

Teaching Bayesian Modeling to Beginners: A Case Study Using Educational Data

Ranadeep Daw¹, Subhash Bagui²

¹Department of Mathematics and Statistics, University of West Florida, Pensacola, FL 32514

Email: [rdaw\[at\]uwf.edu](mailto:rdaw[at]uwf.edu)

²Department of Mathematics and Statistics, University of West Florida, Pensacola, FL 32514

Corresponding Author Email: [sbagui\[at\]uwf.edu](mailto:sbagui[at]uwf.edu)

Abstract: *This study presents a case-based approach to teaching Bayesian statistical modeling to an undergraduate student with no prior exposure within a single semester. The project applied a Beta-binomial model to estimate population proportions using educational data related to student accommodation needs. A comparison between frequentist and Bayesian approaches demonstrated that Bayesian estimates produced narrower interval widths and incorporated prior information effectively. The instructional design combined guided theory with hands-on application, enabling the student to complete a full data analysis cycle, including modeling, interpretation, and presentation. The results suggest that project-based Bayesian instruction is feasible at the undergraduate level and can improve conceptual understanding of uncertainty and inference. This approach provides a practical framework for integrating Bayesian methods into statistics curricula.*

Keywords: Bayesian inference; Beta-binomial model; statistics education; undergraduate pedagogy; proportion estimation; empirical Bayes; credible intervals.

Mathematics Subject Classification: 62C10, 97K70, 97-02.

1. Introduction

Bayesian statistical methods have witnessed a significant increase in relevance and application across diverse fields, including in the social sciences and education (Moore, 1997; Van de Schoot et al., 2014). This growth is driven by advances in computation and methodology, including increased computational capacity, a growing recognition of the advantages of Bayesian frameworks, and their capability to formally incorporate “prior” knowledge into statistical models, especially significant in discerning relevant signals from noise in the high-dimensional and high-dimensional problems nowadays (Bon et al., 2023; Hoegh, 2020). Consequently, there has been a pedagogical undertaking to integrate Bayesian thinking into introductory statistics courses (Albert, 2002; Ferrari, 2022; Jessop, 2010; Wulff & Robinson, 2014). The initial drive was partly motivated by the suggestion that Bayesian inference offers a more intuitive approach to understanding uncertainty compared to traditional frequentist methods, such as the improper overuse of p-values (Dogucu & Hu, 2022). While the utility of Bayesian methods is increasingly recognized, their adoption in undergraduate education remains limited, potentially not meeting the demands of today’s rapidly evolving AI-driven society (Zhu et al., 2024). Therefore, there is often a need to introduce Bayesian methods to individuals with little to no prior experience, especially in the context of short case studies or capstone projects.

This paper presents a case study detailing the experience of providing practical, project-focused instruction in Bayesian modeling to a student with no prior background in Bayesian modeling. The core of this instruction focused on a bachelor’s research project, including analysis, report writing, and an oral presentation of the results. The project focuses on three (population) proportion estimation problems, employing a

conjugate Beta-binomial approach, using self-collected data derived from various accommodation needs of school students - the counts of disability accommodations, McKinney-Vento homeless requirements, and disciplinary actions within a specific school, all related to the student-selected categories.

1.1 Overall Experience

The University of West Florida, Department of Mathematics and Statistics policy requires that all senior students in the Mathematical Sciences department complete “Proseminar”. The Proseminar is a capstone course in which students study a specific topic in mathematics, statistics, or related fields and present their findings to an audience. Students at this stage typically have some experience with introductory statistical techniques, including the calculation of summary statistics, estimation, confidence intervals, hypothesis testing, and basic distributional theory. In this case, the student showed keen interest in analyzing educational requirements across various facets of a chosen school, using self-collected data. However, with only two years of data available, estimating the proportions of requirements in the population and presenting them as a statistics education project posed a significant methodological challenge. Consequently, it was determined that introducing the student to Bayesian methodology, using a portion of the existing data as prior information, would be a suitable approach for both the student’s purpose and the educator’s experience teaching introductory Bayesian methodology required for a data analysis project.

The project management context involved individualized sessions alongside weekly meetings. At the start of the semester, the student had no prior knowledge of the fundamental rules and techniques of Bayesian statistics. Therefore, the initial one-on-one sessions were designed to

Volume 15 Issue 5, May 2026

Fully Refereed | Open Access | Double Blind Peer Reviewed Journal

www.ijsr.net

introduce Bayesian statistical concepts, starting with fundamentals and gradually progressing to Bayes' rule and posterior computation. As the semester progressed, the instructional focus gradually shifted toward applying these techniques within the context of the student's project and formalizing the research report. Ultimately, at the end of the semester, the project was presented to an audience comprising the evaluation committee, regular mathematics faculty, students, and faculty from other disciplines. The written component of the project was submitted to the evaluation committee for final approval, and the committee approved.

