中文书模板

李东风

2020 年 6 月 5 日

目录

简介

R 软件的 bookdown 扩展包是 R Markdown 的增强版，支持自动目录、文献索引、公式编号与引用、定理编号与引用、图表自动编号与引用等功能，可以作为 LaTeX 的一种替代解决方案，在制作用 R 进行数据分析建模的技术报告时，可以将报告文字、R 程序、文字性结果、表格、图形都自动地融合在最后形成的网页或者 PDF 文件中。

Bookdown 使用的设置比较复杂，对初学者不够友好。这里制作了一些模板，用户只要解压缩打包的文件，对某个模板进行修改填充就可以变成自己的中文图书或者论文。Bookdown 的详细用法参见https://bookdown.org/yihui/bookdown/，在李东风的《统计软件教程》也有部分介绍。

一些常用功能的示例在 `0101-usage.Rmd` 文件中，用户可以在编辑器中打开此文件参考其中的做法。

Bookdown 如果输出为网页，其中的数学公式需要 MathJax 程序库的支持，用如下数学公式测试浏览器中数学公式显示是否正常：

$$\text{定积分} = \int_a^b f(x)\,dx$$

如果显示不正常，可以在公式上右键单击，选择“Math Settings–Math Renderer”，依次使用改成“Common HTML”，“SVG”等是否可以变成正常显示。PDF 版本不存在这样的问题。

Chapter 1

中文图书 Bookdown 模板的基本用法

1.1 安装设置

使用 RStudio 软件完成编辑和转换功能。在 RStudio 中，安装 bookdown 等必要的扩展包。

本模板在安装之前是一个打包的 zip 文件，在适当位置解压（例如，在 `C:/myproj` 下），得到 `MathJax`, `Books/Cbook`, `Books/Carticle` 等子目录。本模板在 `Books/Cbook` 中。

为了利用模板制作自己的中文书，将 `Books/Cbook` 制作一个副本，改成适当的子目录名，如 `Books/Mybook`。

打开 RStudio 软件，选选单 "File - New Project - Existing Directory"，选中 `Books/Mybook` 子目录，确定。这样生成一本书对应的 R project（项目）。

为了将模板内容替换成自己的内容，可以删除文件 `0101-usage.Rmd`，然后将 `1001-chapter01.Rmd` 制作几份副本，如 `1001-chapter01.Rmd`, `2012-chapter02.Rmd`，`3012-chapter03.Rmd`。各章的次序将按照前面的数值的次序排列。将每个`.Rmd` 文件内的 `{#chapter01}`, `{#chapter02-sec01}` 修改能够反映章节内容的标签文本。所有的标签都不允许重复。参见本模板中的 `0101-usage.Rmd` 文件。

后面的 §1.3.1 和 §1.3.2 给出了将当前的书转换为网页和 PDF 的命令，复制粘贴这些命令到 RStudio 命令行可以进行转换。

1.2 编写自己的内容

1.2.1 文档结构

除了 `index.Rmd` 以外，每个`.Rmd` 文件是书的一章。每章的第一行是用一个井号（`#`）引入的章标题。节标题用两个井号开始，小节标题用三个井号开始。标题后面都有大括号内以井号开头的标签，标签仅用英文大小写字母和减号。

1.2.2 图形自动编号

用 R 代码段生成的图形，只要具有代码段标签，且提供代码段选项 `fig.cap="图形的说明文字"`，就可以对图形自动编号，并且可以用如`\@ref(fig:label)`的格式引用图形。如：

```
plot(1:10, main="程序生成的测试图形")
```

