I thank ... for .... (Not required.)

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In . . . . on . . . .   Author signature

Název práce: Czech title according to SIS

Autor: First and last name of the author

Katedra: Czech name of the department or institute

Vedoucí bakalářské práce: RNDr. Name Surname, Ph.D., department

Abstrakt: The abstract in Czech, 80 – 200 words long, not a copy of the assignment.

Klíčová slova: 3 to 5 Czech keywords

Title: English title according to the thesis assignment

Author: First and last name of the author

Department: English name of the department or institute

Supervisor of the bachelor thesis: RNDr. Name Surname, Ph.D., department

Abstract: The abstract in English, 80 – 200 words long, not a copy of the assignment.

Keywords: 3 to 5 English keywords

Contents

Chapter 1

Typesetting of mathematical text

This chapter demonstrates a few examples of mathematical text typesetting.

1.1 Basic examples

A number in the mathematical mode with decimal point: $\pi \doteq 3.141\,592\,653\,589$.
Test on a 5% level, 95% confidence interval.
We have $\mathrm{var}(X) = \mathsf{E}\,X^2 - \left(\mathsf{E}\,X\right)^2$.

1.2 Mathematical expressions

Let
$$
\mathbb{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}.
$$
Note the period after the matrix. The mathematical text is a part of the sentence grammar and requires standard punctuation. Expressions that will be referenced should be numbered:
$$
\mathbb{X} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_n^\top \end{pmatrix}. \tag{1.1}
$$
Expression (1.1) defines matrix $\mathbb{X}$. To achieve better readability and organization, it is advised to number only those expressions that are referenced. Do not number all displayed expressions automatically.

Aligning of expressions into several columns:
$$
\begin{aligned}
S(t) &= \mathsf{P}(T > t), & t &> 0 & &\text{(right-continuous)}, \\
F(t) &= \mathsf{P}(T \leq t), & t &> 0 & &\text{(right-continuous)}.
\end{aligned}
$$

Two binded equations:
$$
\left.\begin{aligned}
S(t) &= \mathsf{P}(T > t) \\
F(t) &= \mathsf{P}(T \leq t)
\end{aligned}\right\} \quad t > 0 \quad \text{(right-continuous)}. \tag{1.2}
$$

Two centered unnumbered equations:

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\mathbb{X} = \begin{pmatrix} 1 & \boldsymbol{x}_1^\top \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_n^\top \end{pmatrix}.$$

Two centered numbered equations:

$$\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1.3}$$

$$\mathbb{X} = \begin{pmatrix} 1 & \boldsymbol{x}_1^\top \\ \vdots & \vdots \\ 1 & \boldsymbol{x}_n^\top \end{pmatrix}. \tag{1.4}$$

Definition organized by cases:

$$P_{r-j} = \begin{cases} 0 & \text{if } r - j \text{ is odd,} \\ r!\,(-1)^{(r-j)/2} & \text{if } r - j \text{ is even.} \end{cases}$$

Note the use of punctuation in the equation. Commas and periods are placed according to the standard English language rules.

$$\begin{aligned} x &= y_1 - y_2 + y_3 - y_5 + y_8 - \ldots && \text{by (1.3)} \\ &= y' \circ y^* && \text{by (1.4)} \\ &= y(0)y' && \text{by Axiom 1.} \end{aligned} \tag{1.5}$$

Two unnumbered aligned equations:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_i(y_i; \boldsymbol{\theta}),$$

$$\ell(\boldsymbol{\theta}) = \log\{L(\boldsymbol{\theta})\} = \sum_{i=1}^{n} \log\{f_i(y_i; \boldsymbol{\theta})\}.$$

Two aligned equations, the first numbered:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_i(y_i; \boldsymbol{\theta}), \tag{1.6}$$

$$\ell(\boldsymbol{\theta}) = \log\{L(\boldsymbol{\theta})\} = \sum_{i=1}^{n} \log\{f_i(y_i; \boldsymbol{\theta})\}.$$

Two-line equation, the first line aligned left, the second line aligned right, unnumbered:

$$\ell(\mu, \sigma^2) = \log\{L(\mu, \sigma^2)\} = \sum_{i=1}^{n} \log\{f_i(y_i; \mu, \sigma^2)\}$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.$$

