Things that should have been done better in "A Student's Guide to Bayesian Statistics"

Ben Lambert

March 6, 2019

0.1 Acknowledgements

Rather than "...as close to Bayes as my Mum was as a child..." it should say "...as close to Bayes as my mum was as a child...".

Chapter 1

How to best use this book

Chapter 2

The subjective worlds of Frequentist and Bayesian statistics

Chapter 3

Probability - the nuts and bolts of Bayesian inference

Various distribution's properties. These are only formatting issues not mistakes in the formulae. These should all be consistent. For example, for the uniform distribution, the mean and variance should be stated taking θ as an argument.

Problem 3.7.4 As suggested by Sean Jackson, rephrase the question as "If the officials want to maximize the proportion of people correctly identified as dopers (ie keep down the proportion of false positives), should they test all the athletes or only the winners?"

Problem 3.7.5-6 Note that both answers to these problems were previously incorrect (which was pointed out by an email from Martin Oldfield and Sean Jackson). The problem set answers uploaded on my website are now corrected.

Chapter 4

Likelihoods

Chapter 5

Priors

Chapter 6

The devil is in the denominator

Chapter 7

The posterior - the goal of Bayesian inference

Eqn. (7.3) should read,

$$var(\theta|data) = \int_0^1 p(\theta|data)(\theta - \mathbb{E}[\theta|data])^2 \mathrm{d}\theta. \tag{7.1}$$

Problem 7.2.2 is missing a question. After the eqn. (7.8), there should be the following, "Fit this model to the data using a Frequentist approach. Which model do you prefer, and why?"

Problem 7.2.3 could be clarified by changing the text to, "Construct 80% confidence intervals for (α, β) for the log-log model."

Chapter 8

An introduction to distributions for the mathematically un-inclined

Table 8.1 has a few errors in it:

- LKJ: The mean is the identity matrix; the variance between off-diagonal rows is $var(X_{i,j}) = \frac{4\left(\frac{d}{2}+\nu-1\right)^2}{(d+2\nu-2)^2(d+2\nu-1)}$.
- Poisson: Under 'Prior example uses', it currently says 'Negative Binomial' whereas this should be under 'Preferred distribution.

Chapter 9

Conjugate priors and their place in Bayesian analysis

Chapter 10

Evaluation of model fit and hypothesis testing

Problem 10.4.1 Clarify as "Find the posterior distribution where we specify $\theta \sim beta(a, 1)$ as a prior. Graph how the posterior changes as $a \in [1, 10]$ is varied."

Chapter 11

Making Bayesian analysis objective?

Problem 11.2.1 could have been stated more clearly. Here's a reword of the question, "Suppose that we start by assigning a uniform prior on θ. Using an appropriate probability distribution, calculate the probability that in a sample of two individuals, one person has the disease, and the other does not. Hence find the prior belief that is ascribed to this outcome and comment on the assumption that a uniform prior is uninformative."

Chapter 12

Leaving conjugates behind: Markov Chain Monte Carlo

Chapter goals Spotted by Martin Blais, there is a spelling mistake here and the sentence should read "In most cases, it is difficult to determine what constitutes an adequate sample since..."

Table 12.1 has a few minor rounding errors. The corrected version is shown in Table 12.1.

Problem 12.2 This is minor, but can affect the running of functions in R. The density should be,

$$f(x) = \begin{cases} \frac{1}{1.33485} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, & \text{if } x < 0.9735. \\ 0.186056, & \text{if } 0.9735 \leq x \leq 5. \\ 0, & \text{otherwise.} \end{cases} \tag{12.1}$$

In the text, 1.335 is used which means the PDF doesn't quite normalise.

Problem 12.3.3 There was a mistake in the question in the first edition of the book with $-4x$ opposed to the intended $+4x$.

Figure 12.9 Spotted by Martin Blais, there is a mistake in the leftmost and middle figures, which are missing some edges. The left figure is missing a link between 1-3 and 1-5 and the middle figure is missing links between 4-6 and 2-4.

λ (mean number of BSE cases)	Prior	Likelihood	Prior $\times$ Likelihood	Posterior	$\lambda\times$ Posterior
0.0	0.000	0.000	0.000	0.000	0.000
2.0	0.468	0.003	0.002	0.035	0.069
4.0	0.228	0.060	0.014	0.291	1.160
6.0	0.120	0.138	0.016	0.353	2.120
8.0	0.069	0.140	0.010	0.205	1.640
10.0	0.042	0.090	0.004	0.081	0.813
12.0	0.027	0.044	0.001	0.025	0.306
14.0	0.018	0.017	0.000	0.007	0.096
16.0	0.013	0.006	0.000	0.002	0.026
18.0	0.009	0.002	0.000	0.000	0.007
20.0	0.007	0.001	0.000	0.000	0.002
	1.000	0.500	$Pr(X = 7) = 0.047$	1.000	$\mathbb{E}(\lambda\|X = 7) = 6.240$

Table 12.1:

Chapter 13

Random Walk Metropolis

13.6.3 In the third paragraph, the sentence, "The Metropolis algorithm is characterised by a symmetric proposal distribution, meaning that we are as likely to propose θ_b from θ_a as we are to go in the reverse direction, to θ_a from θ_b." is misleading because "go" implies that we will propose *then* accept. Rephrase this as, "The Metropolis algorithm is characterised by a symmetric proposal distribution, meaning that we are as likely to propose θ_b from θ_a as we are to propose a move in the reverse direction, to θ_a from θ_b." This poorly phrased sentence was spotted by Martin Blais.

