ISyE 6420

Project Report

Improving Sparse Ratings with the Multinomial-Dirichlet Conjugate Pair

Ryan Keeney

12-4-2022

Introduction

Students would like to accurately assess the performance of professors during their course planning. However, public data is often limited to a few observations and frequentist approaches poorly represent the true parameters in these situations. The Multinomial-Dirichlet conjugate pair provide Bayesian statisticians a powerful tool to evaluate categorical and multinomial data, allowing for improved evaluation of sparse data.

The goal of this analysis is estimating the posterior parameters and provide an analysis of the predicted ratings of a given professor. The motivating examples are shown in figures 1 and 2 [6,7]. Example A is a possible poor professor, with only 3 ratings, all of which are 1-star. Example B is a possible good professor, with 8 total ratings.

Figure 1 Example A, possible poor professor.

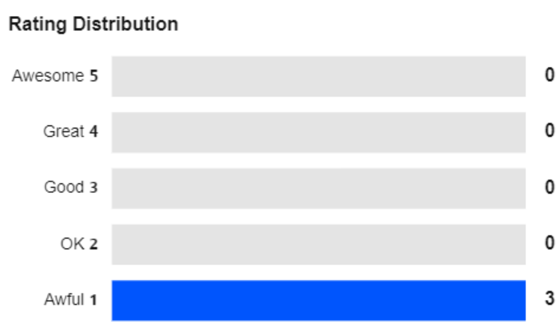
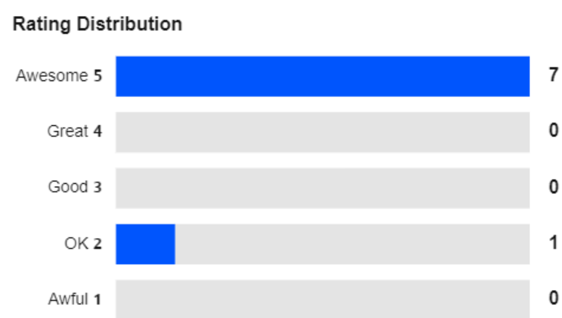


Figure 2 Example B, possible good professor



Methodology (Analysis)

Data sources

Rating data was collected from Mendeley Data [6] and RateMyProfessor.com [7]. The background data from Mendeley Data contains almost 1 million professors' information. The motivating examples A and B are actual ratings of two different professors; however, their names are not relevant to this analysis and have been censored.

Problem setup

The ratings are a multinomial distribution, with $n = 5$. Count data is observed across 5 ratings, and then used to estimate an overall rating, as well as the individual probability of each rating.

$$N_1, \dots, N_k \sim Mult(\theta_1, \dots, \theta_k)$$

Frequentist approach

The frequentist estimation of the multinomial parameters is just the fraction of times rating i is observed. If the data is sparse (few observations), the observed data may not represent the true parameters very well [1].

Deriving the MLE of $\hat{\theta}_i$

Log-likelihood of multinomial

$$\ell(\theta; D) = \log p(D|\theta) = \sum_i N_i \log \theta_i$$

Maximize, subject to the constraint $\sum_k \theta_i = 1$

$$\ell = \sum_{i=1}^{k} N_i \log \theta_i + \lambda \left(1 - \sum_{i=1}^{k} \theta_i\right)$$

Take derivate with respect to θ_k, λ

$$\frac{N_i}{\theta_i} - \lambda = 0, \qquad 1 - \sum_{i=1}^{k} \theta_i = 0$$

Use the constraint, $\sum_K \theta_k = 1$

$$N_i = \lambda \theta_i, \qquad \sum_{i=1}^{k} N_i = \lambda \sum_{i=1}^{k} \theta_i$$

$$N = \lambda$$

$$\hat{\theta}_i = \frac{N_i}{N} \blacksquare$$

Bayesian approach

As will be demonstrated, a Bayesian approach is preferable to the frequentist estimation when there are limited observations since we can incorporate a prior. If a Dirichlet prior [figure 3] is selected, it forms a conjugate pair with a multinomial distribution. It also has the advantages of simple non-informative priors, such as uniform (flat) and the least informative prior (Jeffreys Prior).