Two-line equation, aligned to =, numbered in the middle:

$$\begin{aligned}
\ell(\mu,\,\sigma^2) = \log\big\{L(\mu,\,\sigma^2)\big\} &= \sum_{i=1}^{n} \log\big\{f(y_i;\,\mu,\,\sigma^2)\big\} \\
&= -\frac{n}{2}\,\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2.
\end{aligned} \tag{1.7}$$

1.3   Definitions, theorems, proofs, . . .

Constructions like definitions, theorems, examples, . . . , should be separated from the surrounding text and (usually) numbered, with the use of cross-references. For each of these constructions, the main source file (`BcPrace.tex`) should define an environment providing its visual separation from the surrounding text and numbering/cross-referencing features.

Definition 1. *Let random variables $X_1, \ldots, X_n$ be defined on the same probability space $(\Omega,\, \mathcal{A},\, P)$. Then the vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is called a* random vector.

Definition 2 (random vector)**.** *Let random variables $X_1, \ldots, X_n$ be defined on the same probability space $(\Omega,\, \mathcal{A},\, P)$. Then the vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is called a* random vector.

Definition 1 demonstrates an environment without a subtitle, definition 2 includes a subtitle.

Theorem 1. *The random vector $\boldsymbol{X}$ is a measurable mapping of the space $(\Omega,\, \mathcal{A},\, P)$ to $(\mathbb{R}_n,\, \mathcal{B}_n)$.*

Lemma 2 (Anděl, 2007, str. 29)**.** *The random vector $\boldsymbol{X}$ is a measurable mapping of the space $(\Omega,\, \mathcal{A},\, P)$ to $(\mathbb{R}_n,\, \mathcal{B}_n)$.*

Proof.   The detailed steps of the proof are described in the book Anděl (2007, p. 29).

❑

Theorem 1 demonstrates an environment without a subtitle, lemma 2 includes a reference in the subtitle. Lemmas were defined in the main file so that they share the same numbering with theorems.

Chapter 2

Citations

Citations are created by commands `\citet`, `\citep` etc. (see LaTeX package `natbib`) and processed by running BibTeX. In mathematical publications, citations are usually formatted as "Author name (year of publication)" or "Author name [number of the reference])". Citations by author name are easier to handle in English than in Czech, because of its simpler grammar rules.

2.1 Some examples

Kaplan and Meier (1958); Cox (1972) belong among the most highly cited statistical papers. Student (1908) wrote a paper on the t-test.

Prof. Anděl authored a famous textbook on mathematical statistics (see Anděl, 1998). The theory of estimation is the topic of Lehmann and Casella (1998). When a reference to a specific result (definition, theorem, proof,...) is made it is useful to include the number of the chapter, page, or theorem in the citation, e.g., Anděl (2007, Theorem 4.22) or (see Anděl, 2007, Chapter 4).

Some papers have many coauthors. When citing a paper with three coauthors we usually list all of them at the first ocassion: Dempster, Laird, and Rubin (1977) introduced the concept of the EM algorithm. Alternatively: The concept of the EM algorithm was introduced in the late seventies (Dempster, Laird, and Rubin, 1977). When that same paper is cited again, we use a short citation: Dempster et al. (1977) offer several examples of the use of the EM algorithm.

Papers with more than three coauthors are always cited in the short form: The first results from Project ACCEPT were published by Genberg et al. (2008). *We never cite this paper as* Genberg, Kulich, Kawichai, Modiba, Chingono, Kilonzo, Richter, Pettifor, Sweat, and Celentano (2008).

Chapter 3

Tables, figures, software code

The inclusion of tables and figures in a scientific publication follows certain common and certain specific rules. Tables and figures are not included inside the text but placed either on dedicated pages or floated at the top or the bottom of a text page. LATEX handles floating figures and tables automatically. Every table and figure must be numbered and accompanied with a legend. The legend should describe the contents of the table of figure with enough detail so that the reader can understand them without studying the text of the publication. Each table and figure should be refrenced by its number in the text. The text should summarize the most important conclusions that can be drawn from the table of figure. The text should be easy to follow and understand even without seeingf the figures and tables (and on the contrary, the figures and tables should be easy to understand even without reading the text). Figures and tables should be referenced indirectly in the sentences: instead of *"Table 3.1 shows that men are on average 9.9 kg heavier than women"* we write *"Men are on average 9,9 kg heavier than women (see Table 3.1).*

3.1 Tables

Tables should be formatted according to the following rules:

- Avoid vertical lines. Separate the table from the surrounding text (even the legend) by stronger horizontal lines. Separate the header from the table and different parts of the table from each other by thinner horizontal lines. This table format can be obtained in LATEX by loading the package `booktabs`. If a stronger separation of the columns is desired it can be achieved by including an additional vertical space.