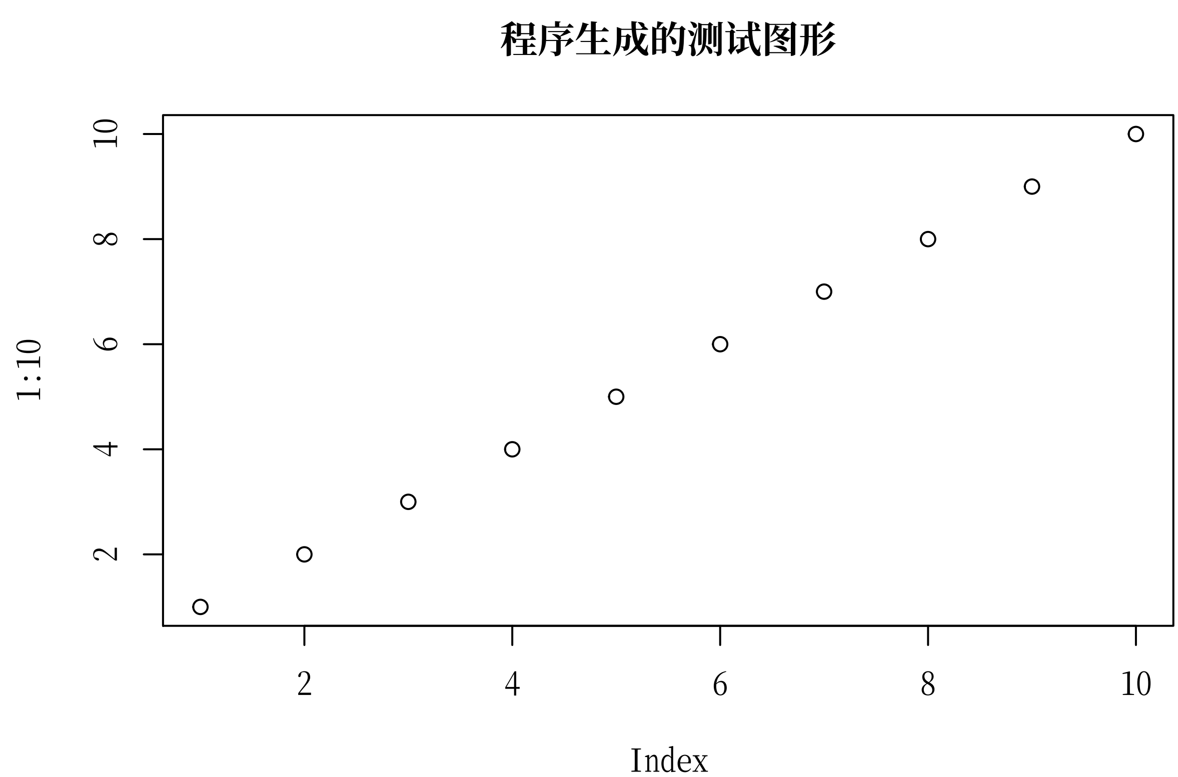

图 1.1: 图形说明文字

引用如：参见图1.1。引用中的 `fig:` 是必须的。

在通过 LaTeX 转换的 PDF 结果中，这样图形是浮动的。

1.2.3 表格自动编号

用 R 代码 `knitr::kable()` 生成的表格，只要具有代码段标签，并且在 `knitr::kable()` 调用时加选项 `caption="表格的说明文字"`，就可以对表格自动编号，并且可以用如`\@ref(tab:label)`的格式引用表格。如：

表 1.1: 表格说明文字

自变量	因变量
1	1
2	4
3	9
4	16
5	25
6	36
7	49
8	64
9	81
10	100

```
d <- data.frame(" 自变量"=1:10, " 因变量"=(1:10)^2)
knitr::kable(d, caption=" 表格说明文字")
```

引用如：参见表1.1。引用中的 `tab:` 是必须的。

在通过 LaTeX 转换的 PDF 结果中，这样的表格是浮动的。

1.2.4 数学公式编号

不需要编号的公式，仍可以按照一般的 Rmd 文件中公式的做法。需要编号的公式，直接写在`\begin{align}`和`\end{align}`之间，不需要编号的行在末尾用`\nonumber`标注。需要编号的行用 `(\#eq:mylabel)`添加自定义标签，如

$$\begin{aligned}\Sigma =&(\sigma_{ij})_{n\times n} \\ =&E[(X-\mu)(X-\mu)^T]\end{aligned} \tag{1.1}$$