Deriving the Multinomial-Dirichlet Conjugate Pair [1,5]

Set up observation and prior

$$N_1, \dots, N_k \sim Mult(\theta_1, \dots, \theta_k)$$

$$\theta_1, \dots, \theta_k \sim Dirichlet(\alpha_1, \dots \alpha_k)$$

Find the posterior

$$f(\theta|n) \propto f(\theta|n) = f(\theta_1, \dots, \theta_k|\alpha_1, \dots, \alpha_1, \dots \alpha_k) \prod_{n_i \in N} f\left(n_j\middle|\theta_1, \dots \theta_k\right)$$

$$\propto \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \prod_{n_j \in N} \prod_{i=1}^{k} \theta_i^{n_j^{(i)}}$$

$$= \prod_{i=1}^{k} \theta_i^{\alpha_i - 1 + \sum_{n_j \in N} n_j^{(i)}}, \qquad \sum_{n_j \in N} n_j^{(i)} = N_i$$

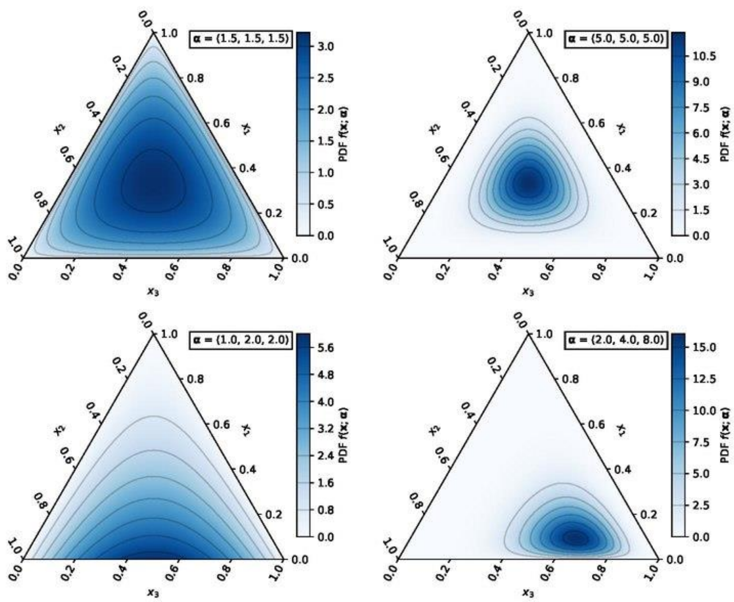
Which is a Dirichlet distribution:

$$\alpha_i^* = \alpha_i + N_i$$

Thus, the posterior is:

$$\pi\big((\theta_1, \dots, \theta_k)|(N_1, \dots, N_k)\big) \sim Dirichlet(\alpha_1 + N_1, \dots, \alpha_k + N_k) \blacksquare$$

Figure 3 Example Dirichlet distribution with n=3.



Possible priors

The Dirichlet prior also has the advantage of simple non-informative priors. The following priors were evaluated.

1. Non-informative, uniform: $\alpha_1 = \cdots \alpha_k = 1$
2. Least-informative, Jeffreys Prior: $\alpha_1 = \cdots \alpha_k = \frac{1}{2}$

3. Informative prior using existing patterns in the data or expert recommendation (not explored in depth).

Evaluation

After the model was constructed, the posterior was sampled 1000 times. The posterior parameter estimations were compared to the frequentist MLE estimations for the examples A and B, and 95% equitailed credible sets were recorded. Additionally, the expected rating and probability of being a good (rating ≥4) or poor (rating ≤ 2) was estimated.

Additionally, parameter convergence and alternative priors were evaluated.

Results

The Bayesian approach is preferable to the frequentist estimation when there are limited observations since we can incorporate a prior. Since the data has few observations and observed data may not represent the true parameters well. If only the prior is used, the expected $\theta_i = \frac{Ni}{N} = 0.2$ (flat or uniform prior) for all i, and the expected rating is 3.0. For the Bayesian approach, the model accounts for how many observations the data has.

Frequentist results

In example A, the possible poor professor with three 1-star ratings, the frequentist estimation is $\hat{\theta}_{i=1} = 1,\ \hat{\theta}_{i\neq 1} = 0$, and the expected rating is 1.0, as shown in table 1.

Table 1 Frequentist MLE estimation of multinomial parameters for example A (possible poor professor)

	1 STAR	2 STARS	3 STARS	4 STARS	5 STARS
N	0	0	0	0	3
$\boldsymbol{\hat{\theta}_i}$	0	0	0	0	1

Bayesian results

When a flat prior is used ($\alpha_i = 1$) the posterior estimation of θ_i is $[0.07, 0.16, 0.78, 0.76, 0.62]$ for example A and $[0.50, 0.12, 0.12, 0.13, 0.13]$ for example B [figure 4, figure 5]. With only three total observations, the prior is still significantly influencing the posterior, providing a more reasonable estimation than the frequentist approach. The 95% equitailed credible sets are also shown in figures 4 and 5.