Effect	Estimate	Std. error[a]	P-value
Intercept	-10.01	1.01	—
Gender (male)	9.89	5.98	0.098
Height (cm)	0.78	0.12	< 0.001

Note: [a] Standard error of the estimator by the Monte Carlo method.

Table 3.1: Maximum likelihood estimates from model M.

- Do not change the type, format and meaning of the cells in a single column (never include means in some cells and percentages in other cells of the same column).
- Do not repeat the same cell contents many times. If the table includes the column "Variance", which contains the value 0.5 in the first ten cells and 1.5 in the next ten cells, delete this column and find another way to communicate the value of the variance. For example, the table can be divided into two separate tables and the variance given in the legend. Or additional rows can be included between the parts of the table informing what was the variance in the subsequent rows.
- Numeric columns in the table should be aligned on the decimal point.
- Tables sometimes include abbreviations that are not used elsewhere in the text. These abbreviations should be explained in the legend or in notes under the table. These notes can be also used to provide more detail on the meaning of selected columns or cells.

3.2 Figures

Some advice on figures and diagrams:
- Create the figure in the same size that will be used in the thesis. Excessive magnification or reduction of figures causes poor readability.
- The axes of a graph must be carefully annotated in the same language the thesis is written in. Units of measurement (kg, minutes, . . .) should be provided when applicable. When the graph plots a function $h(x)$ the axes should be annotated by x a $h(x)$. Each axis must have a clearly defined scale (tickmarks, labels).
- If a two-dimensional scatterplot includes a large number of points make sure that the points do not turn into black cloud. If the number of points is too large reduce the size of the plotting symbol or select a subset of the points. Plots that include thousands of points make problems in electronic documents, they increase the size too much.
- If the thesis is to be printed on a black-and-white printer, do not use colors. Lines can be distinguished by the type (solid, dotted, . . .), areas can be filled by various shades of grey or
  The meaning of the line types and area shading should be explained in the legend or directly in the plot.

The command \psfrag can replace parts of ps/eps files (usually annotations in the graphs) by an arbitrary sequence of LaTeX commands, as the following examples illustrate.

3.3 Software code

Software code or computer output (if needed in the thesis) should be formatted differently from the other text. One option is to use LaTeX package fancyvrb (fancy verbatim), which is used to define the environment PCinout in the master file BcPrace.tex. With this environment, we can generate the following examples:

```
> mean(x)
[1] 158.90
> objekt$mean
[1] 158.90
```