13.6.4 Change "Considering the + case..." to "Considering the case where $r = \frac{p(\theta_{t+1}|data)}{p(\theta_t|data)} + \epsilon$" to avoid ambiguity. Thanks to Martin Blais for pointing this out.

Eq. (13.19) In the second summation on the right hand side, the lower index should be $\tau = 2$. Thanks to Martin Blais here.

Problem 13.1.8 I regret using, "Why can't you use vanilla Random Walk Metropolis for (α, β) here?" This is because, of course, you can use RWM here but, for optimal sampling efficiency, we may want to use Metropolis-Hastings.

Problem 13.2.8 The vector of observations should be using curly braces here for consistency.

Chapter 14

Gibbs sampling

Problem 14.1.3 There is an error in the question here, where I got the positives and negatives the wrong way round. It should read, "Suppose that out of a sample of 100 people, 20 of those tested **positive** and 80 **negative**. Assuming uniform priors on π, S and C, use Gibbs sampling to generate posterior samples for π. What do you conclude?"

Figure 14.6 It's hard to see the path taken by the sampler in both cases. Black line needs to be more prominent.

Chapter summary The Charles Geyer quote is taken from chapter 1 within the Handbook of Markov chain Monte Carlo (by Geyer himself). I should add this detail to the bibliography.

Problem 14.1.4 "Suppose that a previous study that compares the clinical test with a laboratory **iron** standard concludes that..." should be "Suppose that a previous study that compares the clinical test with a laboratory **gold** standard concludes that..."

Chapter 15

Hamiltonian Monte Carlo

Figure 15.7 doesn't have a mistake but in haste to rush to the printers, I think I used a suboptimal value for the number of steps for the HMC panel. See Figure 15.1 for a version which more adequately indicates the behaviour of HMC.

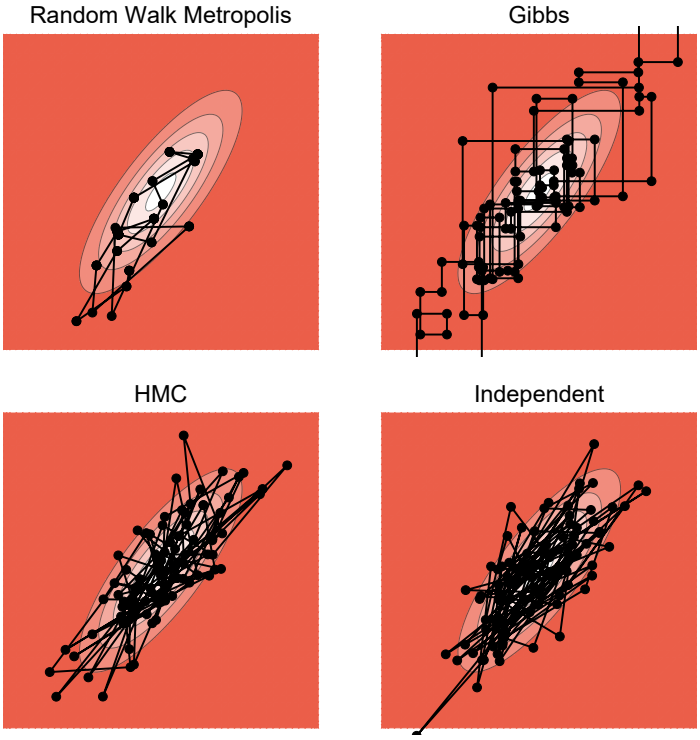

Figure 15.1:

Chapter 16

Stan

16.3.1 There is a spelling mistake here "when when". Good spot by Martin Blais.

16.5.2 The sentence after 8. is slightly misleading and should instead read "It then runs the MCMC using 200 iterations across each of four chains."

16.7.2 There is an odd typeface error with "data" in the first paragraph of page 410. Thanks to Martin Blais.

Chapter 17

Hierarchical models

Chapter 18

Linear regression models

18.2 The sentence "For example, to determine how the weight of an individual is affected by their height" should read "For example, to determine how the height of an individual is affected by their weight." Thanks to Martin Blais for pointing out this inconsistency.

Missing problem between 18.1.3 and 18.1.4. A problem was lost in the move from my Latex document to the main text. Fortunately, nothing depends on it but, nonetheless, I would like it included in the next edition. It should read:

Another model allows there to be time trends in the data,

$$murder_{i,t} \sim \mathcal{N}\left(\alpha_i + \delta_i t + \beta_i penalty_{i,t} + \gamma_i car_{i,t}, \sigma_i\right), \tag{18.1}$$

where $\delta_i \sim \mathcal{N}(\bar{\delta}, \sigma_\delta)$. Again estimate this model and compare the effect size of the death penalty across the three models.

Chapter 19

Generalised linear models and other animals

Problem 19.2 `GLM_metaAnalysis.csv` should be lowercase.

Problem 19.4 `Eurovision.csv` should be `glm_eurovision.csv`.

Problem 19.5 `terrorism.csv` should be `glm_terrorism_pairwise.csv`.

Chapter 20

General issues

The following aren't specific to any particular chapters but I make note of these so that they can be tidied up next time round,

- Consistent use of capitalised Python (there are some upper and lower case versions) and possibly Matlab too.
- Consistent code style using Google style.