Bayesian Case Studies
Final examination

Robin Ryder

January 21, 2020

All documents are allowed. Any attempt to use the Internet, including any form of e-mail or messenging, will result in immediate disqualification.
No phones allowed.
Duration: 3 hours.
Students may answer in English or French. All code must be written in the R language.
At the end of the examination, you must hand in your answers written on paper AND send your R code to `ryder@ceremade.dauphine.fr`*.*
Please contact the examiner if you wish to hand in your answers early. Please make sure that your R code has been correctly received before leaving the room.
Make sure to save your code on a regular basis. Loss of data following computer failure shall not entitle you to extra time.
Sections 2, 3 and 4 are independent and can be treated in any order.

The online encyclopedia Wikipedia contains over 5 million English articles that anyone can edit. Wikipedia articles are ocasionally linked to from the main page. We wish to study how this impacts the edit rate. To this aim, we shall study the amount of time that has elapsed since the last edit to articles mentioned on the English Wikipedia's main page. The data are available on `http://bit.ly/MASH-BCS` and can be read into R by:

```
> data=read.csv('wikipedia.csv', header=F)
```

The first column of the resulting data frame contains $n = 131$ rows; each row corresponds to one article. The first value X_i is the time in days since the article was last edited; the second value Y_i is a category corresponding to a section of the main page[1].
Following standard notation, we say that the real-valued continuous random variables $X \sim \mathcal{E}(\lambda)$ and $Y \sim \Gamma(a, b)$ if X and Y have respective probability density functions

$$f_X(x;\lambda) = \lambda e^{-\lambda x}\mathbb{I}x > 0 \qquad f_Y(y;a,b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}\mathbb{I}_{y>0}$$

where $\Gamma(a)$ denotes the Gamma function implemented by R's `gamma()`.
At each stage of the exam, you may earn bonus points by exploring and handling model misspecifications.

[1] DYK=Did you know; ITN=In the news; OTD=On this day; TFA=Today's featured article; TFL=Today's featured list; TFP=Today's featured picture, OUT=not yet on the main page (can be used as a control).

1 Exponential model

In this section, we suppose that $X_1, \dots, X_n$ are iid realizations of the $\mathcal{E}(\lambda)$ distribution with λ unknown.

1. Verify that the $(\Gamma(a,b))$ family of distributions ($a > 0$, $b > 0$) is conjugate for this model.
2. Calculate Jeffrey's prior. Is this prior proper? If not, is the associated posterior proper?
3. For the prior of your choice, compute the posterior mean and variance of λ. Check the impact of the prior.
4. Give (on paper) the analytical value of the marginal likelihood of the model.
5. For this question, we consider only the data in categories "TFA" and "OUT". We consider two possible models: (1) data in the two categories are realizations from the same exponential distribution $\mathcal{E}(\lambda)$; (2) the data in the two categories are realizations from exponential distributions with different parameters $\mathcal{E}(\lambda_{TFA})$ and $\mathcal{E}(\lambda_{OUT})$. Compute and interpret the Bayes' factor between these two models.
6. Perform and interpret any posterior predictive checks you think are relevant.

2 Mixture model

In this section, we assume that there are two parameters λ_0 and λ_1 and iid latent variables $Z_i \sim Bernoulli(\frac{1}{2})$ for $i = 1 \dots n$. Our model is independent of the Y_i and assumes that

$$X_i|(Z_i = 0) \sim \mathcal{E}(\lambda_0) \qquad X_i|(Z_i = 1) \sim \mathcal{E}(\lambda_1)$$

and we take independent $\Gamma(a,b)$ priors on λ_0 and λ_1.

7. Give the conditional posterior distribution of λ_0 and λ_1 given the (X_i) and the (Z_i).
8. Give the conditional distribution of Z_i given X_i, λ_0 and λ_1.
9. Write a Gibbs' sampler to produce a posterior sample of λ_0 and λ_1.
10. Explain on paper how you checked that your Gibbs's sampler has reached stationarity.
11. Compute the Effective Sample Size of your output.
12. Write an Importance Sampling scheme to estimate the marginal likelihood of this model.
13. We wish to compare this model to the simpler model where $\lambda_0 = \lambda_1$. Compute and interpret the relevant Bayes' factor.

3 Hierarchical model

In this section, we wish to use the category value Y_i in the second column of the data; we assimilate the values of Y_i to $\{1, \ldots, 7\}$. Our model now includes 7 parameters $\lambda_1, \ldots, \lambda_7$ and we have

$$X_i|(Y_i = j) \sim \mathcal{E}(\lambda_j).$$

We use a hierarchical model: the (λ_j) come from a common $\Gamma(a, b)$ distribution, and we wish to estimate a and b. We take the prior $\pi(a, b) \propto \frac{1}{ab}$.

14. Give the conditional distribution of λ_j given a, b and the (X_i) and (Y_i).
15. Give the form of the conditional distribution of a and b given the (λ_j), (X_i) and (Y_i).
16. Write a Gibbs' sampler to sample from the posterior distribution of a and b. You may need to use Metropolis-within-Gibbs.
17. Use your posterior sample to estimate $\lambda_1, \ldots, \lambda_7$.
18. Estimate the posterior probability of the event "the articles that tend to be edited most often are those in category ITN".

4 Gamma model

In this section, we wish to perform a sensitivity analysis and use a heavier tailed distribution. We suppose that the (X_i) are iid realizations of the $\Gamma(a, b)$ distribution with a and b to estimate.

19. *(Difficult question)* Find a conjugate prior family for a and b.
    For the rest of this section, you may use the prior you have found, or the prior $\pi(a, b) \propto \frac{1}{ab}$*, or any other prior you believe is relevant.*
20. Give the form of the posterior distribution of a and b.
21. Write a Metropolis-Hastings algorithm to sample from the posterior distribution of a and b.
22. Explain how you checked that your Metropolis-Hastings algorithm has converged.
23. Compute the Effective Sample Size of your output.
24. Write an Importance Sampling scheme to estimate the marginal likelihood of this model.
25. We wish to compare this model to the simpler model where $a = 1$. Compute and interpret the relevant Bayes' factor.

5 Combining the models

26. If several of the models studied in the previous sections seem relevant to you, propose a model which combines them, and write an inference procedure.
27. Compute the Bayes' factor to compare your combined model to each of the simpler models.