Smaller font:

```
> mean(x)
[1] 158.90
> objekt$mean
[1] 158.90
```

No frame:

```
> mean(x)
[1] 158.90
> objekt$mean
[1] 158.90
```

Narrow frame:
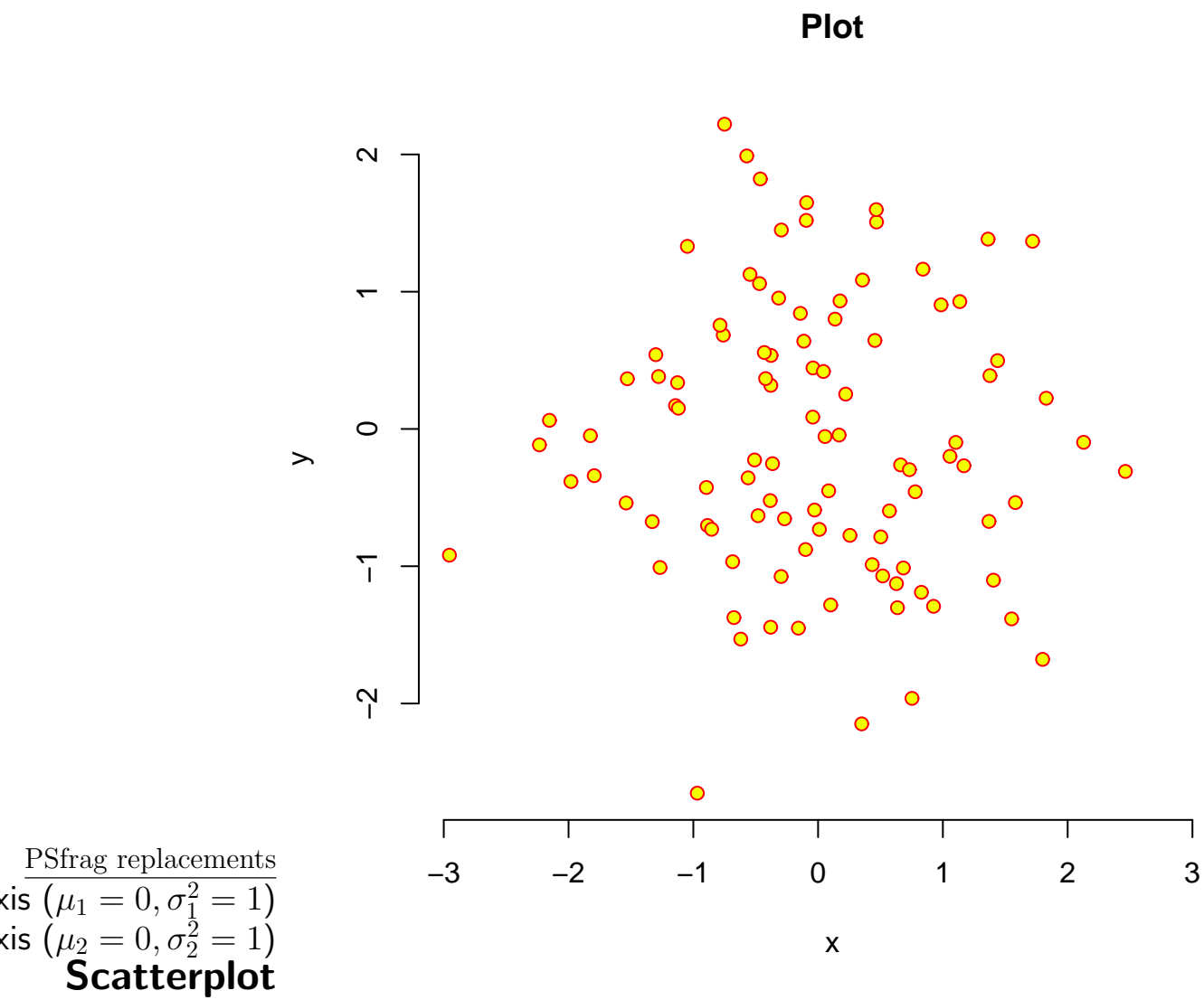
```
> mean(x)
[1] 158.90
> objekt$mean
[1] 158.90
```

