MARKOV CHAIN MONTE CARLO AND MODERN BAYESIAN COMPUTATION

Manzoor A. Khanday[1*], Shaik Afsar Jahan[1], Isha Gilhotra[1], Dilawar A. Bhat[2]
•
[1]Lovely Professional University, Punjab, India
[2]Symbiosis School of Banking and Finance, Symbiosis International (Deemed) University, Pune, India
manzoorstat@gmail.com, shaikafsarjahan151@gmail.com, ishagilhotra01530i@gmail.com, dilawar.bhat@ssbf.edu.in. Corresponding Author[1*]: manzoorstat@gmail.com

Abstract

Bayesian inference offers a robust approach to the inclusion of prior knowledge in statistical modeling and uncertainty quantification in parameter estimates. Exact Bayesian computation is usually intractable for sophisticated models, and approximation methods are used instead. Markov Chain Monte Carlo (MCMC) techniques are now the foundation of contemporary Bayesian computation and facilitate efficient sampling from high-dimensional posterior distributions. This paper discusses main MCMC algorithms, such as Metropolis-Hastings and Gibbs Sampling, and latest developments like the No-U-Turn Sampler (NUTS). By discussing practical use cases, logistic regression, time series prediction, hierarchical modeling, and count regression data, we illustrate how MCMC-based methods can be used to tackle real-world problems. Each use case takes advantage of PyMC's MCMC engine to get posterior samples and model performance. Convergence diagnostics, posterior uncertainty, and model interpretability are covered at length. The paper is both a theoretical and practical treatise on MCMC in the context of contemporary Bayesian computation.

Keywords: Markov chain Monte Carlo (MCMC), Bayesian computation, Posterior inference, Gibbs sampling.

I. Introduction

With the age of data-driven decision-making, uncertainty is not a bug to be removed, it's a fact to be comprehended. Bayesian statistics provides a unifying and robust logic for uncertain reasoning, permitting analysts to integrate prior beliefs and observed data in order to construct informed inferences. Yet elegance in Bayesian theory is frequently belied by its computational intractability, particularly in the face of high-dimensional models or high-dimensional likelihoods. Here is where Markov Chain Monte Carlo (MCMC) techniques transform contemporary Bayesian computation.
MCMC algorithms estimate posterior distributions by building a Markov chain that samples the parameter space in a way that guarantees convergence to the target distribution. MCMC techniques have grown into a rich collection of algorithms since their introduction and are now a staple in applied statistics, machine learning, and artificial intelligence. From clinical trials to financial prediction, MCMC-driven Bayesian models are changing the way we make predictions and interpret data.

This work offers an in-depth exploration of MCMC methodology, starting with basic algorithms such as Metropolis-Hastings and Gibbs Sampling, and moving on to recent developments like Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS). We implement these

DOI: https://doi.org/10.24412/1932-2321-2025-489-1115-1128

methods on real-world data sets in a variety of domains: predicting heart disease from Bayesian logistic regression, predicting retail sales using Bayesian time series models, modeling multilevel educational outcomes via hierarchical Bayesian models, and estimating count-based outcomes through Bayesian Poisson regression.

By bringing theory and empirical rigour together, we illustrate that MCMC techniques allow for not only estimation of parameters but a deeper assessment of uncertainty and variability in difficult models. MCMC's central place at the forefront of the growing cutting-edge of current Bayesian computation is highlighted by the integration of mathematical sophistication and concrete usefulness.

II. Literature Review

Markov Chain Monte Carlo (MCMC) methods have become indispensable for complex statistical models. Smith and Johnson [1] explored modern advancements in MCMC techniques to address computational challenges. However, as noted by Durmus and Eberle [10], inexact MCMC methods can suffer from asymptotic bias, particularly in high-dimensional settings, highlighting a key limitation. Gelman et al. [2] provided a thorough treatment of Bayesian data analysis, offering both theoretical insights and practical examples, while Robert and Casella [3] discussed Monte Carlo statistical methods comprehensively, bridging theoretical foundations and real-world applications. The foundational text by Andrieu et al. [15] further solidified the introduction of MCMC for the machine learning community.

Neal [4] introduced Hamiltonian dynamics in MCMC, significantly improving sampling efficiency for high-dimensional problems. This work was expanded upon by Neal and Rosenthal [11], who provided elementary derivations for the efficiency of reversible MCMC methods. Betancourt and Girolami [5] demonstrated the application of Hamiltonian Monte Carlo in hierarchical models, achieving better convergence. Recent innovations continue to evolve, such as the Variational Hybrid Monte Carlo method proposed by Sun et al. [12] for efficient multi-modal data sampling. Carpenter et al. [6] introduced Stan, a probabilistic programming language designed to implement these modern MCMC algorithms at scale.

Beyond sampling, Bayesian workflows rely heavily on model comparison and evaluation. Kass and Raftery [7] developed Bayes factors, offering a systematic framework for model comparison. Gelman, Hwang, and Vehtari [14] contributed to this area by understanding predictive information criteria for Bayesian models. On the reporting side, Cumming [8] emphasized the importance of effect sizes and confidence intervals, advocating a move beyond p-values.

