

A Bayesian hierarchical model of categorical data rating and classification

Bob Carpenter

Center for Computational Mathematics, Flatiron Institute

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur posuere vestibulum facilisis. Aenean pretium orci augue, quis lobortis libero accumsan eu. Nam mollis lorem sit amet pellentesque ullamcorper. Curabitur lobortis libero eget malesuada vestibulum. Nam nec nibh massa. Pellentesque porttitor cursus tellus. Mauris urna erat, rhoncus sed faucibus sit amet, venenatis eu ipsum.

Keywords: crowdsourcing, Bayesian modeling, hierarchical priors, categorical data

Contents

1	Introduction	2
1.1	About this document	2
1.2	Advice for writing your manuscript	2
2	Formatting	2
2.1	Basic markdown formatting	2
2.2	Mathematics	2
2.2.1	Mathematical formulae	2
2.2.2	Theorems and other amsthm-like environments	3
2.3	Code	3
2.3.1	R	4
2.3.2	Python	4
2.4	Figures	4
2.5	Tables	5
2.6	Handling references	6
2.6.1	Bibliographic references	6
2.6.2	Other cross-references	6
	Session information	6

1 Introduction

1.1 About this document

This document provides a template based on the [quarto system](#) for contributions to **Computo** (Computo Team (2021)). We show how Python (Perez, Granger, and Hunter 2011) or R (R Core Team 2020) code can be included.

1.2 Advice for writing your manuscript

First make sure that you are able to build your manuscript as a regular notebook on your system. Then you can start configure the binder environment.

2 Formatting

This section covers basic formatting guidelines. [Quarto](#) is a versatile formatting system for authoring HTML based on markdown, integrating LaTeX and various code block interpreted either via Jupyter or Knitr (and thus deal with Python, R and many other languages). It relies on the [Pandoc Markdown](#) markup language.

To render/compile a document, run `quarto render`. A document will be generated that includes both content as well as the output of any embedded code chunks within the document:

```
quarto render content.qmd # will render to html
```

2.1 Basic markdown formatting

Bold text or *italic*

- This is a list
- With more elements
- It isn't numbered.

But we can also do a numbered list

1. This is my first item
2. This is my second item
3. This is my third item

2.2 Mathematics

2.2.1 Mathematical formulae

[LaTeX](#) code is natively supported¹, which makes it possible to use mathematical formulae:

will render

¹We use [katex](#) for this purpose.

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

It is also possible to cross-reference an equation, see Equation 1:

$$\begin{aligned} D_{x_N} &= \frac{1}{2} \begin{bmatrix} x_L^\top & x_N^\top \end{bmatrix} \begin{bmatrix} L_L & B \\ B^\top & L_N \end{bmatrix} \begin{bmatrix} x_L \\ x_N \end{bmatrix} \\ &= \frac{1}{2} (x_L^\top L_L x_L + 2x_N^\top B^\top x_L + x_N^\top L_N x_N), \end{aligned} \tag{1}$$

2.2.2 Theorems and other amsthm-like environments

Quarto includes a nice support for theorems, with predefined prefix labels for theorems, lemmas, proposition, etc. see [this page](#). Here is a simple example:

Theorem 2.1 (Strong law of large numbers). *The sample average converges almost surely to the expected value:*

$$\bar{X}_n \xrightarrow{a.s.} \mu \quad \text{when } n \rightarrow \infty.$$

See Theorem 2.1.

2.3 Code

Quarto uses either Jupyter or knitr to render code chunks. This can be triggered in the yaml header, e.g., for Jupyter (should be installed on your computer) use

```
---
title: "My Document"
author "Jane Doe"
jupyter: python3
---
```

For knitr (R + knitr must be installed on your computer)

```
---
title: "My Document"
author "Jane Doe"
---
```

You can use Jupyter for Python code and more. And R + Knitr for if you want to mix R with Python (via the package reticulate Ushey, Allaire, and Tang (2020)).

2.3.1 R

R code (R Core Team 2020) chunks may be embedded as follows:

```
x <- rnorm(10)
```

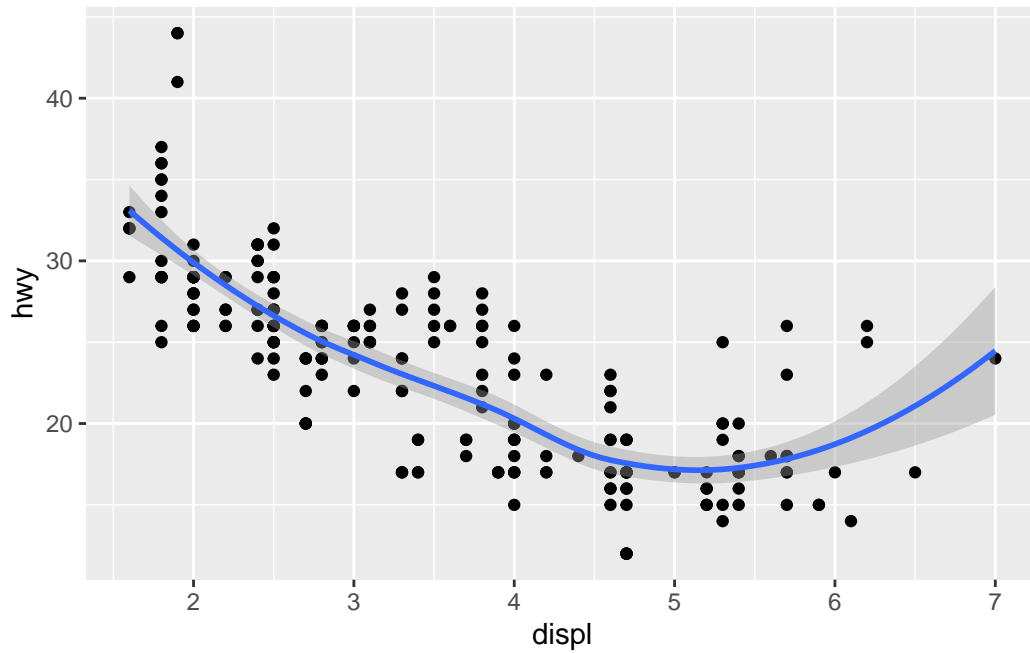
2.3.2 Python

```
---  
title: "My Document"  
author "Jane Doe"  
jupyter: python3  
---  
  
import matplotlib.pyplot as plt  
import numpy as np  
  
fig, ax = plt.subplots()  
ax.plot(np.arange(10))
```