Figure 4 Bayesian estimation of the posterior I, example A (possible poor professor)

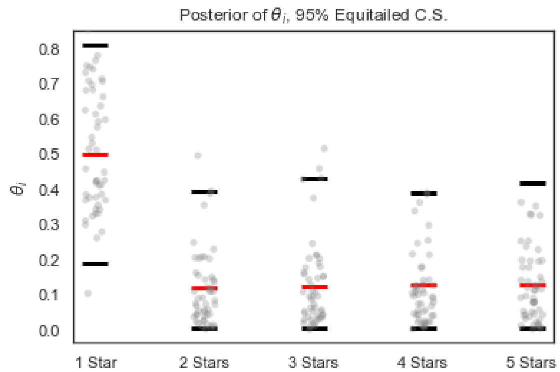
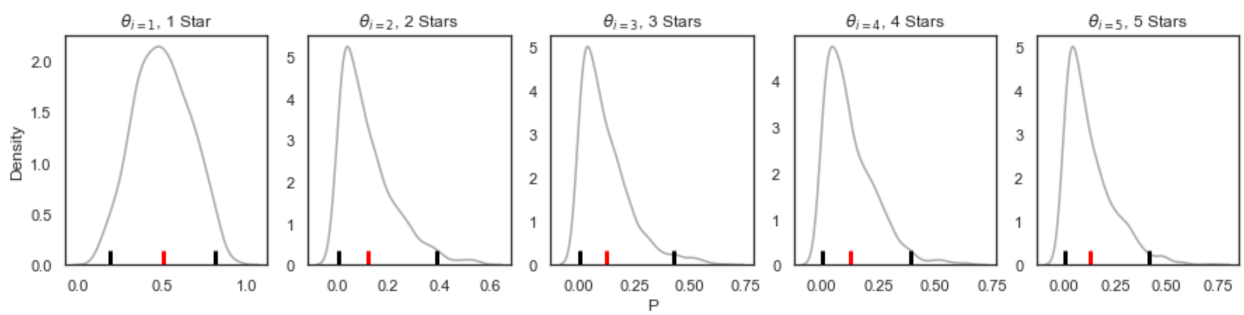


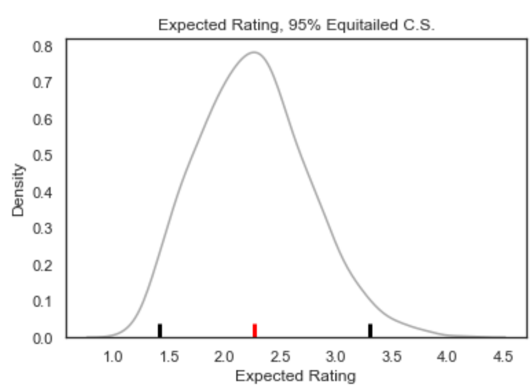
Figure 5Bayesian estimation of the posterior II, example A (possible poor professor)



Predicting professor ratings

Compared to the frequentist approach where the expected rating was 1.0 stars for example professor A, the Bayesian estimation is $E[rating] = 2.25$ stars, with a 95% equitailed credible set of : $[1.49, 3.4]$, as shown in figure 6.

Figure 6 The posterior prediction of expected stars for example A is closer to the prior mean than the MLE estimate.



To evaluate the probability of each example being a good (rating ≥4 stars) or poor (rating ≤ 2 stars) was estimated from the posterior samples. For examples A and B, the professor's predicted probability of having a 1- or 2-star rating is shown in tables 2 and 3.

Table 2 Example A P [$\theta_{i\leq 2}$]

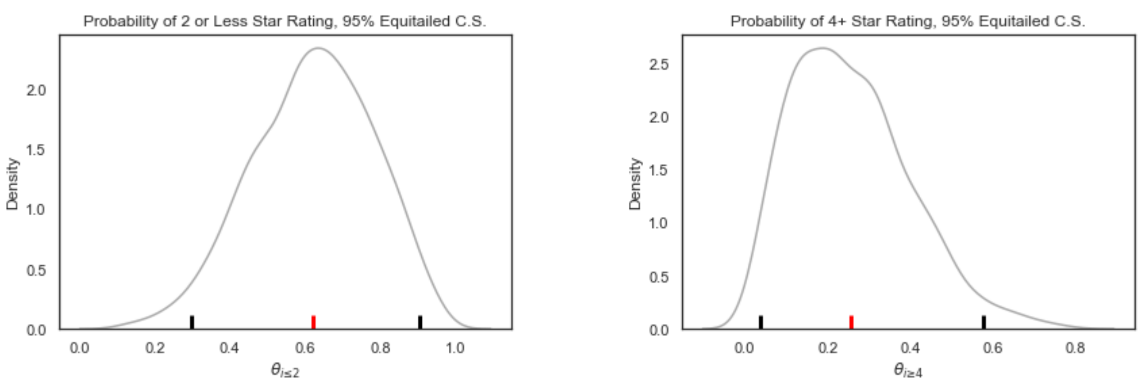
Prior	Posterior	Freq., MLE
0.40	0.62	1.00

