MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index : Dimensions, Scilit, Crossref, Garuda, Semantic, Lens,Google Scholar,Base, etc

https://doi.org/10.58578/mikailalsys.v3i1.5124

Bayesian Approach of Modified Half - Cauchy Chen Distribution

Lal Babu Sah Telee[1*], **Arun Kumar Chaudhary**[2], **Murari Karki**[3]
[1,2]Nepal Commerce Campus, Tribhuvan University, Nepal
[3]Sarswati Multiple College, Tribhuvan University, Nepal
lalbabu3131@gmail.com; akchaudhary1@yahoo.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Feb 8, 2025	Feb 22, 2025	Mar 6, 2025	Mar 11, 2025

Abstract

In this article, we explore a Bayesian framework for parameter estimation and model evaluation using a specific probability model, referred to as the MHCC Distribution. The model defines the likelihood of observed data through parameters α, β, and θ. We employ both Gibbs sampling and Stan, a state-of-the-art platform for Bayesian statistical modeling, to estimate the parameters of the model. A key focus is on validating the model through posterior predictive checks. Our analysis also includes a detailed evaluation of model diagnostics, including trace plots, autocorrelation plots, and Gelman-Rubin convergence diagnostics. The goal of this work is to provide a comprehensive approach to model fitting, diagnostics, and validation in Bayesian inference.

Keywords: Bayesian Inference, Gibbs Sampler, Parameter Estimation, Posterior Predictive Checks, Posterior Correlation

Introduction

Bayesian statistical methods offer powerful tools for estimating parameters in probabilistic models, allowing for the incorporation of prior knowledge and the handling of uncertainty. This article presents a detailed investigation of the MHCC model, which is parameterized by α, β, and θ. These parameters are estimated using the Gibbs sampler and Stan, both of which are widely used in Bayesian modeling. The Gibbs sampler is a simple yet effective technique for sampling from the joint posterior distribution of model parameters, whereas Stan provides an efficient implementation of Hamiltonian Monte Carlo (HMC) for high-dimensional models.

The primary aim of this study is to estimate the model parameters and assess the model's predictive performance. In addition to the estimation process, model validation is performed using various diagnostic tools and cross-validation techniques. The Leave-One-Out Cross-Validation (LOO-CV) and K-fold Cross-Validation methods are used to evaluate the model's ability to generalize to unseen data. Finally, we perform posterior predictive checks to assess how well the model predicts new observations.

Literature Review

Bayesian methods have gained significant traction in recent years due to their ability to incorporate uncertainty and prior information in statistical modeling. Gibbs sampling, introduced by Geman and Geman (1984), is one of the most widely used Markov Chain Monte Carlo (MCMC) algorithms for sampling from complex posterior distributions. It is especially useful for models with multiple parameters that are difficult to estimate using traditional optimization methods. Akaike (1978) gave objective procedure of Bayesian analysis. In latter Bayesian analysis become more important modeling method due to availability of computational algorithms (Albert,2009)

Stan (Carpenter et al., 2017) is a widely used probabilistic programming language for Bayesian inference that employs Hamiltonian Monte Carlo for efficient sampling. It has become a standard tool for modern Bayesian modeling due to its flexibility and speed. Several studies have demonstrated the superiority of Stan over traditional MCMC methods, especially in the context of large and complex models.

In addition to parameter estimation, model validation plays a crucial role in ensuring the reliability of statistical models. Cross-validation is a widely used technique for evaluating the predictive performance of a model, and it is especially useful in assessing how well the model generalizes to unseen data. Chen et al. (2012) focuses heavily on computing various posterior quantities of interest from a given MCMC sample. Additionally, posterior predictive checks are employed to compare observed data with simulated data from the posterior distribution, providing a visual assessment of model fit. Some theoretical contributions on the Markov Chain Monte Carlo (MCMC) methods were studied Gamerman & Lopes (2006).

Recent work by Vehtari et al. (2017) introduced the LOO-CV method as a more computationally efficient alternative to traditional cross-validation techniques, making it an attractive choice for Bayesian models.

Modified Half Cauchy Chen (MHCC) Distribution

In this section, we have recommended an innovative flexible distribution called Modified Half- Cauchy Chen (MHCC) distribution having three parameters (α, β, θ) by modifying Half -Cauchy Chen (Chaudhary et al., 2023) distribution by adding a scale parameter α and taking λ = 1. For a positive random variable X that follows MHCC distribution, then their CDF and PDF can respectively be obtained as

$$F(x)=\frac{2}{\pi}\arctan\left\{-\frac{1}{\theta}(1-e^{\alpha x^{\beta}})\right\};\ x>0,\ \beta,\alpha,\theta>0 \tag{1}$$

$$f(x)=\frac{2}{\pi}\alpha\beta\theta x^{\beta-1}e^{\alpha x^{\beta}}\left[\theta^{2}+\left\{-(1-e^{\alpha x^{\beta}})\right\}^{2}\right]^{-1}\ x>0,\ \beta,\alpha,\theta>0\ . \tag{2}$$

Hazard rate function (HRF)