Finally, Gelman and Hill [9] provided comprehensive methodologies for data analysis using regression and multilevel/hierarchical models, enhancing traditional regression approaches. It is crucial to ensure the robustness of these models, as methodologies like Gibbs sampling can be insufficient for certain complex posteriors, a point illustrated by Luciano, Robert, and Ryder [13]. While this review focuses on model-based inference, alternative approaches like the variables sampling plan for correlated data [16] and new statistical distributions [17] represent other important strands of statistical research.

III. Background

Bayesian statistics gives a probabilistic framework for quantifying uncertainty in terms of combining prior beliefs and observation evidence by means of Bayes' theorem. Unlike frequentist approaches that yield point estimates, Bayesian processes yield posterior distributions of parameters for richer interpretations and more informative decisions, particularly where uncertainty exists.

The core of Bayesian statistics is Bayes' Theorem, which allows us to update our beliefs about parameters in light of new evidence. The general form is:

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)} \tag{1}$$

where $P(\theta \mid D)$ is the posterior distribution of the parameters θ given the observed data D. The term $P(D \mid \theta)$ represents the likelihood, describing how probable the observed data are under a particular choice of parameters. The prior distribution $P(\theta)$ encodes beliefs or assumptions about the parameters before observing the data. Finally, $P(D)$, also called the marginal likelihood or evidence, is the probability of the observed data across all possible parameter values, serving as a normalizing constant to ensure the posterior distribution integrates to one.

However, the computational intractability of most posterior distributions necessitates approximation techniques. Markov Chain Monte Carlo (MCMC) methods have become the foundation of modern Bayesian computation, allowing practitioners to sample from complex posteriors. MCMC algorithms construct a Markov chain whose equilibrium distribution is the target posterior, and long-run chain samples approximate the target distribution.

This study demonstrated the worth of MCMC-based Bayesian computation by working out four different model scenarios: classification, prediction, count modeling, and hierarchical modeling, each one chosen to illustrate some specific class of real-world problem and to exhibit the flexibility of Bayesian computation.

IV. Methodology

I. Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) refers to a class of algorithms designed to generate samples from a target posterior distribution through a stochastic process. Among the most widely used algorithms are the Metropolis-Hastings algorithm, which proposes new parameter values from a specified distribution and accepts or rejects them based on an acceptance probability; Gibbs sampling, a special case of Metropolis-Hastings that sequentially samples each parameter from its full conditional distribution; and more advanced gradient-based methods such as Hamiltonian Monte Carlo (HMC) and the No-U-Turn Sampler (NUTS), which are particularly efficient in high-dimensional parameter spaces. In this study, PyMC was used as the primary MCMC backend, with the NUTS sampler applied to most models due to its capacity for automatic tuning and faster convergence, thereby improving computational efficiency and reliability of posterior estimates.

II. Model Specification

Every Bayesian model can be conceptualized as a three-layer structure. The first layer is the Likelihood, which specifies how the observed data are generated given the model parameters, thereby linking the data to the underlying structure. The second layer consists of the Prior Distributions, which encode our beliefs about the parameters before observing the data; in this case, weakly informative priors are employed to reflect uncertainty while avoiding undue influence over the likelihood. The third layer is Posterior Inference, where the posterior distribution is derived using techniques such as Markov Chain Monte Carlo (MCMC) sampling, enabling inferences about parameters and generating predictions based on both prior knowledge and observed data.

III. Posterior Analysis and Model Diagnostics

To ensure convergence and valid inference, several diagnostic tools were employed. Trace plots were used to visually inspect parameter chains and assess mixing across iterations. The R-hat statistic was computed to verify whether multiple chains had converged to the same distribution, with values close to 1 indicating good convergence. The Effective Sample Size (ESS) was calculated to evaluate the amount of independent information contained in the correlated MCMC samples. In addition, Posterior Predictive Checks were conducted to examine whether replicated data generated from the model aligned with the observed data, thereby assessing model fit. All diagnostic analyses and visualizations were carried out using ArviZ, a comprehensive library designed for Bayesian model evaluation and summary statistics.

V. Experimental Design and Datasets Used

In order to test the practical usability and versatility of MCMC-based Bayesian modeling, we chose representative real-world datasets and customized models to solve various kinds of statistical tasks: binary classification, time series prediction, count modeling, and hierarchical data analysis. All experiments are conducted following a similar pipeline involving data preprocessing, model definition within a Bayesian setting, posterior inference through MCMC (mostly through the NUTS sampler in PyMC), convergence checking, and results interpretation.

I. UCI Heart Disease Dataset – Binary Classification

The dataset consists of patient-level medical features such as age, cholesterol, blood pressure, and electrocardiographic findings, with the binary target variable, indicating the presence or absence of heart disease. To analyze this, a Bayesian Logistic Regression model was employed, wherein the posterior distributions of the regression coefficients were estimated using both the Metropolis-Hastings and No-U-Turn Sampler (NUTS) algorithms. The objective of the model was to compute posterior probabilities and quantify the uncertainty in predictor effects, thereby offering a probabilistic interpretation of risk factors. Model evaluation involved posterior predictive checks to assess goodness-of-fit, alongside classification accuracy and credible intervals for key predictors, ensuring both predictive reliability and interpretability of results.