1.2 Problem Formulation

The project focused primarily on a proportion estimation problem. As an introductory exercise, the goal was to use a conjugate prior, i.e., one for which the prior and posterior belong to the same family of distributions. For data collection, various requirements were targeted to address educational needs – namely the accommodation qualifications (AQ), McKinney-Vento and free/reduced-lunch qualifications (MV+), and disciplinary action records (DA). AQ included students with various accommodations, such as IEPs, 504 Plans, ELL status, and medical conditions. MV+ covered students eligible for the McKinney-Vento Act (i.e., support for students experiencing homelessness or housing instability) and those who qualified for free or reduced-price lunch. DA referred to students facing multiple disciplinary actions, i.e., those with two or more days of suspension or more severe penalties.

The student collected the data from selected surveyed classes, which were kept anonymous to protect privacy. The student was explicitly told that this sample, while practical for an introductory exercise, might not accurately represent the broader population, such as the entire district, county, or state. The primary pedagogical goal was to introduce a practical application of Bayesian analysis in the context of a capstone project. The investigation focused on estimating the proportion of students in the defined class who met the criteria for each of the three categories (AQ, MV+, and DA). For each category, a student meeting the specified criteria was considered a "success," and the rest were counted as "failures." Following this formal definition of the problem, the project was intended to conduct both a frequentist and a Bayesian analysis for comparison. Furthermore, the corresponding confidence and credible intervals were compared to better understand the uncertainty, and a comprehensive comparison of the frequentist and Bayesian methods was presented. Table 1 presents the data used:

Table 1: Data table for Project 1

Category	Prior year	Current year
Total Students	776	153
AQ	302	27
MV+	286	36
DA	211	26

Table 1 uses pseudo data for illustration purposes only. Actual values are concealed for confidentiality reasons.

2. Methodology

This section outlines the statistical methodologies used in the project, specifically chosen to introduce both frequentist and Bayesian approaches to proportion estimation.

2.1 Frequentist Approach

This subsection introduces the frequentist framework for estimating the population proportion of students who qualify for McKinney-Vento accommodations. At this stage of the project, the student was already familiar with the fundamental concepts of the Bernoulli and binomial distributions, which formed the basis of the analysis. The frequentist approach focused on estimating the proportion of "successes" – defined in Section 3 as a student receiving accommodations – using only the "Current Year" data and intentionally excluding the "Prior Year" information from this analysis. A single trial was modeled using a Bernoulli random variable X with success probability θ as:

$$\mathbb{P}(X = x) = \theta^x(1 - \theta)^{1-x}; \quad x = 0,1.$$

Its mean and variance are given by $E[X] = \theta$ and $V[X] = \theta(1 - \theta)$. For a sample with n students, the total number of successes (i.e., the total number of students requiring accommodation), Y , is modeled as a binomial distribution:

$$\mathbb{P}(Y = y) = \binom{n}{y} \theta^y(1 - \theta)^{n-y}; \quad y = 0,1, \dots, n,$$

with its mean $E[Y] = n\theta$ and variance $\text{Var}[Y] = n\theta(1 - \theta)$.

To estimate the unknown proportion θ , two classical frequentist methods were introduced: Maximum Likelihood Estimation (MLE) and the Method of Moments (MoM). The MLE approach maximizes the likelihood of observing the data under the binomial model, while the MoM equates sample estimates to population parameters. For the binomial case, both methods yield the same estimator with the same variance:

$$\hat{\theta} = \frac{y}{n}; \quad s(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}.$$

Finally, confidence intervals for the proportions θ -s were constructed using the Central Limit Theorem (CLT, see Chapter 7 of Montgomery & Runger, 2019; Bagui et al., 2013). Under the CLT, the sampling distribution of $\hat{\theta}$ is approximately normal for large n . This leads to the computation of a $100(1 - \alpha)\%$ confidence interval of θ

using the formula $\hat{\theta} \pm z_{\alpha/2} \sqrt{\frac{\hat{\theta}(1 - \hat{\theta})}{n}}$. For this particular project, the interest was for $\alpha = 0.05$, i.e., the 95% confidence interval, for which $z_{0.05/2} \approx 1.96$. Table 2 shows the results for the frequentist study.

