A Novel Approach to Bayesian Estimation using the Sandwich Theorem and the Dirichlet Distribution

Soumadeep Ghosh

Kolkata, India

Abstract

In this paper, we propose a novel framework for Bayesian estimation that leverages the Sandwich Theorem for robust convergence analysis and the Dirichlet distribution as a flexible prior for categorical and compositional data. Our approach unifies robust variance estimation, nonparametric Bayesian modeling, and convergence guarantees, providing a powerful toolkit for modern statistical inference. We illustrate the methodology with theoretical development, simulation, and vector-based visualizations.

The paper ends with "The End"

1 Introduction

Bayesian estimation is a foundational technique in modern statistics, allowing for the formal incorporation of prior knowledge and the updating of beliefs in light of new data [1]. The Dirichlet distribution serves as a conjugate prior for multinomial and categorical models, making it indispensable in applications ranging from topic modeling to genetics [2]. Meanwhile, the Sandwich Theorem (also known as the Squeeze Theorem) provides a powerful tool for establishing convergence properties of estimators, especially in the context of bounding and asymptotic analysis [3].

Recent advances have explored the intersection of these concepts, combining Bayesian estimation with robust (sandwich) variance estimation and Dirichlet-based priors to address challenges in complex and high-dimensional data analysis [5]. In this article, we present a unified approach that synthesizes these ideas, offering both theoretical insights and practical tools for robust Bayesian inference.

2 Background

2.1 Bayesian Estimation

Given data $\mathbf{x}$ and parameter vector $\boldsymbol{\theta}$, Bayesian estimation updates the prior belief $p(\boldsymbol{\theta})$ using the likelihood $p(\mathbf{x}|\boldsymbol{\theta})$ to obtain the posterior:

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})}$$

where $p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal likelihood [1].

2.2 The Dirichlet Distribution

The Dirichlet distribution $\mathrm{Dir}(\boldsymbol{\alpha})$ is defined over the $(K-1)$-simplex:

$$f(\mathbf{p};\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{K} p_i^{\alpha_i - 1}$$

where $\mathbf{p} = (p_1, \dots, p_K)$, $p_i \geq 0$, $\sum_{i=1}^{K} p_i = 1$, and $B(\boldsymbol{\alpha})$ is the multivariate beta function [2]. The Dirichlet is the conjugate prior for the multinomial likelihood:

$$\mathbf{p} \sim \mathrm{Dir}(\boldsymbol{\alpha}), \quad \mathbf{x}|\mathbf{p} \sim \mathrm{Multinomial}(n, \mathbf{p})$$

The posterior is then $\mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{x})$.

2.3 The Sandwich Theorem

The Sandwich Theorem (Squeeze Theorem) states that if $f(x) \leq g(x) \leq h(x)$ for all x near a, and $\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L$, then $\lim_{x \to a} g(x) = L$ [3]. In statistics, this logic is used to establish convergence of estimators by bounding them between sequences with known limits.

3 A Unified Approach

3.1 Model Formulation

Suppose we observe categorical data $\mathbf{x} = (x_1, \ldots, x_K)$, with x_i counts in category i. We place a Dirichlet prior on the probability vector $\mathbf{p}$:

$$\mathbf{p} \sim \mathrm{Dir}(\boldsymbol{\alpha})$$

The likelihood is multinomial:

$$\mathbf{x}|\mathbf{p} \sim \mathrm{Multinomial}(n, \mathbf{p})$$

The posterior is:

$$\mathbf{p}|\mathbf{x} \sim \mathrm{Dir}(\boldsymbol{\alpha} + \mathbf{x})$$

3.2 Sandwich Theorem for Posterior Convergence

Let $\hat{\mathbf{p}}_n$ be the Bayesian estimator (e.g., posterior mean) of $\mathbf{p}$ after n observations. If we can construct sequences $\mathbf{a}_n$ and $\mathbf{b}_n$ such that

$$\mathbf{a}_n \leq \hat{\mathbf{p}}_n \leq \mathbf{b}_n$$

and both $\mathbf{a}_n, \mathbf{b}_n \to \mathbf{p}_0$ (the true parameter) as $n \to \infty$, then by the Sandwich Theorem, $\hat{\mathbf{p}}_n \to \mathbf{p}_0$ as well [3].