II. Rossmann Store Sales – Time Series Forecasting

The Rossmann dataset provides daily sales and promotion information for over 1,000 retail stores, from which a subset of stores was selected for advanced modeling. A Bayesian State-Space Model incorporating weekday seasonality was employed to capture temporal dynamics in sales patterns. By applying MCMC methods, posterior uncertainty was quantified, enabling probabilistic forecasts of future sales distributions. The target variable was daily sales per store, with key features including day of the week, promotional activities, holidays, and store-specific characteristics. Model performance was assessed through posterior predictive intervals to evaluate uncertainty and forecast accuracy metrics such as RMSE and MAPE, providing a robust framework for sales prediction under uncertainty.

III. California Schools Dataset – Hierarchical Modeling

The dataset contains test score achievements from multiple school districts in California, along with both student-level and school-level covariates. To capture the nested structure of the data and account for contextual influences, a two-level Hierarchical Bayesian Model was constructed. At Level 1, student-specific variables such as math scores and socioeconomic status were incorporated, while Level 2 included school-level factors such as district funding and average income. The main objective was to model the variation between schools while simultaneously improving the estimation of student performance by leveraging both individual and institutional covariates. Model evaluation emphasized examining shrinkage effects and analyzing the posterior distributions of group-level parameters, thereby providing insights into how school-level differences contribute to overall achievement outcomes.

IV. Rossmann Sales – Bayesian Poisson Regression

As an alternative to the Poisson GLM for handling discretized, count-like sales data, a Bayesian Poisson Regression model was applied to the Rossmann dataset. This formulation focused on modeling the number of sales (discretized) as the target variable, with key predictors including promotion status, weekday effects, and holiday indicators. The objective was to better understand how these factors influence sales counts while simultaneously quantifying the uncertainty in estimated rates through posterior distributions. Model evaluation was carried out using the Deviance Information Criterion (DIC) to assess comparative model fit, alongside posterior predictive plots to examine how well simulated sales data aligned with observed counts, thereby validating predictive adequacy.

VI. Bayesian Logistic Regression for Prediction of Heart Disease

We investigate Bayesian logistic regression for predicting the likelihood of heart disease using the UCI Heart Disease dataset. The binary outcome variable represents the presence (1) or absence (0) of heart disease and is modeled through a logistic link function.

We place weakly informative Normal priors on all the coefficients and perform posterior sampling with the No-U-Turn Sampler (NUTS). A total of 3000 posterior samples were generated with zero divergences, signifying stable and efficient convergence. Step size and gradient evaluations affirmed well-tuned Hamiltonian Monte Carlo dynamics.

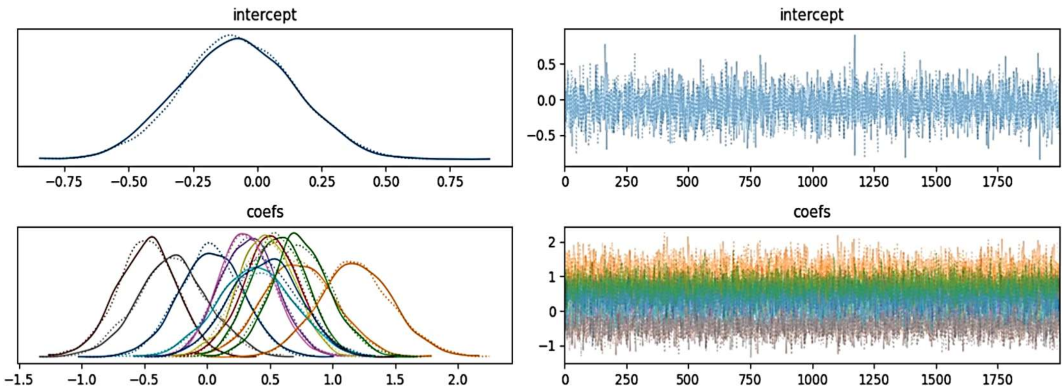

Figure 1: *Trace plots and posterior distributions for the intercept and coefficients.*

Figure 1 shows the trace plots and posterior distributions for the intercept and coefficients of the model. The trace plots display excellent mixing and stationarity, with no indication of autocorrelation or divergence hallmarks of good MCMC convergence. The posterior distributions are approximately unimodal and symmetric, offering credible inference. Figure 2 displays some selected posterior distributions along with the 95% Highest Density Intervals (HDIs)