Table 2: Frequentist results

Category	Estimate ($\hat{\theta}$)	Standard Error $s(\hat{\theta})$	Confidence interval	Interval Width
AQ	0.1765	0.0308	(0.1161, 0.2369)	0.1208
MV+	0.2353	0.0343	(0.1681, 0.3025)	0.1344
DA	0.1699	0.0304	(0.1104, 0.2294)	0.1190

2.2 Bayes Beta-Binomial Model

Since the central pedagogical objective of this project was to introduce Bayesian statistical inference, this section focuses on modeling binomial outcomes within a Bayesian framework. For ease of implementation, only a conjugate prior model – the Beta-binomial conjugate model for proportion data – was considered appropriate.

Since we model the count of students requiring accommodations with a Binomial distribution, i.e., $Y \sim \text{Bin}(n, \theta)$, the prior distribution is chosen as a Beta distribution, $\theta \sim \mathbb{B}(\alpha, \beta)$ (\mathbb{B} denotes a Beta distribution), because the Beta distribution naturally models proportions with support on (0,1). The probability density function (pdf) of the Beta prior for θ is given by:

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1}; \quad 0 < \theta < 1; \alpha, \beta > 0,$$

where $\Gamma(\cdot)$ is the Gamma function. The mean and variance of the prior distribution are respectively, given by $E[\Theta] = \frac{\alpha}{\alpha + \beta}$

and $\text{Var}[\Theta] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{E[\Theta](1 - E[\Theta])}{\alpha + \beta + 1}$. Under this

formulation and applying Bayes' rule, the posterior distribution of θ is proportional to the product of the prior and the likelihood, i.e., this expression is recognized as the core form of a Beta distribution (see Appendix A), which confirms that the posterior distribution of θ given Y is also a Beta distribution (i.e., conjugate family) with parameters $(y + \alpha)$

and $(n - y + \beta)$. As a result, the posterior mean and variance of θ are given by:

$$\hat{p}_{\text{Bayes}} = E[\Theta | Y = y] = \frac{y + \alpha}{n + \alpha + \beta} \quad \text{and}$$

$$\text{Var}[\Theta | Y = y] = \frac{(y + \alpha)(n - y + \beta)}{(n + \alpha + \beta)^2(n + \alpha + \beta + 1)} = \frac{\hat{p}_{\text{Bayes}}(1 - \hat{p}_{\text{Bayes}})}{(n + \alpha + \beta + 1)},$$

respectively.

To guide the student in applying Bayesian inference meaningfully, the “Prior Year” data from 2023–2024 were used to inform the selection of the prior parameters alpha (α) and beta (β). This approach resembles an empirical Bayes strategy (Carlin & Louis, 2000), in which the prior distribution is informed by an initial data-driven estimate. Because the Beta distribution models proportions, the goal was to select Beta priors whose means closely matched the observed success rates from the prior year, while keeping the standard deviations small (around 0.05) to emphasize the influence of prior information. (Another way to choose hyperparameters for the prior is to fix the ratio $\frac{\alpha}{\beta}$ and set the mean using the observed success rates.) The student was asked to experiment with various (α, β) combinations to achieve these targets, although the formal implications of this were not fully discussed during the initial instruction, in line with the pedagogical strategy of emphasizing hands-on learning first. As an example, see Figure 1 for the scaling difference between two Beta distributions with parameters (2, 3) and (20, 30). A thorough, formal mechanism for prior selection is provided in Appendix B for those interested in a deeper understanding.

