Style Guide for Loss Data Analytics

An open text authored by the Actuarial Community

2018-07-22

Contents

Preface

It is important for chapters in the Open Actuarial Textbooks project to have a consistent look and feel. The Style Guide for Loss Data Analytics aims to assist contributors in developing chapters consistently throughout the Loss Data Analytics book.

This guide is set up as sample chapters containing

- information regarding suitable content for a chapter,
- methods to implement chapter elements and
- conventions for consistent notation.

Chapter 1

Chapter Structure

In this chapter, you learn how to:

- Determine what and what not to include in a chapter
- Include technical supplements as needed
- Assess types of exercises and book resources that are appropriate for a chapter

---

1.1 Chapter Preview and Learning Objectives

- Chapter Preview. Begin with a chapter preview to set the stage of the chapter.
- Learning Objectives. At the beginning of each chapter and section, describe in a few bullet points what the reader can expect to learn.

1.2 Main Body

Split the chapter into 4-7 sections; within each section, introduce 0-5 subsections. Do not develop a deeper hierarchy (e.g., a "sub-subsection"). Use nonlinear aspects of the web. For example, detailed mathematical developments can go into a technical appendix or are simply hidden (using javascript "hide/show" tools) unless the viewer really wants to see the details. Case studies and historical references can be included in "side-bars," a supporting webpage. For the main body of the chapter, think about "25 pages" in length (whatever that means....).

1.2.1 What to Include

- Within the chapter, use boxed and numbered lists of procedures for easy reference.
- It is certainly okay (and expected) to use mathematical notation although please adhere to the conventions described in Section 3. Each chapter should have examples interwoven within theory, allowing readers to see the development of the theory along with the importance of the applications.
- Distinguish between an "Example" and a "Special Case". The former shows how to relate the mathematics to a practical situation likely to be encountered by a practicing actuary. The latter looks at a

subset of a general (usually) mathematical result. A few special cases are certainly acceptable but we want to focus on developing examples.

- Think of graphical ways to visualize/summarize relationships that you want to emphasize.
- Begin each section with a short bullet list describing the learning objectives of that section. Finish each section with a short quiz on these learning objectives. As of this writing (July 2018), quizzes are multiple-choice.
- Include short exercises/examples/special cases that can be readily solved by the viewer (with solutions using "hide/show" features) within the main body. These serve to reinforce concepts and provide benchmarks for understanding.

1.2.2 What Not to Include

- Do not include development of equations/formulas in the main body of the text. The main body of the text will be devoted to presenting results, providing context and intuition as to the importance of the results.
- Do not include references to the literature. This will appear in the last section on "Further Reading and References."
- Do not include graphs whose information could easily be summarized by a table.

1.3 Technical Supplements

We want our viewers to understand the underpinnings of the theory (the old analogy of "what is going on under the hood to see how the engine works" - no black boxes.) So, there will be occasions when you feel like a short development or "proof"; is reasonable. Put this in an appendix. Technical supplements should develop the theory in a step-by-step fashion, building on each concept in a crisp, mathematical fashion.

1.4 Contributors and Further Resources

1.4.1 Contributors

Make sure that contributors are listed at the end of the chapter. The following provides an example.

```
####Contributors {-}

- Edward W. (Jed) Frees, University of Wisconsin-Madison,
is the principal author of the initital version of this chapter.
Email: jfrees@bus.wisc.edu for chapter comments and
suggested improvements. Helpful improvements provided by
Alyaa Nuval Binti Othman and Aisha Nuval Binti Othman.
```

1.4.2 Further Reading and References

Do not finish with a "preview of upcoming chapter"; finish instead with a "Further Reading and References." This consists of a series of references with one or two lines of annotation for each reference that the interested reader could follow up on (self-citations are okay!). Historical developments are particularly nice in this section.

1.5 Other Supporting Elements

1.5.1 Exercises

We anticipate that substantial exercise banks will be built over time by users, professional associations, and those with commercial interests. At this early stage of developing the chapter foundations, we recommend developing the following types of exercises. Exercises will be segmented by section (not subsection) and will be positioned at the end of each chapter.

- **Hand Calculation.** Similar to those appearing within the chapter, include at the end of the chapter short exercises/examples/special cases that can be readily solved by the viewer.
- **Software.** Include exercises that ask the viewer to work with "R" software, such as calculating a function or reproducing a graph.
- **Data.** The need for working with real data is well documented; for example, see Hogg (1972), Moore and Roberts (1989) or Singer and Willett (1990). By providing detailed guided tutorials that work with theory and data, we teach our students the essence of Loss Data Analtyics. Of course, there are some important disadvantages to working with real data. Data sets can quickly become outdated. Further, the ideal data set to illustrate a specific statistical issue is difficult to find. Data exercises are complex and can span several chapter sections as well as chapters.