Plot

xis $(\mu_1 = 0, \sigma_1^2 = 1)$
xis $(\mu_2 = 0, \sigma_2^2 = 1)$
Scatterplot

Figure 3.1: Random sample from distribution $\mathcal{N}_2(\mathbf{0}, I)$.
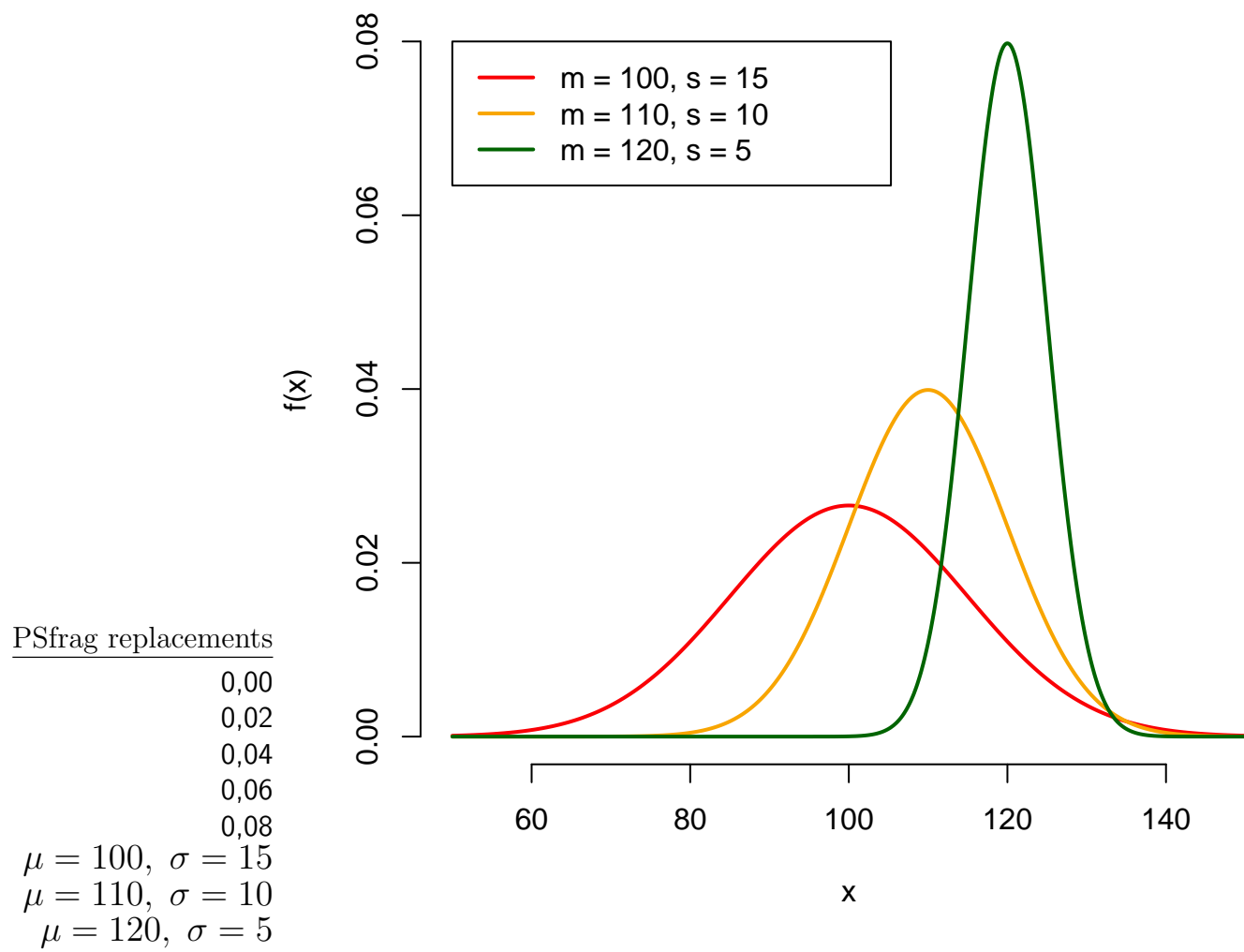
Figure 3.2: Densities of several normal distributions.

m = 100, s = 15

0,00
0,02
0,04
0,06
0,08
60
80
100
120
140
f(x)