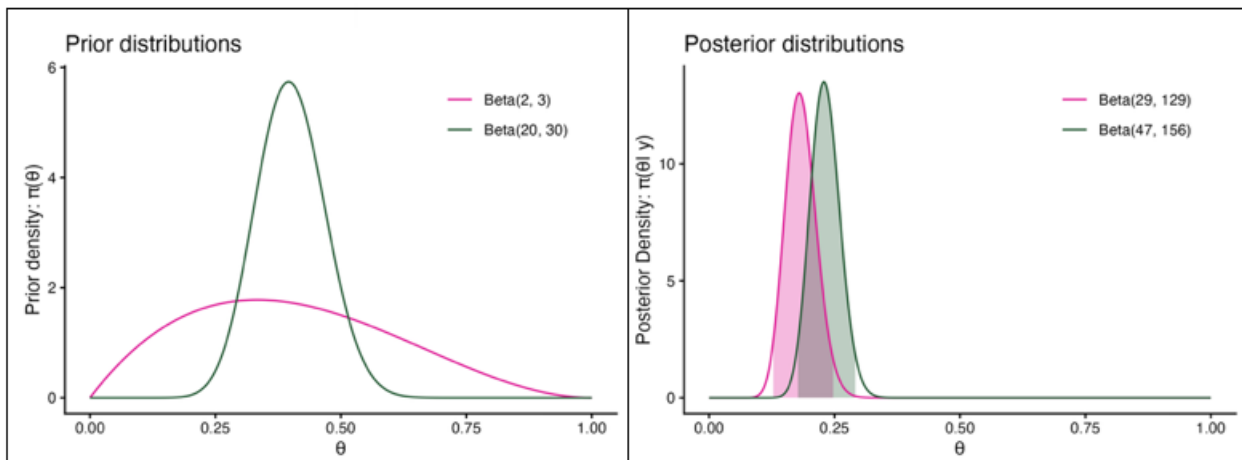


Figure 1: Plots of Prior and Posterior distributions

Figure 1: The left plot shows two Beta prior candidates for the AQ category, differing in variance. The right plot shows the corresponding posterior Beta distributions under the respective priors, with shaded 95% credible intervals. The Beta (20, 30) distribution was selected for this project.

Table 3 presents the final prior parameter choices for the three categories, along with the observed prior-year statistics and the resulting prior summaries.

Table 3: Prior parameters from prior year data

Category	Total Students	Total successes	Success proportion	Prior parameters	Prior Mean	Prior SD
AQ	776	302	0.39	$\alpha = 20, \beta = 30$	0.4	0.07
MV+	776	286	0.37	$\alpha = 35, \beta = 65$	0.35	0.05
DA	776	211	0.27	$\alpha = 21, \beta = 49$	0.30	0.05

Combining the prior distributions with the “Current Year” data (in Table 1 from Section 3), the posterior distributions for each category were derived using the Beta-binomial

updating formula described above. The posterior means and standard deviations were computed from the formulas given above. We computed both approximate and exact credible

intervals for comparison purposes. Here, our posterior distribution is a Beta distribution. When the hyperparameters (α, β) of the Beta distribution are large, the CLT implies that the distribution approaches a normal distribution. Thus, 95% approximate credible intervals can be computed by applying the CLT to the posterior distribution, using the corresponding

means and variances, as our hyperparameters are large. See Table 4 for the results. The posterior hyperparameters in column 2 are obtained by combining data from column 3 of Table 1 and the prior hyperparameter numbers from column 5 of Table 3. From Table 1, note the sample size $n = 153$ and $z_{0.025} = 1.96$.

Table 4: Posterior estimation

Category	Posterior distribution	Posterior mean	Posterior SD	Credible interval	Interval width
AQ	$\mathbb{B}(47, 156)$	0.2315	0.0295	(0.1737, 0.2893)	0.1156
MV+	$\mathbb{B}(71, 182)$	0.2806	0.0282	(0.2253, 0.3359)	0.1104
DA	$\mathbb{B}(47, 176)$	0.2108	0.0273	(0.1573, 0.2643)	0.1070

Note that we could have used the exact posterior distribution (i.e., Beta) to construct the 95% credible interval by noting the 2.5th and 97.5th quantiles from the corresponding beta distributions. Beta quantiles can be obtained using standard statistical software such as R, SAS, or Minitab. Another possibility is to use the transformation to an F distribution based on the following identity, avoiding the need for statistical software and instead using the F critical value table to obtain Beta critical values:

$$X = \frac{n_1 F_{n_1, n_2}}{n_1 F_{n_1, n_2} + n_2} \square \text{Beta}\left(\frac{n_1}{2}, \frac{n_2}{2}\right).$$

Table 5: Exact 95% credible intervals using Beta distribution

Category	Posterior distribution	Exact credible interval	Exact interval width
AQ	$\mathbb{B}(47, 156)$	(0.1762, 0.2918)	0.1156
MV+	$\mathbb{B}(71, 182)$	(0.2271, 0.3375)	0.1104
DA	$\mathbb{B}(47, 176)$	(0.1599, 0.2665)	0.1066

While both approaches yield nearly identical results, the exact distribution yields slightly more accurate tail estimates. In our example, the resulting exact intervals are summarized in Table 5 above.