1.5.2 Additional Book Resources Supporting Each Chapter

There will be several resources support the book that will appear outside of the chapter structure, including:

- **Case Studies and Historical Vignettes.** Similar to those appearing within the chapter, include short exercises/examples/special cases that can be readily solved by the viewer. These serve to reinforce concepts and provide benchmarks for understanding. Case studies can be used to emphasize different practices in different countries. Historical vignettes can be interesting in their own right and remind us all of the foundations of our discipline.
- **Data.** We anticipate developing a library of data sets that can be used by instructors who wish to emphasize different areas of practice.
- **Technical Supplements, Lists, and Tables.** The roles of technical supplements has already been described and there could be many. As is common in textbooks, we will also provide a place for lists or tables of organized facts for learners.

1.5.3 Software Support

We do not focus on developing `R` tutorials but will provide guides and links to people who wish to learn `R`. Our focus is on teaching statistical methods and actuarial issues, not software. Over time, the project may also provide support for users of other software environments, such as Microsoft's Excel or Python.

Chapter 2

Samples of Writing in `R bookdown`

In this chapter, you learn how to:

- Reference other sections and equations
- Include in-text citation that links to the bibliography
- Include tables and figures not generated by `R` code
- Include a footnote

As we expand our contributor and reviewer base, it will be helpful to know more about the conventions used in the series regarding the details of `R markdown` and `R bookdown` used in the series. This chapter summarizes these conventions.

2.1 Section Labels and Learning Objectives

The following shows how to code Section titles and refer to them.

```
## Section Labels {#S:SectionLabels}
```

With that reference, one can readily refer to Section 2.1 in your text, as follows:

```
With that reference, one can readily refer to
Section \\ref{S:SectionLabels} in your text, as follows:
```

The following shows how to code learning objectives:

```
***
In this chapter, you learn how to:

- Reference other sections and equations
- Include in-text citation that links to the bibliography
- Include tables and figures
- Include a footnote

***
```

2.2 Equation References

Here is an example of a latex equation produced in `R markdown`, with reference number.

$$x + y = 1 \tag{2.1}$$

You can produce that equation using the following code.

```
\begin{equation}
  x + y = 1
\label{eq:ExampleEquation}
\end{equation}
```

With this, equation (2.1) can be referred to using the following code:

```
With this, equation \\eqref{eq:ExampleEquation} can be
referred to using the following code:
```

2.3 In-text Citations

Here is an example of an in-text citation made possible by `R bookdown` (Xie, 2015). This links to the bibliography where the full referece is displayed. As a convention we use the APA style citation.

```
Here is an example of an in-text citation made possible by
`R bookdown` [@xie2015]. This links to the bibliography
where the full reference is displayed.
As a convention we use the *APA* style citation.
```

2.4 Including Tables

In order to include table not generated by `R` such as a Latex table, we have to make some adjustments to regular Latex syntax.

Policyholder	Number of claims
X	1
Y	2

Table 2.1 : An example of including tables using Latex in an `R markdown` document

`R markdown` does not have a convention for referencing non-R generated tables. For now, we reference them manually as in refer to Table 2.1. We do this by manually inserting an html anchor tag.

The following code produces this table.

```
<a id=tab:2.1></a>

[Table 2.1]: \#tab:2.1

$$
\begin{matrix}
    \begin{array}{c|c} \hline
    \text{Policyholder} & \text{Number of claims} \\\hline
```

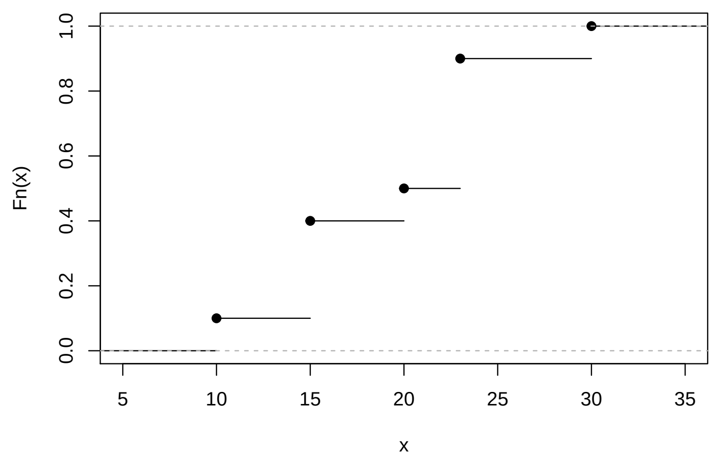

Figure 2.1: Empirical Distribution Function of a Toy Example

```
    \textbf{X} & 1 \\\hline
    \textbf{Y} & 2 \\\hline
    \end{array}
\end{matrix}
$$

[Table 2.1] : An example of including tables using Latex in an `R markdown` document
```