Table 3 Example B P [$\theta_{i\leq 2}$]

Prior	Posterior	Freq., MLE
0.40	0.27	0.125

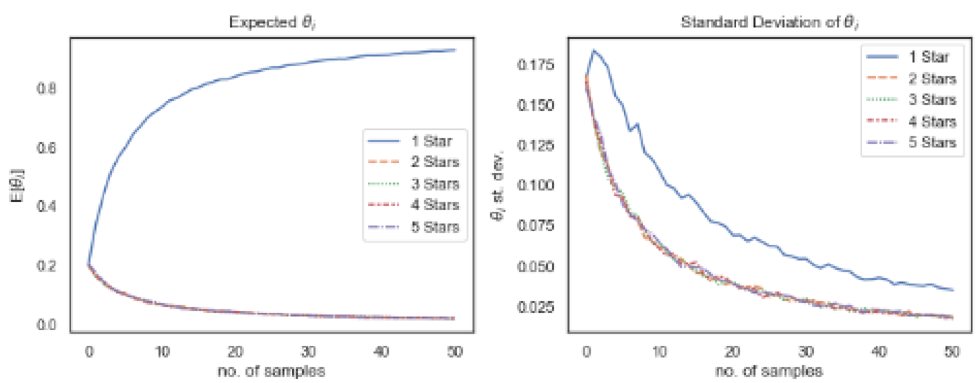
The 95% equitailed credible sets for the posterior prediction of the probability that example A is a good or bad professor are shown in figure 7.

Figure 7 Example A, the possible poor professor, is still a 'good' professor about 1/4 of the time.



$95\%\ Equitailed\ CS[0.29, 0.90]$ $95\%\ Equitailed\ CS[0.03, 0.58]$

Figure 8 About 10 ratings need to be collected before the posteriors stabilizes



Posterior stability

The Bayesian Multinomial-Dirichlet model's posterior predictions stabilize with more observations. With a flat prior, about 10 observations are required before the posterior estimates stabilize [figure 8]. With a Jeffreys prior, about 5-6 observations are required before the posterior estimates stabilize.

Conclusion

The Bayesian approach is preferable to the frequentist estimation when there are limited observations since we can incorporate a prior. Since the data has few observations and observed data may not represent the true parameters well. In both examples, the Bayesian approach provided more reasonable estimations of a professors rating, especially when limited observations existed.

Prior selection

Based posterior stabilization curves, I recommend a flat prior which is informative than Jeffreys prior while not still avoiding strong assumptions. For future analysis, more informative priors may be evaluated.

Bias

Students rate professors based on a variety of factors [2], these ratings may involve biases which are not accounted for in this analysis. Some examples of biases and obstacles are listed below.

- Overall ratings may be confounded with other factors such as topic, class structure, and requirements, and personal biases such as race, sex, and communication style.
- Students are more likely to rate professors they feel strongly about
- Students self-select for classes (e.g., select easy classes)

Extensions

The Multinomial-Dirichlet conjugate pair provide Bayesian statisticians a powerful tool to evaluate categorical and multinomial data. This approach may be adapted for product recommendations, although it is best applied where data is limited, such as performance reviews or to supplement qualitative research applications.

Appendix

Code

Code submitted separately as .ipynb.

Citations

[1] Vidakovic, B. (2017). *Engineering biostatistics: An introduction using Matlab and WinBUGS*. Wiley.
[2] James Otto, Douglas A. Sanford Jr & Douglas N. Ross (2008) Does ratemyprofessor.com really rate my professor?, Assessment & Evaluation in Higher Education, 33:4, 355-368, DOI: 10.1080/02602930701293405
[3] Nerebur, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons
[4] Murphy, Kevin P. The University of British Columbia. (2007) Frequentist parameter estimation, https://www.cs.ubc.ca/~murphyk/Teaching/CS340-Fall07/reading/paramEst.pdf
[5] Tu, Stephen. The Dirichlet-Multinomial and Dirichlet-Categorical models for Bayesian inference. https://stephentu.github.io/writeups/dirichlet-conjugate-prior.pdf
[6] He, Jibo (2020), "Big Data Set from RateMyProfessor.com for Professors' Teaching Evaluation", Mendeley Data, V2, doi: 10.17632/fvtfjyvw7d.2
[7] *Find and rate your professor or campus.* Rate My Professors. (n.d.). Retrieved December 1, 2022, from https://www.ratemyprofessors.com/