3. Concluding Remarks

This study demonstrates that Bayesian statistical modeling can be effectively introduced to undergraduate students within a single semester through a structured, project-based approach. The use of a Beta-binomial framework enabled intuitive understanding of prior and posterior reasoning while allowing comparison with classical methods. However, the findings are limited by the small sample size and the single-student case study design, which restricts generalizability. Future work should evaluate this instructional approach across multiple cohorts and incorporate formal assessment metrics to measure learning outcomes. Expanding the framework to include computational tools and more complex Bayesian models could further enhance its applicability in modern statistics education.

From a pedagogical standpoint, this approach required the instructor to introduce Bayesian concepts in an applied, incremental manner. The analysis began with a familiar frequentist framework for estimating binomial proportions and then naturally transitioned to the Bayesian framework using conjugate priors. This allowed direct comparisons

between frequentist and Bayesian results and helped build an understanding of how Bayesian techniques might reduce uncertainty through the choice of prior parameters. The project was successfully concluded with both a public presentation and a formal written report.

More broadly, this experience shows that students with a limited background in Bayesian methods can complete a complete research project – including data collection, cleaning, modeling, interpretation, and communication – when supported by individualized guidance and context-specific learning. In a rapidly evolving data-driven era, Bayesian methods offer not only computational efficiency but also pedagogical value. Integrating them into undergraduate education – particularly in capstone settings – can enhance students’ statistical reasoning and autonomy for modern data challenges.

Authors’ Contribution

Subhash C. Bagui supervised and instructed the undergraduate student throughout the project. Ranadeep Daw contributed uniquely to the preparation and completion of this manuscript and was not involved in the instructional component.

References

- [1] Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. *Proceedings of the Sixth International Conference on Teaching Statistics*, 1–14.
- [2] Bagui, S.C., Bhaumik, D.K. & Mehra, K.L. (2013). A few counterexamples useful in teaching central limit theorems, *The American Statistician*, **67**(1), 49–56, DOI: 10.1080/00031305.2012.755361.
- [3] Bon, J. J., Bretherton, A., Buchhorn, K., Cramb, S., Drovandi, C., Hassan, C., Jenner, A. L., Mayfield, H. J., McGree, J. M., Mengersen, K., et al. (2023). Being Bayesian in the 2020s: Opportunities and challenges in the practice of modern applied Bayesian statistics. *Philosophical Transactions of the Royal Society A*, **381**(2247), 20220156.
- [4] Carlin, B. P., & Louis, T. A. (2000). Empirical Bayes: Past, present, and future. *Journal of the American Statistical Association*, **95**(452), 1286–1289.
- [5] Dogucu, M., & Hu, J. (2022). The current state of undergraduate Bayesian education and recommendations for the future. *The American Statistician*, **76**(4), 405–413.

- [6] Ferrari, D. (2022). Teaching Bayesian statistics. *PS: Political Science & Politics*, 55(1), 230–235.
- [7] Hoegh, A. (2020). Why Bayesian ideas should be introduced in the statistics curricula and how to do so. *Journal of Statistics Education*, 28(3), 222–228.
- [8] Jessop, A. (2010). Bayes Ice-Braker, *Teaching Statistics*, 32(1), 13-16.
- [9] Montgomery, D. C., & Runger, G. C. (2019). Applied statistics and probability for engineers. *John Wiley & Sons*.
- [10] Moore, D. S. (1997). Bayes for beginners? Some pedagogical questions. *Springer*.
- [11] Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860.
- [12] Wulff, S. S., & Robinson, T. J. (2014). What is the probability that you are a Bayesian? *Journal of Statistics Education*, 22(2).
- [13] Zhu, T., Zhang, K., & Wang, W. Y. (2024). Embracing AI in education: Understanding the surge in large language model use by secondary students. *arXiv Preprint arXiv:2411.18708*, <https://doi.org/10.48550/arXiv.2411.18708>.

Appendix A

I. To derive the posterior distribution for a Beta-Binomial family, we first explicitly express the likelihood function as a function of θ as the following:

$$P(Y = y | \theta) \propto \theta^y (1 - \theta)^{n-y}$$

Now, with a beta prior distribution on θ and applying Bayes' theorem, we get the following:

$$\begin{aligned} P(\theta | Y = y) &\propto P(Y = y | \theta)\pi(\theta) \\ &\propto \theta^y (1 - \theta)^{n-y} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \theta^{y+\alpha-1} (1 - \theta)^{n-y+\beta-1} \end{aligned}$$

The last line is the kernel part of a Beta distribution with parameters $(y + \alpha)$ and $(n - y + \beta)$. This shows that the Beta-Binomial is a conjugate prior family, i.e., the Beta prior on the parameters of a Binomial distribution leads to the same family of Beta posterior of the parameter of interest θ .