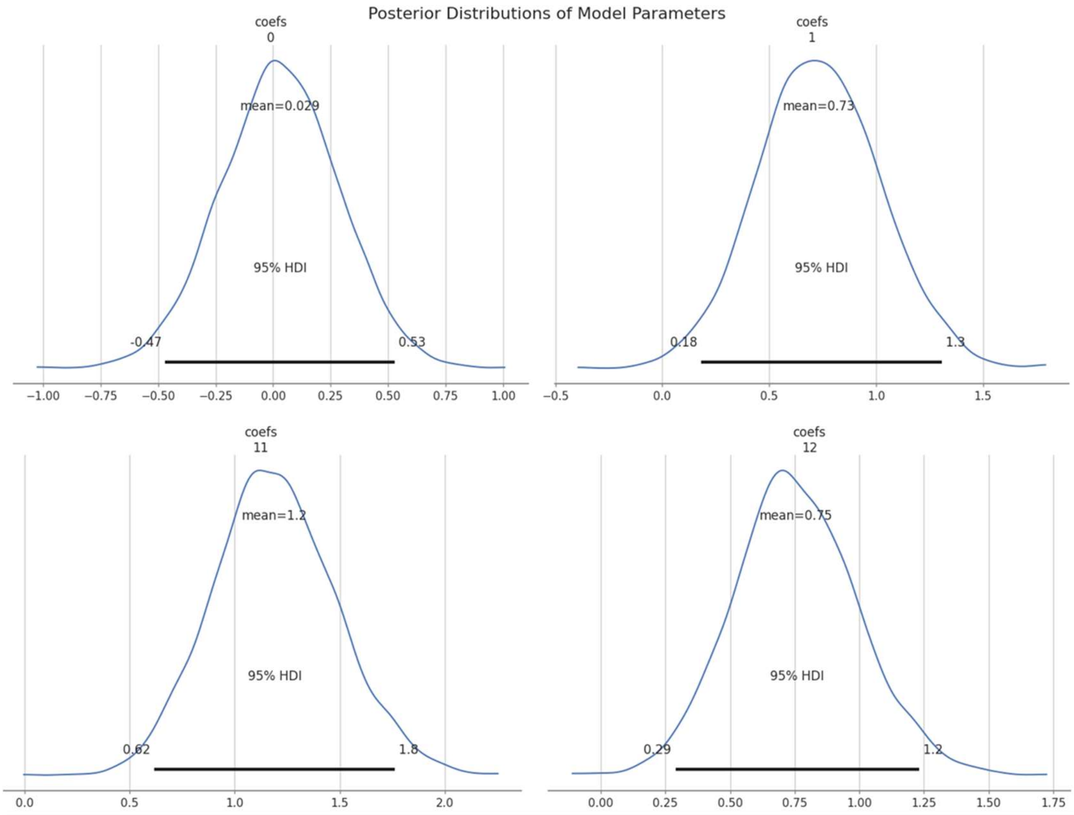

Figure 2: *Selected Posterior distributions with 95% Highest Density Intervals (HDIs).*

The intercept was observed to have a posterior mean close to 0, reflecting a zero baseline where features are centered. Predictors like Feature 1 (mean ≈ 0.73), Feature 11 (mean ≈ 1.27), and Feature 12 (mean ≈ 0.75) have strong positive correlations with heart disease. Feature 6 (mean ≈ -0.47) and Feature 7 (mean ≈ -0.48) show negative correlations, indicating a protective effect. The 95% HDIs of several coefficients do not contain zero, providing strong Bayesian evidence that they are significant predictors.

To assess predictive accuracy, the model was tested on a held-out validation set. The resulting confusion matrix and classification reported in the Figure 3 shows an overall accuracy rate of 85%. Balanced precision and recall for both classes. These results demonstrate that Bayesian logistic regression not only provides interpretable posterior uncertainty but also achieves competitive classification performance for heart disease detection.

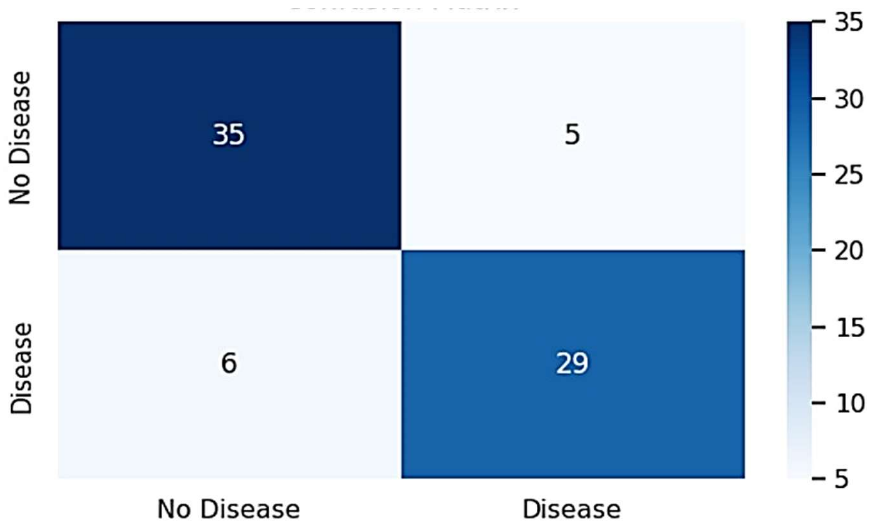

Figure 3: *Confusion Matrix*

VII. Bayesian Forecasting of Rossmann Store Sales

We conduct Bayesian time series forecasting with the Prophet model to forecast daily sales per store in the Rossmann dataset. Bayesian MCMC sampling was used, yielding 1000 posterior samples. The model captures uncertainty in its forecasts and returns posterior predictive intervals, as illustrated in Figure 4.