For reference, then use

```
`R markdown` does not have a convention for referencing
non-R generated tables. For now, we reference them manually
as in refer to [Table 2.1].
```

2.5 Including Figures

2.5.1 Figures Generated by R

Most figures are generated using R. Here is an illustrative figure.

that we refer to as Figure 2.1. Here is the code for producing the figure:

```
```{r EDFToy, echo = FALSE,
   fig.cap = 'Empirical Distribution Function of a Toy Example',
   out.width = '60%', fig.asp = 0.75, fig.align = 'center'}
xExample <- c(10,rep(15,3),20,rep(23,4),30)
PercentilesxExample <- ecdf(xExample)
plot(PercentilesxExample, main = "", xlab = "x")
```
```

Figure 2.2: An example of including figures in an R Markdown document

Here is is the code for referencing the Figure 2.1:

```
Here is is the code for referencing the Figure \\ref{fig:EDFToy}:
```

2.5.2 Figures Not Generated by R

For figures, we store the figures as png or jpeg files in a separate folder called "Figures". Then we use R code to call those figures for display so that we can reference them.

Here is such a figure:

And here is the code that generates the figure:

```
"three backticks"{r, ExampleFigure, fig.cap = 'An example
of including figures in an R Markdown document',
out.width = '5%', fig.align = 'center', echo = FALSE}
knitr::include_graphics("Figures/RStudio-Ball.png")
"three backticks"
```

Here is is the code for referencing the Figure 2.2:

```
Here is is the code for referencing the Figure \\ref{fig:ExampleFigure}:
```

2.6 Including Footnotes

Try to minimize the use of footnotes. But, if you need them, here is how you can include a footnote [1].

```
Here is how you can include a footnote [^1].

[^1]: the footnote displays at the end of the chapter
```

2.7 Useful Links

Naturally, you will want to learn more about coding in `R markdown`, `R bookdown` and so forth. The following provide some useful links for taking the next step.

- For an `R markdown` guide refer https://rmarkdown.rstudio.com/authoring_pandoc_markdown.html.
- For a `R bookdown` guide, see https://bookdown.org/yihui/bookdown/.
- For best practices in coding `R`, we suggest http://r-pkgs.had.co.nz/style.html.
- See also our online actuarial text resources at https://sites.google.com/a/wisc.edu/loss-data-analytics/online-actuarial-text-resources.

[1] the footnote displays at the end of the chapter

Chapter 3

Conventions for Notation

Chapter Preview. **Loss Data Analytics** will serve as a bridge between actuarial problems and methods and widely accepted statistical concepts and tools. Thus, the notation should be consistent with standard usage employed in probability and mathematical statistics. See, for example, (Halperin et al., 1965) for a description of one standard.

3.1 General Conventions

- Random variables are denoted by upper-case italicized Roman letters, with X or Y denoting a claim size variable, N a claim count variable, and S an aggregate loss variable. Realizations of random variables are denoted by corresponding lower-case italicized Roman letters, with x or y for claim sizes, n for a claim count, and s for an aggregate loss.
- Probability events are denoted by upper-case Roman letters, such as Pr(A) for the probability that an outcome in the event ''A" occurs.
- Cumulative probability functions are denoted by $F(z)$ and probability density functions by the associated lower-case Roman letter: $f(z)$.
- For distributions, parameters are denoted by lower-case Greek letters. A caret or ''hat" indicates a sample estimate of the corresponding population parameter. For example, $\hat{\beta}$ is an estimate of β .
- The arithmetic mean of a set of numbers, say, $x_1, \ldots, x_n$, is usually denoted by $\bar{x}$; the use of x, of course, is optional.
- Use upper-case boldface Roman letters to denote a matrix other than a vector. Use lower-case boldface Roman letters to denote a (column) vector. Use a superscript prime ''$\prime$" for transpose. For example, $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a quadratic form.
- Acronyms are to be used sparingly, given the international focus of our audience. Introduce acronyms commonly used in statistical nomenclature but limit the number of acronyms introduced. For example, pdf for probability density function is useful but GS for Gini statistic is not.

3.2 Abbreviations

Here is a list of abbreviations that we adopt. We italicize these acronyms. For example, we can discuss the goodness of fit in terms of the AIC criterion.

AIC	Akaike information criterion
BIC	(Schwarz) Bayesian information criterion
cdf	cumulative distribution function
df	degrees of freedom
iid	independent and identically distributed
glm	generalized linear model
mle	maximum likelihood estimate
ols	ordinary least squares
pdf	probability density function
pf	probability function
pmf	probability mass function
rv	random variable

3.3 Common Statistical Symbols and Operators

Here is a list of commonly used statistical symbols and operators, including the latex code that we use to generate them (in the parens).

