Bayesian Data Analysis

Anthony R. Colombo, Dr. Paul Marjoram

2022-05-27

Contents

About

This is an independent study of the text **Bayesian Data Analysis** by Andrew Gelman and the more introductory text **A First Course In Bayesian Statistical Methods** by Peter D. Hoff. We will read through most of the chapters and typeset the major definitions. The Gelman text can be difficult, and for difficult chapters, we will lean more on the Hoff textbook.

Independent study

We will typeset each chapter of the Gelman text book, unless the chapter is too difficult, then we will use the introductory text **A First Course In Bayesian Statistical Methods** by Hoff. Each chapter will summarize the definitions, and attempt several problems selected.

Supervised learning

Dr. Paul Marjoram will supervise the learning and have a general oversight to the learning process.

Chapter 1

Fundamentals of Bayesian Inference

The first few chapters of Gelman's text are introductory, and we attempt to highlight the key definitions and summarize each chapter. At the end of each chapter we attempt several problems. Probability and inference is defined using three steps

1. setting up the full probability model for a joint distribution for all observable and unobservable quantities.
2. Conditioning on observed data: computing the appropriate *posterior* distribution, the conditional probability distribution of the unobserved quantities of oltimate interest, given the observed data.
3. Evaluating the fit of the model.

1.1   General notation for statistical inference

There are two different kinds of estimands, the first are potentially observable quantities, such as future observations of a process, and the second are quantities that are not directly observable, namely the parameters that govern a process being investigated.

Exchangeability

One key assumption is that the n values y_i are regarded as *exchangeable*, meaning that the uncertainty can be expressed as a joint probability $p(y_1, ..., y_n)$ that is invariant to permutations of indexes. Often times the exchangeable distribution is modeled as *iid*.

Explanatory variables

It is common to have observations on each unit which have non-random variables called *explanatory variables* or *covariates*. The explanatory variables are usually denoted by X. However treating X as random then exchangeability can be extended $(x, y)_i$ which is invariant to permutations of the indexes. Further, it is always appropriate to assume exchangeability of y, conditioned on sufficient information of X, where the indexes can be thought of as randomly assigned. It follows that if two units have the same value of x, then the distributions of y are the same.

Hierarchical modeling

for a model across patients across different cities, we can assume exchangability to patients within a city. Further conditioned on the explanatory variables at the individual, the conditional distribution given these explanatory variables would be exchangeable.

1.2   Bayesian inference

The prior, p(θ), and the sampling distribution, or the *data distribution*, $p(y|\theta)$ is

$$p(\theta, y) = p(\theta)p(y|\theta)$$

Where using Bayes' rule the posterior distribution

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{1.1}$$

Where $p(y) = \int p(\theta)p(y|\theta)d\theta$, or a sum in discrete case. An equivalent form of (1.1) is the *unnormalized posterior density* given as

$$p(\theta|y) \propto p(\theta)p(y|\theta) \tag{1.2}$$

Note that $p(y|\theta)$ is taken as a function of θ, not of y.

Prediction

Inferences about an unknown *observable* variable, are called predictive inferences. Beofre the data y are considered, the distribution of the unknown, observable, y is

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta$$

this is defined as the marginal distribution of y, and also called *prior predictive distribution.* Prior refers that the data is not conditional on any previous observation, and predictive refers to the data being observable.

The *posterior predictive distribution* is conditional on the observed y, but is predictive because it is predicting observable values.

$$
\begin{aligned}
p(\hat{y}|y) &= \int p(\hat{y}, \theta|y)d\theta \\
&= \int p(\hat{y}|\theta, y)p(\theta|y)d\theta \\
&= \int p(\hat{y}|\theta)p(\theta|y)d\theta
\end{aligned}
\tag{1.3}
$$

Likelihood

The data y affects the posterior inference only through (1.2) likelihood function $p(y|\theta)$ which is regarded as a function of θ for fixed y. The *likelihood function* is defined as $p(y|\theta)$, and the *likelihood principle* is for any given sample, and any two likelihood models $p(y|\theta)$, two models with the same likelihood will have the same inference for θ.

Subjectivity and Objectivity

The frequentist models, MLEs, have subjectivity in their assumptions because they rely on long sequence of identical trials, that are iid. The Bayesian model relies on the prior distribution. If any experiment is repeatable and can be replicated, the the prior distribution can be estimated from the data themselves and the analysis is more 'objective'. Replication increases objectivity of a given model. However the Bayesian approach allows for (1) the ability to combine information from multiple sources (allowing for greater objectivity) and (2) more encompassing by accounting for uncertainity about the unknowns in a statistical problem.

It is important to include as much background information as possible

1.3   Exercises

1. (a) for $\sigma = 2$; $p(y) = \int p(y|\theta)p(\theta)d\theta = (1/2)N(1, \sigma^2) + (1/2)N(2, \sigma^2)$ as
    the marginal density

```r
fy<-function(y) {return(0.5*dnorm(y,mean=1,sd=2)+0.5*dnorm(y,2,sd=2))}
```

(b) $P(\theta = 1|y = 1) = \frac{p(\theta=1)p(y|\theta=1)}{p(y)} = \frac{(1/2)N(1,4)}{(1/2)N(1,\sigma^2)+(1/2)N(2,\sigma^2)} = 0.53$

```r
dy<-function(y){ return( (1/2)*dnorm(y,mean=1,sd=2)/fy(y))}
dy(1)
```

```
## [1] 0.5312094
```

Chapter 2

Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

2.1 Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
   - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world` = `# Hello world {#hello-world}`.
   - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.

2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter 2.
   - If you prefer text as the link instead of a numbered reference use: any text you want can go here.

2.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 2.1.

```r
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```


Figure 2.1: Here is a nice figure!

Don't miss Table 2.1.

```r
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

Table 2.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Chapter 3

Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 4

Footnotes and citations

4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2022] (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

---

[1] This is a footnote.

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 5.1.

Theorem 5.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

5.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

Chapter 6

Sharing your book

6.1 Publishing

HTML books can be published online, see: https://bookdown.org/yihui/bookdown/publishing.html

6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book-all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/gitbook.html

Or use:

```
?bookdown::gitbook
```

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL http://yihui.org/knitr/. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2022. URL https://CRAN.R-project.org/package=bookdown. R package version 0.26.