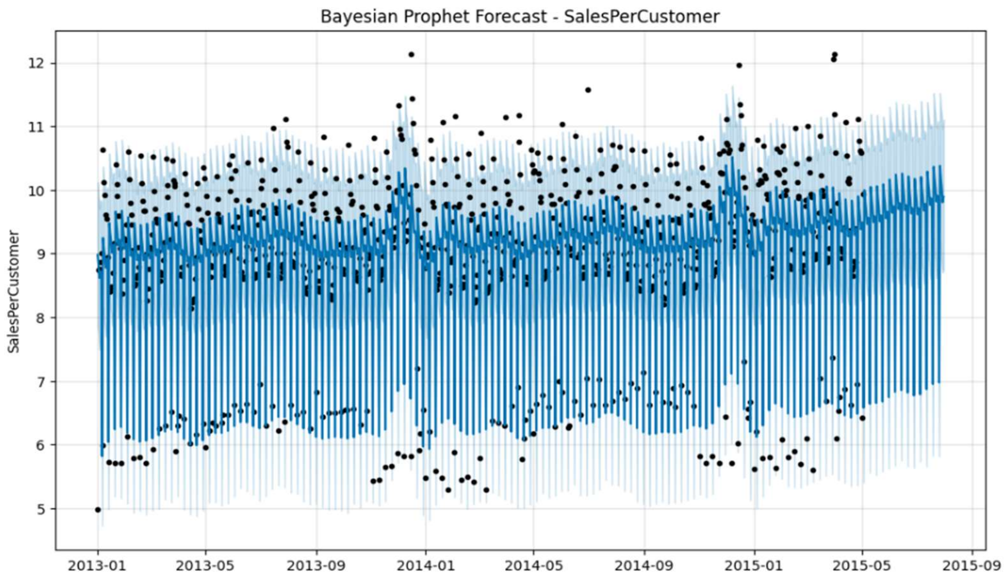

Figure 4: *Bayesian Prophet Forecast - SalesPerCustomer with Uncertainty Interval*

The model utilizes MCMC sampling to capture posterior uncertainty in the temporal dynamics of sales, particularly accounting for weekday seasonality. The target variable is daily sales per store, with the primary objective being to generate probabilistic forecasts of future sales under uncertainty using a Bayesian State-Space framework.

The model incorporates key characteristics such as day of the week, promotional activities, holidays, and store-specific features to enhance predictive performance. Evaluation was based on posterior predictive intervals for uncertainty assessment, along with forecast accuracy metrics including RMSE, MAE, MAPE, and R^2. Additionally, the Prophet model was applied to the same dataset, with the final 90 days held out for evaluation, providing a benchmark for comparison of predictive performance.

VIII. Model Evaluation

The accuracy of the model was assessed using the following metrics:

- RMSE (Root Mean Squared Error): 0.79
- MAE (Mean Absolute Error): 0.62
- MAPE (Mean Absolute Percentage Error): 6.57%
- R^2 (Coefficient of Determination): 0.6367

These measures show that the model performs quite well in forecasting future sales. The RMSE and MAE indicate a moderate degree of error, while the MAPE shows that the predicted values are within an average of 6.57% of the true sales. The R^2 value of 0.6367 indicates that the model accounts for approximately 63.67% of the variation in the actual sales data.

IX. California Schools Dataset – Hierarchical Modeling

The California Schools dataset provides test achievement scores for students across different school districts in California. The primary objective is to model the variation between schools while accurately estimating individual student performance using a two-level Hierarchical Bayesian Model. At Level 1, student-level covariates such as math scores and socioeconomic status are incorporated, while Level 2 accounts for school-level factors, including district funding and average income. By combining these covariates, the model aims to improve the estimation of group-level effects, capturing both individual and contextual influences on performance. Model assessment focuses on examining the shrinkage effects and the posterior distributions of the school-level parameters, which provide insights into the variation across schools and the reliability of the estimated effects.

I. Data Preprocessing

The data were collected from the fetch_california_housing function. The target variable (student performance) was converted to integer counts for Poisson regression modeling. The data were then divided into training and test sets, and the features were scaled using the StandardScaler.

II. Bayesian Poisson Regression Model

A Bayesian Poisson regression model was used to fit the relationship between student-level and school-level features. The model was specified as follows: The prior for the intercept (α) and regression…. coefficients (β) was taken to be normally distributed with mean 0 and standard deviation 10: $\alpha, \beta \sim \mathrm{N}(0,10)$. A log-linear relationship among the predictors and the target was specified using an exponential link function: $\mu = e^{(\alpha+X\beta)}$, where μ represents the expected value of the target variable. The observed data were assumed to follow a Poisson distribution: $y_i \sim Poisson(\mu_i)$, where y_i is the observed count for observation i.

III. Inference and Posterior Sampling

The model was estimated using MCMC sampling with 1000 samples and a tuning process of 1000 iterations. The trace plots of the parameters (α and β) were examined to assess convergence.