2.4 Figures

Plots can be generated as follows:

```
library("ggplot2")  
p <- ggplot(mpg, aes(displ, hwy)) +  
  geom_point() +  
  geom_smooth()  
p
```



It is also possible to create figures from static images:



Figure 1: SFdS logo (c.a. 2021)

2.5 Tables

Tables (with label: @tbl-mylabel renders Table 1) can be generated with markdown as follows

Table 1: my table caption

Tables	Are	Cool
col 1 is	left-aligned	\$1600
col 2 is	centered	\$12
col 3 is	right-aligned	\$1

Table can also be generated by some code, for instance with knitr here:

```
knitr::kable(summary(cars), caption = "Table caption.")
```

Table 2: Table caption.

speed	dist
Min. : 4.0	Min. : 2.00
1st Qu.:12.0	1st Qu.: 26.00
Median :15.0	Median : 36.00
Mean :15.4	Mean : 42.98
3rd Qu.:19.0	3rd Qu.: 56.00
Max. :25.0	Max. :120.00


2.6 Handling references

2.6.1 Bibliographic references

References are displayed as footnotes using [BibTeX](#), e.g. `[@computo]` will be displayed as (Computo Team 2021), where `computo` is the bibtex key for this specific entry. The bibliographic information is automatically retrieved from the `.bib` file specified in the header of this document (here: `references.bib`).

2.6.2 Other cross-references

As already (partially) seen, Quarto includes a mechanism similar to the bibliographic references for sections, equations, theorems, figures, lists, etc. Have a look at [this page](#).

 For more information

[Check our mock version of the t-SNE paper](#) for a full and advanced example using the Jupyter kernel.

[The template available in the Computo Quarto extension](#) uses advanced features and the Knitr kernel (interactive plots and pseudocode).

Session information

```
sessionInfo()
```

```
R version 4.2.2 (2022-10-31)
```

```
Platform: x86_64-conda-linux-gnu (64-bit)
```

```
Running under: Ubuntu 22.04.1 LTS
```

```
Matrix products: default
```

```
BLAS/LAPACK: /home/runner/micromamba-root/envs/computorbuild/lib/libopenblaspr0.3.21.so
```

```

locale:
  [1] LC_CTYPE=C.UTF-8      LC_NUMERIC=C          LC_TIME=C.UTF-8
  [4] LC_COLLATE=C.UTF-8    LC_MONETARY=C.UTF-8  LC_MESSAGES=C.UTF-8
  [7] LC_PAPER=C.UTF-8      LC_NAME=C            LC_ADDRESS=C
  [10] LC_TELEPHONE=C        LC_MEASUREMENT=C.UTF-8 LC_IDENTIFICATION=C

```

```

attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods    base

```

```

other attached packages:
[1] ggplot2_3.4.0

```

```

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.9      highr_0.10      compiler_4.2.2  pillar_1.8.1
 [5] tools_4.2.2     digest_0.6.31   nlme_3.1-161    jsonlite_1.8.4
 [9] evaluate_0.19   lifecycle_1.0.3 tibble_3.1.8    gtable_0.3.1
[13] lattice_0.20-45 mgcv_1.8-41     pkgconfig_2.0.3 png_0.1-8
[17] rlang_1.0.6     Matrix_1.5-3    DBI_1.1.3       cli_3.5.0
[21] yaml_2.3.6      xfun_0.36       fastmap_1.1.0   withr_2.5.0
[25] stringr_1.5.0   dplyr_1.0.10    knitr_1.41      generics_0.1.3
[29] vctrs_0.5.1     tidyselect_1.2.0 grid_4.2.2      reticulate_1.26
[33] glue_1.6.2      R6_2.5.1        fansi_1.0.3     rmarkdown_2.19
[37] farver_2.1.1    magrittr_2.0.3  splines_4.2.2   scales_1.2.1
[41] htmltools_0.5.4 assertthat_0.2.1 colorspace_2.0-3 labeling_0.4.2
[45] utf8_1.2.2      stringi_1.7.8   munsell_0.5.0

```

- Computo Team. 2021. “Computo: Reproducible Computational/Algorithmic Contributions in Statistics and Machine Learning.” *Computo*.
- Perez, Fernando, Brian E Granger, and John D Hunter. 2011. “Python: An Ecosystem for Scientific Computing.” *Computing in Science & Engineering* 13 (2): 13–21.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ushey, Kevin, JJ Allaire, and Yuan Tang. 2020. *Reticulate: Interface to Python*. <https://github.com/rstudio/reticulate>.