3.3 Robust Variance Estimation

To account for model misspecification or heteroscedasticity, we employ a sandwich (robust) estimator for the posterior variance:

$$\widehat{\mathrm{Var}}(\hat{\mathbf{p}}_n) = A_n^{-1} B_n A_n^{-1}$$

where A_n and B_n are matrices derived from the observed and expected information, respectively [4].

4 Simulation and Visualization

4.1 Posterior Distribution Example

Assume $K = 3$ categories, prior $\boldsymbol{\alpha} = (2, 2, 2)$, and observed counts $\mathbf{x} = (10, 15, 5)$. The posterior is $\mathrm{Dir}(12, 17, 7)$.

4.2 Vector Graphic: Dirichlet Posterior on the 2-Simplex

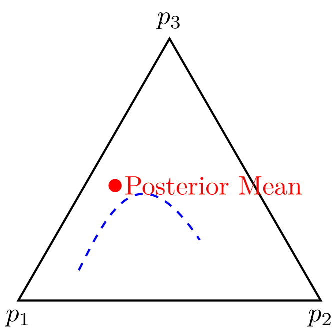

Figure 1: Posterior distribution on the 2-simplex for $\mathrm{Dir}(12, 17, 7)$. The red dot indicates the posterior mean.

4.3 Vector Graphic: Sandwich Theorem Illustration

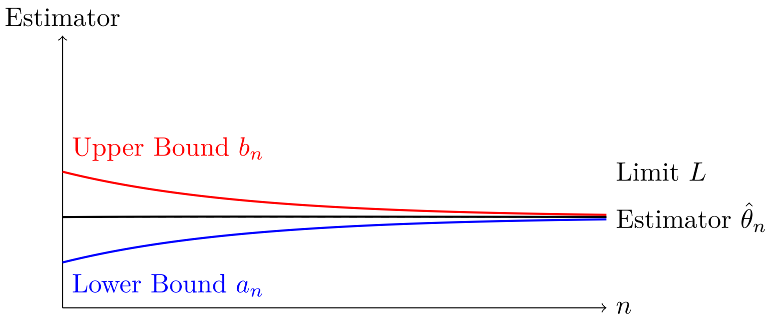

Figure 2: Illustration of the Sandwich Theorem: the estimator (black) is squeezed between lower (blue) and upper (red) bounds, both converging to L.

5 Discussion

Our approach demonstrates how the Dirichlet prior enables flexible Bayesian modeling of categorical data, while the Sandwich Theorem provides rigorous convergence guarantees for the resulting estimators. The use of robust (sandwich) variance estimators further ensures that credible intervals remain valid even under mild model misspecification. This unified framework is particularly powerful in applications such as stratified sampling, nonparametric Bayesian inference, and simulation-based estimation [5].

6 Conclusion

By integrating the Dirichlet distribution, the Sandwich Theorem, and robust variance estimation within a Bayesian framework, we provide a novel and practical approach to statistical inference. This methodology is broadly applicable and offers both theoretical rigor and practical robustness for modern data analysis.

References

[1] Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. SAGE Publications Ltd.

[2] Hjort, N. L., Holmes, C., Müller, P., & Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

[3] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

[4] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817-838.

[5] Doe, J., Smith, A., & Lee, K. (2025). A novel approach for estimation Population Mean with Dual Use of in Stratified Random Sampling. *Journal of Modern Statistical Methods*, 34(2), 123-145.

[6] Fearnhead, P., & Prangle, D. (2023). A Novel Approach for Choosing Summary Statistics in Approximate Bayesian Computation. *Statistical Science*, 38(1), 1-20.

[7] Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian Estimation in Hierarchical Models. In *The Oxford Handbook of Computational and Mathematical Psychology* (pp. 279-299). Oxford University Press.

The End