$\mu = 100, \sigma = 15$
$\mu = 110, \sigma = 10$
$\mu = 120, \sigma = 5$

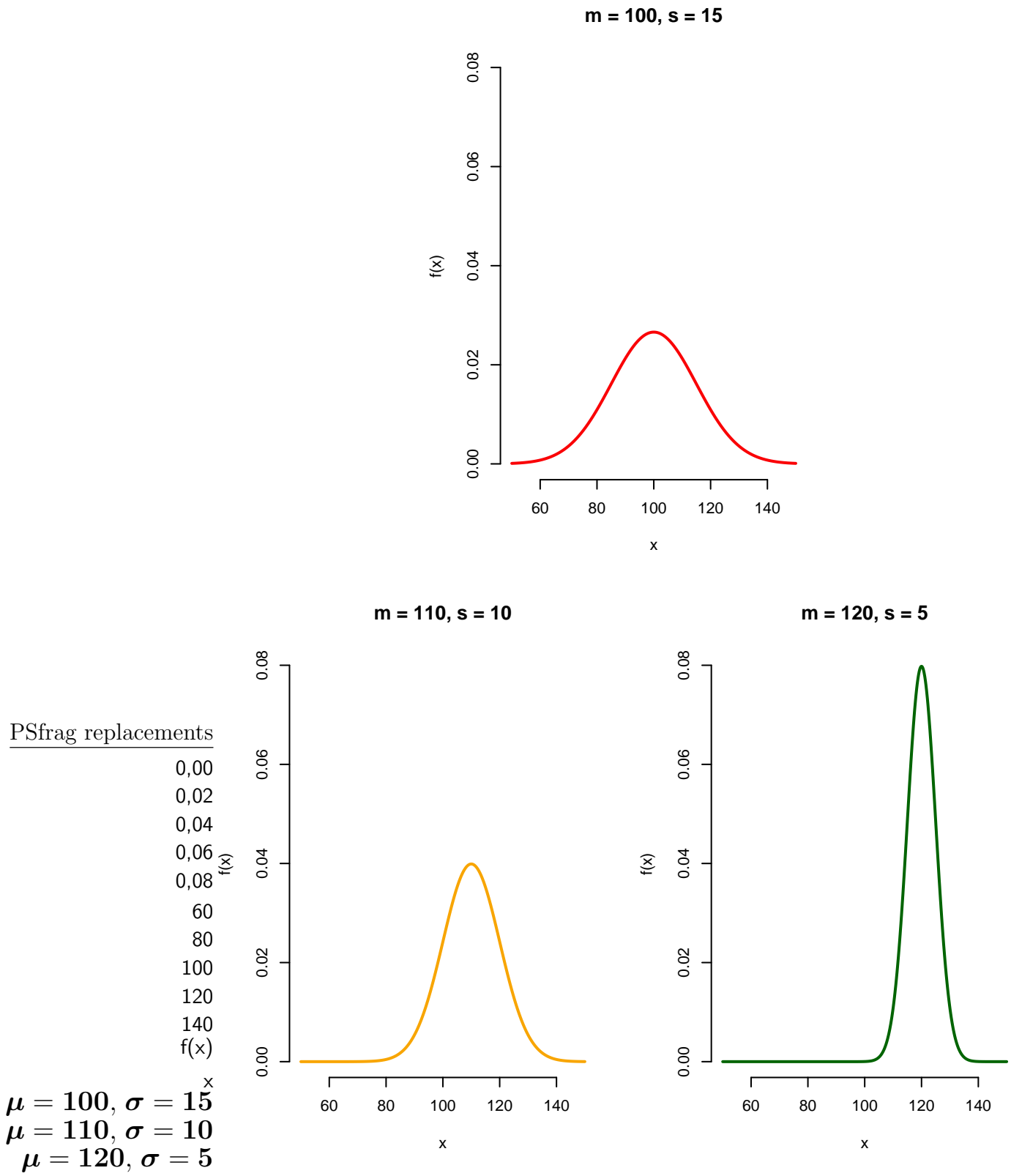**m = 110, s = 10**

m = 120, s = 5

Figure 3.3: Densities of several normal distributions.

Bibliography

J. Anděl. *Statistické metody.* Druhé přepracované vydání. Matfyzpress, Praha, 1998. ISBN 80-85863-27-8.

J. Anděl. *Základy matematické statistiky.* Druhé opravené vydání. Matfyzpress, Praha, 2007. ISBN 80-7378-001-1.

D. R. Cox. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, 34(2):187–220, 1972.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

B. L. Genberg, M. Kulich, S. Kawichai, P. Modiba, A. Chingono, G. P. Kilonzo, L. Richter, A. Pettifor, M. Sweat, and D. D. Celentano. HIV risk behaviors in sub-Saharan Africa and Northern Thailand: Baseline behavioral data from project Accept. *Journal of Acquired Immune Deficiency Syndrome*, 49:309–319, 2008.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

E. L. Lehmann and G. Casella. *Theory of Point Estimation.* Second Edition. Springer-Verlag, New York, 1998. ISBN 0-387-98502-6.

Student. On the probable error of the mean. *Biometrika*, 6:1–25, 1908.

List of Figures

List of Tables