引用如：协方差定义见(1.1)。

1.2.5 文献引用与文献列表

将所有文献用 bib 格式保存为一个`.bib`文献库，如模板中的样例文件 `mybib.bib`。可以用 JabRef 软件来管理这样的文献库，许多其它软件都可以输出这样格式的文件库。

为了引用某一本书，用如：参见 (Wichmann & Hill, 1982)。

被引用的文献将出现在一章末尾以及全书的末尾，对 PDF 输出则仅出现在全书末尾。

1.3 转换

1.3.1 转换为网页

用如下命令将整本书转换成一个每章为一个页面的网站，称为 gitbook 格式：

```
bookdown::render_book("index.Rmd",
  output_format="bookdown::gitbook", encoding="UTF-8")
```

为查看结果，在 _book 子目录中双击其中的 index.html 文件，就可以在网络浏览器中查看转换的结果。重新编译后应该点击"刷新"图标。

在章节和内容较多时，通常不希望每次小修改之后重新编译整本书，这时类似如下的命令可以仅编译一章，可以节省时间，缺点是导航目录会变得不准确。命令如：

```
bookdown::preview_chapter("1001-chapter01.Rmd",
  output_format="bookdown::gitbook", encoding="UTF-8")
```

单章的网页可以通过网络浏览器中的"打印"功能，选择一个打印到 PDF 的打印机，可以将单章转换成 PDF 格式。

1.3.2 生成 PDF

如果想将 R Markdown 文件借助于 LaTeX 格式转换为 PDF，需要在系统中安装一个 TeX 编译器。现在的 rmarkdown 包要求使用 tinytex 扩展包以及配套的TinyTeX 软件包，好像不再支持使用本机原有的 LaTex 编译系统，如果不安装 tinytex，编译为 PDF 格式时会出错。TinyTeX 优点是直接用 R 命令就可以安装，更新也由 R 自动进行，不需要用户干预。但是，安装时需要从国外网站下载许多文件，有因为网络不畅通而安装失败的危险。

为了安装 R 的 tinytex 扩展包和单独的 TinyTeX 编译软件，应运行：

```
install.packages('tinytex')
tinytex::install_tinytex()
```

安装过程需要从国外的服务器下载许多文件，在国内的网络环境下有可能因为网络超时而失败。如果安装成功，TinyTeX 软件包在 MS Windows 系统中一般会安装在 C:\Users\用户名\AppData\Roaming\MikTex 目录中，其中"用户名"应替换成系统当前用户名。如果需要删除 TinyTeX 软件包，只要直接删除那个子目录就可以。

为了判断 TinyTeX 是否安装成功，在 RStudio 中运行

```
tinytex::is_tinytex()
```

结果应为 TRUE, 出错或者结果为 FALSE 都说明安装不成功。在编译 `pdf_book` 时，可能会需要联网下载 LaTeX 所需的格式文件。

Bookdown 借助操作系统中安装的 LaTeX 编译软件 TinyTeX 将整本书转换成一个 PDF 文件，这需要用户对 LaTeX 有一定的了解，否则一旦出错，就完全不知道如何解决。用户如果需要进行 LaTeX 定制，可修改模板中的 `preamble.tex` 文件。

转换为 PDF 的命令如下：

```
bookdown::render_book("index.Rmd",
  output_format="bookdown::pdf_book", encoding="UTF-8")
```

在 `_book` 子目录中找到 `CBook.pdf` 文件，这是转换的结果。CBook.tex 是作为中间结果的 LaTeX 文件，如果出错可以从这里查找错误原因。

转换 PDF 对于内容多的书比较耗时,不要过于频繁地转换 PDF,在修改书的内容时,多用 `bookdown::preview_chapter` 和转换为 gitbook 的办法检验结果。定期地进行转换 PDF 的测试。每增加一章后都应该试着转换成 PDF 看有没有错误。命令如：

```
bookdown::preview_chapter("1001-chapter01.Rmd",
  output_format="bookdown::gitbook", encoding="UTF-8")
```

1.3.3 上传到网站

如果书里面没有数学公式，则上传到网站就只要将 `_book` 子目录整个地用 ftp 软件传送到自己的网站主目录下的某个子目录即可。但是，为了支持数学公式，就需要进行如下的目录结构设置：