$I(\cdot)$	binary indicator operator (I). For example, $I(A)$ is one if an outcome in event A occurs and is 0 otherwise.
$\Pr(\cdot)$	probability (`\Pr`)
$\mathrm{E}(\cdot)$	expectation operator (`\mathrm{E}`). For example, $\mathrm{E}(X) = \mathrm{E}\ X$ is the expected value of the random variable X, commonly denoted by μ.
$\mathrm{Var}(\cdot)$	variance operator (`\mathrm{Var}`).For example, $\mathrm{Var}(X) = \mathrm{Var}\ X$ is the variance of the random variable X, commonly denoted by σ^2.
$\mu_k = \mathrm{E}\ X^k$	kth moment of the random variable X. For k=1, use $\mu = \mu_1$.
$\mathrm{Cov}(\cdot,\cdot)$	covariance operator (`\mathrm{Cov}`). For example, $\mathrm{Cov}(X,Y) = \mathrm{E}\left\{(X - \mathrm{E}\ X)(Y - \mathrm{E}\ Y)\right\} = \mathrm{E}(XY) - (\mathrm{E}\ X)(\mathrm{E}\ Y)$ is the covariance between random variables X and Y.
$\mathrm{E}(X\|\cdot)$	conditional expectation operator. For example, $\mathrm{E}(X\|Y=y)$ is the conditional expected value of a random variable X given that the random variable Y equals y.
$\Phi(\cdot)$	standard normal cumulative distribution function (`\Phi`)
$\phi(\cdot)$	standard normal probability density function (`\phi`)
$\sim$	means is distributed as (`\sim`). For example, $X \sim F$ means that the random variable x has distribution function F.
$se(\hat{\beta})$	standard error of the parameter estimate $\hat{\beta}$ (`\hat{\beta}`), usually an estimate of the standard deviation of $\hat{\beta}$, which is $\sqrt{Var(\hat{\beta})}$.
H_0	null hypothesis
H_a or H_1	alternative hypothesis

3.4 Common Mathematical Symbols and Functions

Here is a list of commonly used mathematical symbols and functions, including the latex code that we use to generate them (in the parens).

$\equiv$	identity, equivalence (\equiv)
$a := b$	defines a in terms of b
$\Longrightarrow$	implies (\implies)
$\Longleftrightarrow$	if and only if (\iff)
$\to, \longrightarrow$	converges to (\to, \longrightarrow)
$\mathbb{N}$	natural numbers $1, 2, \ldots$ (\mathbb{N})
$\mathbb{R}$	real numbers (\mathbb{R})
$\in$	belongs to (\in)
$\notin$	does not belong to (\notin)
$\subseteq$	is a subset of (\subseteq)
$\subset$	is a proper subset of (\subset)
$\cup$	union (\cup)
$\cap$	intersection (\cap)
$\emptyset$	empty set (\emptyset)
A^c	complement of A
$g * f$	convolution $(g * f)(x) = \int_{-\infty}^{\infty} g(y) f(x-y) dy$
$\exp$	exponential (\exp)
$\log$	natural logarithm (\log)
$\log_a$	logarithm to the base a
!	factorial
$\text{sgn}(x)$	sign of x(sgn)
$\lfloor x \rfloor$	integer part of x, that is, largest integer $\leq x$ (\lfloor, \rfloor)
$\lvert x \rvert$	absolute value of scalar x
$\varGamma(x)$	gamma (generalized factorial) function (\varGamma), satisfying $\varGamma(x+1) = x\varGamma(x)(\varGamma)$
$B(x, y)$	beta function, $\varGamma(x)\varGamma(y)/\varGamma(x+y)$

3.5 Further Readings

To make connections to other literatures, see (Abadir and Magnus, 2002) http://www.janmagnus.nl/misc/notation.zip for a summary of notation from the econometrics perspective. This reference has a terrific feature that many latex symbols are defined in the article. Further, there is a long history of discussion and debate surrounding actuarial notation; see (Boehm et al., 1975) for one contribution.

Bibliography

Abadir, K. and Magnus, J. (2002). Notation in econometrics: a proposal for a standard. The Econometrics Journal, 5(1):76–90.

Boehm, C., Engelfriet, J., Helbig, M., Kool, A. I., Leepin, P., Neuburger, E., and Wilkie, A. (1975). Thoughts on the harmonization of some proposals for a new international actuarial notation. Blätter der DGVFM, 12(2):99–129.

Halperin, M., Hartley, H. O., and Hoel, P. G. (1965). Recommended standards for statistical symbols and notation: Copss committee on symbols and notation. The American Statistician, 19(3):12–14.

Xie, Y. (2015). Dynamic Documents with R and knitr. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.