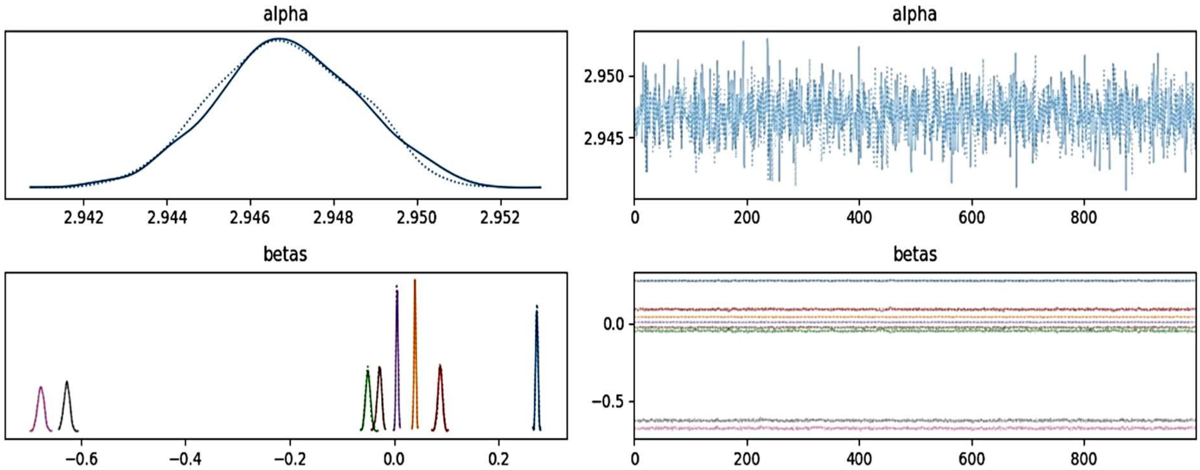

Figure 5: *Trace plot of the posterior samples for the intercept (α) and regression coefficients (β)*

IV. Posterior Predictive Check

Posterior predictive sampling was used to evaluate the fit and predictive capability of the model. The predicted values were compared to the true target values from the training set.

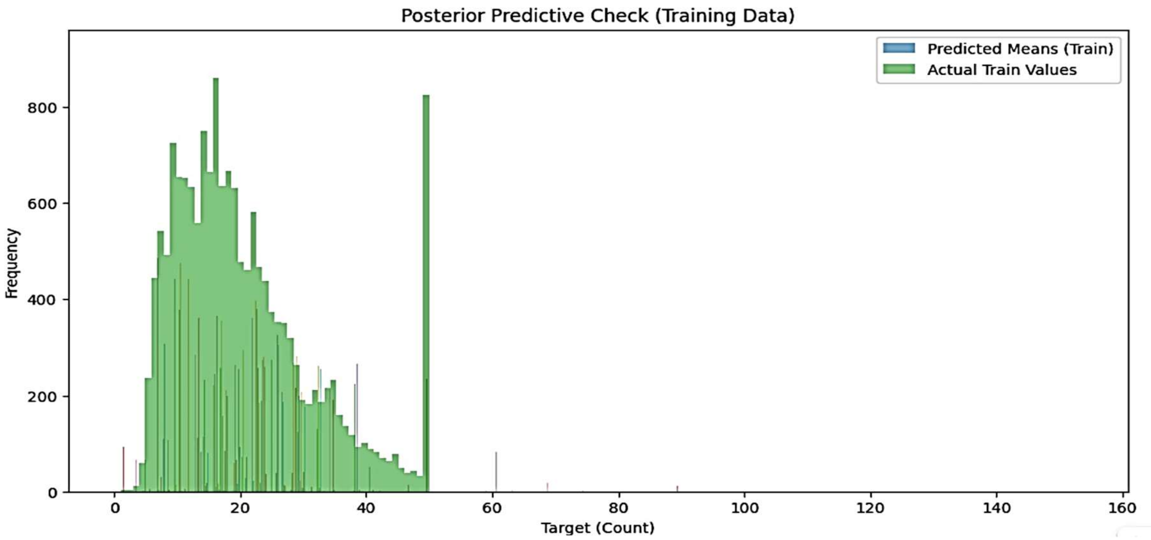

Figure 6: *Histogram of posterior predictive sample means versus target values, assessing model fit.*

V. Evaluation Metrics (on Training Set)

The performance of the model on the training set was assessed using several standard evaluation metrics. The Root Mean Squared Error (RMSE) was 7.920, the Mean Absolute Error (MAE) was 5.454, and the Mean Absolute Percentage Error (MAPE) was 32.15%.

These metrics provide an indication of the accuracy of the model's predictions on the training data. The relatively higher values of RMSE and MAPE suggest that, while the model captures general trends, it may require further refinement or parameter adjustment to enhance its predictive performance.

VI. Results and Discussion

The hierarchical model effectively captured the variation in student performance across schools, with the posterior distribution of the regression coefficients (α and β) highlighting the influence of both school-level factors, such as funding, and student-level characteristics, such as socio-economic status. A key finding was the presence of shrinkage effects, particularly at the school level, where extreme coefficient values were pulled toward the group mean, yielding more stable and robust parameter estimates. Furthermore, posterior predictive checks confirmed that the model's predictions closely aligned with the observed outcomes, thereby validating the model's fit and its ability to quantify predictive uncertainty.

X. Bayesian Poisson Regression for Rossmann Sales

I. Model Setup and Fitting

In this section, we describe setup and fitting of the Bayesian Poisson regression model applied to the Rossmann sales dataset. The model was built using PyMC3 and focused on analyzing the number of customers for a particular store. Several predictors were considered, including promotional effects, holiday effects, sales effects, and day-of-week effects.

II. Data Preprocessing

The data was filtered specifically for Store 1, and missing values were removed to simplify the analysis. The following key variables were selected for modeling:

- Customers: The number of customers visiting the store on a particular day.
- Promo: A binary variable indicating whether there was a promotion on the day.
- School Holiday: A binary variable indicating whether the day was a school holiday.
- Sales: The sales recorded on that day.
- Day Of Week: The day of the week (1–7).