II. When choosing prior parameters for a Beta distribution, it is often helpful to think in terms of a target mean (μ) and standard deviation (σ), informed by domain knowledge or prior data. For example, in our case, we used the prior year's proportions and targeted a standard deviation of approximately 0.05 (see Section 4). Recall that for a Beta(α, β) distribution, the mean and standard deviation are given by:

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}}$$

Given a desired mean μ and variance σ^2 , the parameters α and β can be solved as:

This approach allows one to select priors that reflect both your best estimate for the proportion and your confidence in that estimate. Below is a small example with an R implementation for a targeted mean 0.1 and standard deviation 0.05:

```
mu = 0.1; sigma = 0.05
alpha = mu * ((mu * (1 - mu) / sigma^2) - 1)
beta = (1 - mu) * ((mu * (1 - mu) / sigma^2) - 1)
cat(sprintf("Alpha = %.2f, Beta = %.2f",
           alpha, beta))
# Alpha = 3.50, Beta = 31.50 (not run)
```

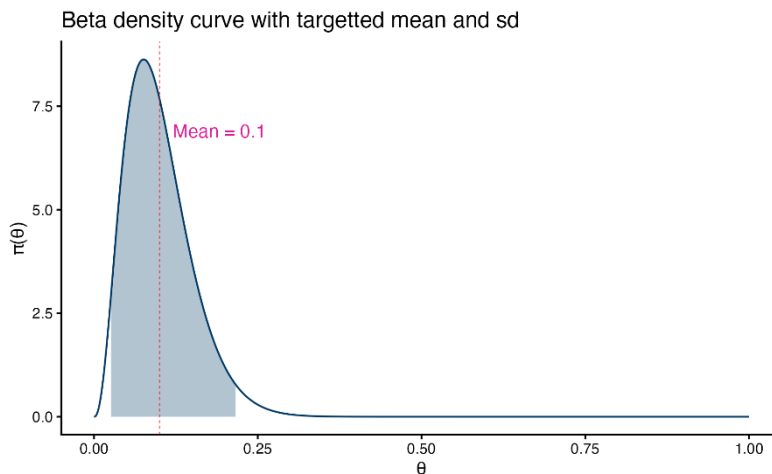


Figure A1: Plot of Beta density with targeted mean and s.d.

III. The Bayes estimate with quadratic loss is the posterior mean, which can be expressed as

$$\delta(y) = E(\theta | Y = y) = \hat{p}_{Bayes} = \frac{y + \alpha}{n + \alpha + \beta} = \left(\frac{\alpha + \beta}{n + \alpha + \beta} \right) \left(\frac{\alpha}{\alpha + \beta} \right) + \left(\frac{n}{n + \alpha + \beta} \right) \hat{p}_{mle},$$

which shows that Bayes' estimate (\hat{p}_{Bayes}) is a weighted average of prior mean $\alpha/(\alpha + \beta)$ and the maximum likelihood estimate $\hat{p}_{mle} = y/n$ with weights $w_1 = (\alpha + \beta)/(n + \alpha + \beta)$ given to prior mean and $w_2 = (n)/(n + \alpha + \beta)$ for the MLE \hat{p}_{mle} . The uncertainty in the Bayes estimate corresponds to the standard deviation of the posterior distribution, which is $\sqrt{(y + \alpha)(n - y + \beta) / (n + \alpha + \beta)^2 (\alpha + \beta + 1)}$. The change in prior mean to posterior mean can be assessed as

$$D = \frac{|\text{prior mean} - \text{posterior mean}|}{\text{posterior std. dev.}}$$

The changes between prior and posterior means, relative to the posterior s.d., across various categories are shown in the following table:

Table A1: Changes between prior and posterior means in relation to posterior s.d.

Category	Prior distribution	Prior mean	Posterior distribution	Posterior mean	Posterior SD	D
AQ	$\mathbb{B}(20, 30)$	0.40	$\mathbb{B}(47, 156)$	0.2315	0.0295	5.712
MV+	$\mathbb{B}(35, 65)$	0.35	$\mathbb{B}(71, 182)$	0.2806	0.0282	2.461
DA	$\mathbb{B}(21, 49)$	0.30	$\mathbb{B}(47, 176)$	0.2108	0.0273	3.267