1. 设自己的网站服务器目录为`/home/abc`，将 MathJax 目录上传到这个目录中。
2. 在`/home/abc` 中建立新目录 `Books/Mybook`。
3. 将 `_book` 子目录上传到`/home/abc/Books/Mybook` 中。
4. 这时网站链接可能类似于 `http://dept.univ.edu.cn/~abc/Books/Mybooks/_book/index.html`，具体链接地址依赖于服务器名称与主页所在的主目录名称。

如果有多本书，`MathJax` 仅需要上传一次。因为 `MathJax` 有三万多个文件，所以上传 `MathJax` 会花费很长时间。

Chapter 2

格兰格因果性

2.1 介绍

考虑两个时间序列之间的因果性。这里的因果性指的是时间顺序上的关系，如果 $X_{t-1}, X_{t-2}, \dots$ 对 Y_t 有作用，而 $Y_{t-1}, Y_{t-2}, \dots$ 对 X_t 没有作用，则称 $\{X_t\}$ 是 $\{Y_t\}$ 的格兰格原因，而 $\{Y_t\}$ 不是 $\{X_t\}$ 的格兰格原因。如果 $X_{t-1}, X_{t-2}, \dots$ 对 Y_t 有作用，$Y_{t-1}, Y_{t-2}, \dots$ 对 X_t 也有作用，则在没有进一步信息的情况下无法确定两个时间序列的因果性关系。

注意这种因果性与采样频率有关系，在日数据或者月度数据中能发现的领先——滞后性质的因果关系，到年度数据可能就以及混杂在以前变成同步的关系了。

2.2 格兰格因果性的定义

设 $\{\xi_t\}$ 为一个时间序列，$\{\eta_t\}$ 为向量时间序列，记

$$\bar{\eta}_t = \{\eta_{t-1}, \eta_{t-2}, \dots\}$$

记 $\text{Pred}(\xi_t|\bar{\eta}_t)$ 为基于 $\eta_{t-1}, \eta_{t-2}, \dots$ 对 ξ_t 作的最小均方误差无偏预报，其解为条件数学期望 $E(\xi_t|\eta_{t-1}, \eta_{t-2}, \dots)$，在一定条件下可以等于 ξ_t 在 $\eta_{t-1}, \eta_{t-2}, \dots$ 张成的线性 Hilbert 空间的投影（比如，(ξ_t, η_t) 为平稳正态多元时间序列），即最优线性预测。直观理解成基于过去的 $\{\eta_{t-1}, \eta_{t-2}, \dots\}$ 的信息对当前的 ξ_t 作的最优预测。

令一步预测误差为

$$\varepsilon(\xi_t|\bar{\eta}_t) = \xi_t - \text{Pred}(\xi_t|\bar{\eta}_t)$$

令一步预测误差方差，或者均方误差，为

$$\sigma^2(\xi_t|\bar{\eta}_t) = \text{Var}(\varepsilon_t(\xi_t|\bar{\eta}_t)) = E\left[\xi_t - \text{Pred}(\xi_t|\bar{\eta}_t)\right]^2$$