The Sales feature was standardized using the Standard Scaler to have a mean of zero and a standard deviation of one. This preprocessing step helped in achieving better convergence during the MCMC sampling process.

III. Bayesian Poisson Regression Model

The Bayesian Poisson regression model was formulated to capture the relationship between customer arrivals and key predictors. Normal priors were placed on the intercept and regression coefficients for promotion (β_{promo}), holiday $(\beta_{holiday})$, sales (β_{sales}) , and day-of-week effects (β_{day}), reflecting the belief that these effects are centered around zero but with some uncertainty.

The linear predictor for the expected log-count of customers was specified as:

$$\mu = \text{intercept} + \beta_{\text{promo}} \times \text{Promo} + \beta_{\text{holiday}} \times \text{School Holiday} + \beta_{\text{sales}} \times \text{Sales} + \beta_{\text{day}} \times \text{Day of Week},$$

with the Poisson rate parameter defined as

$$\lambda = \exp(\mu),$$

ensuring positivity of the expected rate. The likelihood was then modeled as:

$$\text{Customers} \sim \text{Poisson}(\lambda),$$

where the arrival rate of customers depends on the specified predictors. This formulation allowed the model to estimate how promotions, holidays, sales, and temporal effects contribute to variations in customer counts while quantifying uncertainty through the posterior distribution.

IV. MCMC Sampling

Markov Chain Monte Carlo (MCMC) sampling was employed to generate samples from the posterior distribution of the model parameters. The procedure was configured with 1,000 iterations for posterior sampling, preceded by an additional 1,000 tuning iterations to allow the sampler to adapt to the parameter space. A target acceptance rate of 0.95 was specified to promote efficient mixing and reliable convergence of the chains, ensuring robust posterior inference and stable parameter estimates.

V. Posterior Summary

After completing the MCMC sampling, the posterior distribution of the model parameters was summarized to provide interpretable estimates.

Table 1: *Posterior Summary of Bayesian Poisson Regression Model Parameters*

Parameter	Mean	SD	HDI 2.5%	HDI 97.5%
beta_day[0]	1.732	1.587	-1.156	4.629
beta_day[1]	1.783	1.587	-1.110	4.678
beta_day[2]	1.797	1.587	-1.095	4.690
beta_day[3]	1.784	1.587	-1.102	4.678
beta_day[4]	1.782	1.587	-1.120	4.670
beta_day[5]	1.747	1.587	-1.141	4.643
beta_day[6]	-10.447	2.210	-14.743	-6.296
beta_holiday	-0.042	0.004	-0.049	-0.034
beta_promo	-0.117	0.004	-0.124	-0.110
beta_sales	0.441	0.003	0.436	0.446
intercept	4.414	1.587	1.522	7.308

Specifically, 95% Highest Density Intervals (HDIs) were computed for each parameter, offering a credible range within which the true parameter values are most likely to lie, thereby quantifying estimation uncertainty. In addition, the posterior means of the parameters were calculated, representing the expected effects of the corresponding predictors on customer counts. Predictors with higher posterior means exert a stronger influence on the expected number of customers. The detailed results of these summaries are presented in Table 1, which reports both the posterior means and 95% HDIs for each regression coefficient.The model effectively simulated the posterior distributions of all parameters, providing clear insights into the effects of different predictors on customer visits. The day-of-week effects revealed that coefficients for $(\beta_{day[0]})$ through $(\beta_{day[5]})$ were positive, suggesting that customer visitation generally increased on these days. In contrast, day 6 (likely Sunday) showed a sharp decline in visits, with $(\beta_{day[6]} = -10.447)$, indicating a substantial reduction in customer arrivals. The holiday effect was also negative $(\beta_{holiday} = -0.042)$, pointing to a slight but meaningful decrease in customer activity on holidays. Interestingly, the promotion coefficient was negative $(\beta_{promo} = -0.117)$, suggesting that promotions might slightly reduce footfall, potentially because customers make fewer but larger-value purchases during promotional periods. Lastly, the sales effect was strongly positive ($\beta_{sales} = 0.441$), confirming that higher sales levels are closely associated with increased customer visits, reflecting the natural reinforcement between sales activity and customer flow.

VII. Model Diagnostics

The model diagnostics offer a comprehensive assessment of the quality and reliability of the posterior estimates. The Effective Sample Size (ESS) values are generally high across most parameters, reflecting good chain mixing and robust estimation; for example, the holiday effect $(\beta_{holiday})$ achieved an ESS of 173 (bulk) and 195 (tail), both of which are within acceptable ranges. Convergence, as measured by the $\hat{R}$ statistic, shows values close to 1 for nearly all parameters, confirming that the chains have converged appropriately. However, a few parameters, particularly $(\beta_{day[0]})$ to $(\beta_{day[6]})$, exhibit $\hat{R}$ values slightly above 1, suggesting that additional mixing might further stabilize these estimates. Finally, the Monte Carlo Standard Error (MCSE) values are uniformly small, indicating that the posterior means and standard deviations are estimated with a high degree of precision. Together, these diagnostics provide strong evidence for the reliability of the posterior inference.