考虑两个时间序列 $\{X_t\}$ 和 $\{Y_t\}$，$\{(X_t, Y_t)\}$ 宽平稳或严平稳。

- 如果
$$\sigma^2(Y_t|\bar{Y}_t, \bar{X}_t) < \sigma^2(Y_t|\bar{Y}_t)$$
则称 $\{X_t\}$ 是 $\{Y_t\}$ 的**格兰格原因**，记作 $X_t \Rightarrow Y_t$。这不排除 $\{Y_t\}$ 也可以是 $\{X_t\}$ 的格兰格原因。
- 如果 $X_t \Rightarrow Y_t$，而且 $Y_t \Rightarrow X_t$，则称互相有**反馈**关系，记作 $X_t \Leftrightarrow Y_t$。
- 如果
$$\sigma^2(Y_t|\bar{Y}_t, X_t, \bar{X}_t) < \sigma^2(Y_t|\bar{Y}_t, \bar{X}_t)$$
即除了过去的信息，增加同时刻的 X_t 信息后对 Y_t 预测有改进，则称 $\{X_t\}$ 对 $\{Y_t\}$ 有瞬时因果性。这时 $\{Y_t\}$ 对 $\{X_t\}$ 也有瞬时因果性。
- 如果 $X_t \Rightarrow Y_t$，则存在最小的正整数 m，使得
$$\sigma^2(Y_t|\bar{Y}_t, X_{t-m}, X_{t-m-1}, \dots) < \sigma^2(Y_t|\bar{Y}_t, X_{t-m-1}, X_{t-m-2}, \dots)$$
称 m 为**因果性滞后值** (causality lag)。如果 $m > 1$，这意味着在已有 $Y_{t-1}, Y_{t-2}, \dots$ 和 $X_{t-m}, X_{t-m-1}, \dots$ 的条件下，增加 X_{t-1}, ..., X_{t-m+1} 不能改进对 Y_t 的预测。

例 2.1. 设 $\{\varepsilon_t, \eta_t\}$ 是相互独立的零均值白噪声列，$\text{Var}(\varepsilon_t) = 1$, $\text{Var}(\eta_t) = 1$, 考虑
$$\begin{aligned} Y_t =& X_{t-1} + \varepsilon_t \\ X_t =& \eta_t + 0.5\eta_{t-1} \end{aligned}$$

用 $L(\cdot|\cdot)$ 表示最优线性预测，则
$$\begin{aligned} & L(Y_t|\bar{Y}_t, \bar{X}_t) \\ =& L(X_{t-1}|X_{t-1}, \dots, Y_{t-1}, \dots) + L(\varepsilon_t|\bar{Y}_t, \bar{X}_t) \\ =& X_{t-1} + 0 \\ =& X_{t-1} \\ \sigma(Y_t|\bar{Y}_t, \bar{X}_t) =& \text{Var}(\varepsilon_t) = 1 \end{aligned}$$

而
$$Y_t = \eta_{t-1} + 0.5\eta_{t-2} + \varepsilon_t$$

有
$$\gamma_Y(0) = 2.25, \gamma_Y(1) = 0.5, \gamma_Y(k) = 0, k \geq 2$$

所以 $\{Y_t\}$ 是一个 MA(1) 序列，设其方程为
$$Y_t = \zeta_t + b\zeta_{t-1}, \zeta_t \sim \text{WN}(0, \sigma_\zeta^2)$$

可以解出
$$\begin{aligned} \rho_Y(1) =& \frac{\gamma_Y(1)}{\gamma_Y(0)} = \frac{2}{9} \\ b =& \frac{1 - \sqrt{1 - 4\rho_Y^2(1)}}{2\rho_Y(1)} \approx 0.2344 \\ \sigma_\zeta^2 =& \frac{\gamma_Y(1)}{b} \approx 2.1328 \end{aligned}$$

于是

$$\sigma(Y_t|\bar{Y}_t) = \sigma_\zeta^2 = 2.1328$$

所以

$$\sigma(Y_t|\bar{Y}_t, \bar{X}_t) = 1 < 2.1328 = \sigma(Y_t|\bar{Y}_t)$$

即 X_t 是 Y_t 的格兰格原因。

反之，X_t 是 MA(1) 序列，有

$$\eta_t = \frac{1}{1+0.5B} X_t = \sum_{j=0}^{\infty} (-0.5)^j X_{t-j}$$

其中 B 是推移算子（滞后算子）。于是

$$\begin{aligned}
L(X_t|\bar{X}_t) =& L(\eta_t|\bar{X}_t) + 0.5L(\eta_{t-1}|\bar{X}_t) \\
=& 0.5 \sum_{j=0}^{\infty} (-0.5)^j X_{t-1-j} \\
=& -\sum_{j=1}^{\infty} (-0.5)^j X_{t-j} \\
\sigma(X_t|\bar{X}_t) =& \mathrm{Var}(X_t - L(X_t|\bar{X}_t)) \\
=& \mathrm{Var}(\eta_t) = 1
\end{aligned}$$

而

$$\begin{aligned}
L(X_t|\bar{X}_t, \bar{Y}_t) =& L(\eta_t|\bar{X}_t, \bar{Y}_t) + 0.5L(\eta_{t-1}|\bar{X}_t, \bar{Y}_t) \\
=& 0 + 0.5L(\sum_{j=0}^{\infty} (-0.5)^j X_{t-1-j}|\bar{X}_t, \bar{Y}_t) \\
=& -\sum_{j=1}^{\infty} (-0.5)^j X_{t-j} \\
=& L(X_t|\bar{X}_t)
\end{aligned}$$

所以 Y_t 不是 X_t 的格兰格原因。

考虑瞬时因果性。

$$\begin{aligned}
L(Y_t|\bar{X}_t, \bar{Y}_t, X_t) =& X_{t-1} + 0(\text{注意}\varepsilon_t\text{与}\{X_s, \forall s\}\text{不相关} \\
=& L(Y_t|\bar{X}_t, \bar{Y}_t)
\end{aligned}$$

所以 X_t 不是 Y_t 的瞬时格兰格原因。

例 2.2. 在例2.1中，如果模型改成

$$\begin{aligned}
Y_t =& X_t + \varepsilon_t \\
X_t =& \eta_t + 0.5\eta_{t-1}
\end{aligned}$$

有怎样的结果？

这时

$$Y_t = \varepsilon_t + \eta_t + 0.5\eta_{t-1}$$

仍有

$$\gamma_Y(0) = 2.25, \gamma_Y(1) = 0.5, \gamma_Y(k) = 0, k \geq 2$$

所以 Y_t 还服从 MA(1) 模型

$$Y_t = \zeta_t + b\zeta_{t-1}, b \approx 0.2344, \sigma_\zeta^2 \approx 2.1328$$

$$\begin{aligned}
L(Y_t|\bar{Y}_t, \bar{X}_t) =& L(X_t|\bar{Y}_t, \bar{X}_t) + 0 \\
=& L(\eta_t|\bar{Y}_t, \bar{X}_t) + 0.5L(\eta_{t-1}|\bar{Y}_t, \bar{X}_t) \\
=& 0 + 0.5L(\sum_{j=0}^{\infty}(-0.5)^j X_{t-1-j}|\bar{Y}_t, \bar{X}_t) \\
=& -\sum_{j=1}^{\infty}(-0.5)^j X_{t-j} \\
=& X_t - \eta_t \\
\sigma(Y_t|\bar{Y}_t, \bar{X}_t) =& \text{Var}(\varepsilon_t + \eta_t) = 2
\end{aligned}$$

而

$$\sigma(Y_t|\bar{Y}_t) = \sigma_\zeta^2 \approx 2.1328 > \sigma(Y_t|\bar{Y}_t, \bar{X}_t) = 2$$

所以 X_t 是 Y_t 的格兰格原因。

反之，

$$\begin{aligned}
L(X_t|\bar{X}_t, \bar{Y}_t) =& -\sum_{j=1}^{\infty}(-0.5)^j X_{t-j} \\
=& L(X_t|\bar{X}_t)
\end{aligned}$$

所以 Y_t 不是 X_t 的格兰格原因。

考虑瞬时因果性。

$$\begin{aligned}
L(Y_t|\bar{X}_t, \bar{Y}_t, X_t) =& X_t + 0(\text{注意}\varepsilon_t\text{与}\{X_s, \forall s\}\text{不相关} \\
=& X_t \\
\sigma(Y_t|\bar{X}_t, \bar{Y}_t, X_t) =& \text{Var}(\varepsilon) \\
=& 1 < 2 = \sigma(Y_t|\bar{X}_t, \bar{Y}_t)
\end{aligned}$$

所以 X_t 是 Y_t 的瞬时格兰格原因。

[aaa]

Bibliography

Wichmann, B. A., & Hill, I. D. (1982). Algorithm as 183: An efficient and portable pseudo-random number generator. *Applied Statistics*, *31*, 188–190. Remarks: 34, 198 and 35, 89.