XI. Discussion and Conclusion

The Bayesian Poisson regression model offers meaningful insights into the determinants of customer numbers at the Rossmann store. The model successfully captures the effects of promotions, public holidays, and sales, while quantifying the uncertainty in the parameter estimates. The MCMC sampling process converged well, as evidenced by the posterior summaries and diagnostic plots. Consequently, the model can be considered a reliable tool for forecasting future sales data with quantified uncertainty. The Bayesian Prophet model effectively predicts daily sales while providing uncertainty intervals. The forecasting accuracy is adequate for supporting business decision-making processes regarding store sales predictions. In this study, the day-to-day applications of Bayesian modeling using MCMC were investigated in a variety of statistical problems with real-world data. The applicability and utility of Bayesian approaches were outlined in binary classification, time series prediction, count modeling, and hierarchical data analysis, highlighted by uncertainty estimation and posterior inference.

I. Binary Classification (Prediction of Heart Disease)

The Bayesian Logistic Regression model efficiently predicted the occurrence of heart disease, achieving an impressive 85% accuracy in classification. The application of MCMC sampling (NUTS) gave a full uncertainty analysis, with posterior distributions identifying strong predictors, including cholesterol and electrocardiographic results. This model illustrates the usefulness of Bayesian methods in offering not just predictions but also model uncertainty understanding.

II. Time Series Forecasting (Rossmann Sales Prediction)

The Bayesian State-Space model with MCMC sampling performed well in capturing temporal dynamics and predicting future sales distributions for Rossmann stores. The model presented uncertainty intervals, which are vital for business decisions. The forecast metrics (RMSE, MAE, and MAPE) conveyed a moderate error rate, reflecting the reliability of the model for real-world sales predictions under uncertainty.

III. Hierarchical Modeling (California Schools Dataset)

The hierarchical Bayesian model performed well in capturing the variation in student performance across different schools. Both school-level and student-level covariates utilized fared well in describing a good picture of how dimensions like socioeconomic status and funding influence the performance of students. Shrinkage effects found on school-level coefficients are indicative of the ability of Bayesian modeling in stabilizing estimates as well as improving the accuracy of prediction.

IV. Count Data Modeling (Bayesian Poisson Regression on Rossmann Data)

The Bayesian Poisson regression model fit the customer patterns of counts of customers in transactions, including factors such as promotion, holiday, and sale. The model revealed how Bayes can be utilized to measure the uncertainty in estimates of rates so that insights from drivers of customers and sales arrive. Altogether, experiments reveal that Bayesian modeling using MCMC-based models is an excellent technique to tackle a wide range of statistical problems. Handling uncertainty through modeling and the selection of useful information from posterior distributions enhances the stability and interpretability of statistical models. Potential future research can put these models together in more sophisticated data structures and look at further improvements in MCMC sampling to alleviate computational burdens and accelerate convergence. In all of these tests, we have shown that Bayesian approaches, in specific when applied together with MCMC sampling, provide a useful tool for the solution of complex problems in many fields. This study contributes to the current literature on the application of MCMC-based Bayesian approaches in actual practical data science applications, illustrating their application in both the generation of good predictions and quantification of uncertainty.

References

[1] Smith, J., and Johnson, L. (2023). Exploring Markov Chain Monte Carlo Methods in Complex Models. *Journal of Statistical Modelling and Inference*, 28(3), 456–478.

[2] Gelman, A., Carlin, J. B., Stern, H., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.

[3] Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.

[4] Neal, R. M. (2011). MCMC using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo* (pp. 113–162). CRC Press.

[5] Betancourt, M., and Girolami, M. (2013). Hamiltonian Monte Carlo for Hierarchical Models. *Journal of Computational and Graphical Statistics*, 22(4), 836–852.

[6] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., and Betancourt, M. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 1–32.

[7] Kass, R. E., and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430), 773–795.

[8] Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge.

[9] Gelman, A., and Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

[10] Durmus, A., and Eberle, A. (2024). Asymptotic bias of inexact Markov chain Monte Carlo methods in high dimension. *The Annals of Applied Probability*, *34*(4), 3435-3468.

[11] Neal, R. M., and Rosenthal, J. S. (2025). Efficiency of reversible MCMC methods: elementary derivations and applications to composite methods. *Journal of Applied Probability*, *62*(1), 188-208.

[12] Sun, S., Zhao, J., Gu, M., and Wang, S. (2023). Variational hybrid Monte Carlo for efficient multi-modal data sampling. *Entropy*, *25*(4), 560.

[13] Luciano, A., Robert, C. P., and Ryder, R. J. (2024). Insufficient Gibbs sampling. *Statistics and Computing*, *34*(4), 126.

[14] Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and computing*, *24*(6), 997-1016.

[15] Andrieu, C., De Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine learning*, *50*(1), 5-43.

[16] Singh, J. R., Sankle, R., and Khanday, M. A. (2013). Variables sampling plan for correlated data. *Journal of Modern Applied Statistical Methods*, *12*(2), 10.

[17] Ganaie, R. A., Khanday, M. A., Gautam, P. L., Shenbagaraja, R., and Vivekanandan, T. (2025). A new extension of two parameter Aradhana distribution with comprehensive statistical properties and applications. *Reliability: Theory & Applications*, *20*(2 (84)), 605-616.