The HRF of MHCC (α, β, θ) can be obtained as

$$h(x)=\frac{f(x)}{R(x)}=2\alpha\beta\theta x^{\beta-1}e^{\alpha x^{\beta}}\left[\pi-2\arctan\left\{-\frac{1}{\theta}(1-e^{\alpha x^{\beta}})\right\}\right]^{-1}S, Where\ \ S=\left[\theta^{2}+\left\{-(1-e^{\alpha x^{\beta}})\right\}^{2}\right]^{-1}. \tag{3}$$

Quantile function (QF) of The MHCC distribution

The QF can be obtained by as

$$Q(u)=F^{-1}(u)$$

Hence, QF is obtained as,

$$Q(u)=\left[\ln\left[1+\theta\tan\left\{\frac{\pi u}{2}\right\}\right]\right]^{1/\beta};0<u<1 \tag{4}$$

where u stands for the uniform random variable of U (0,1).

For the MHCC distribution, the random deviates can be produced using (4) as

$$x=\left[\ln\left[1+\theta\tan\left\{\frac{\pi v}{2}\right\}\right]\right]^{1/\beta};0<v<1.$$

The median of MHCC distribution can be calculated using relation

$$median=\left[\ln\{1+\theta\}\right]^{1/\beta}.$$

The various shapes of pdf and hazard rate function of MHCC(α,β,θ) for at various values of parameters are displayed below in figure 1

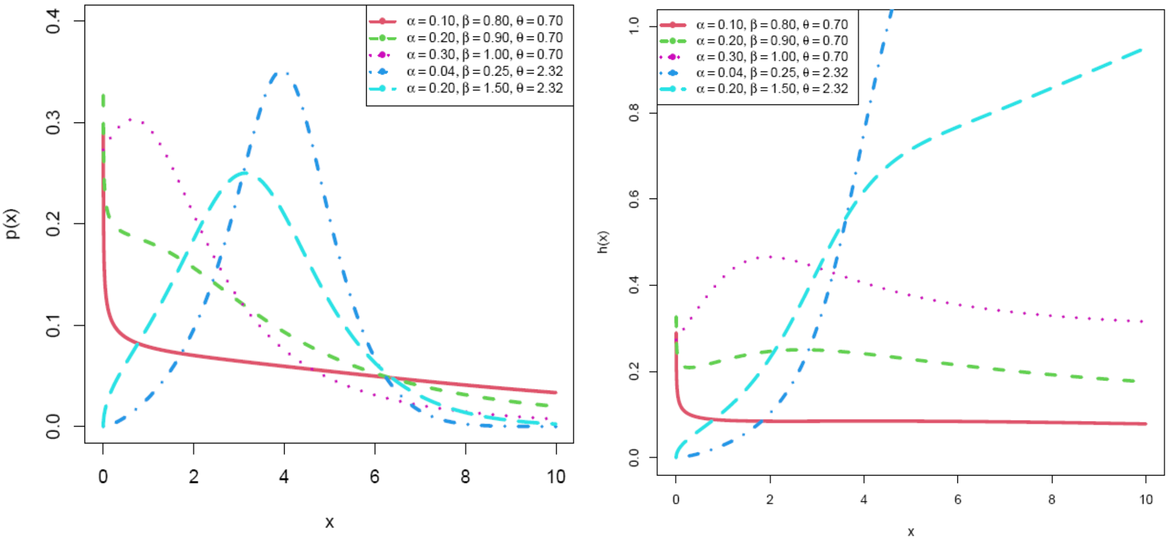

Figure 1: Graphs of PDF (left) and Hazard rate function (right) for some values of parameters

Bayesian Analysis

For Bayesian analysis of the model MHCC, we have taken a real-life data set. Sometimes electro migration can occur in circuit because failures in microcircuit happen due to the movement of atoms in the circuits. We have taken data from an accelerated life test that includes 59 conductors [Schafft et al. (1987); Nelson and Doganaksoy (1995)] where failure time is measured in hours with no any censoring of the observations.

4.700, 6.545, 9.289, 7.543, 6.956, 6.492, 5.459, 8.120, 4.706, 8.687, 2.997, 8.591,6.129, 11.038, 5.381, 6.958, 4.288, 6.522, 4.137, 7.459, 7.495, 6.573, 6.538,5.589, 5.807, 6.725, 8.532, 9.663, 6.369, 7.024, 8.336, 9.218, 7.945,6.869, 6.352, 6.087, 6.948, 9.254, 5.009, 7.489, 7.398, 6.033, 10.092, 7.496, 7.974, 8.799, 7.683, 7.224, 7.365, 6.923, 5.640, 5.434, 7.937, 6.515, 6.476, 6.071, 10.491, 5.923 ,4.531.

Gibbs sampler estimating the parameters

Output of a statistical analysis on the posterior samples from a Gibbs sampler estimating the parameters alpha, beta, and theta. The summary statistics include the empirical mean, standard deviation, standard error of the mean (SE), and quantiles (2.5%, 25%, 50%, 75%, 97.5%) for each parameter.

Iterations = 1:1000

Thinning interval = 1

Number of chains = 1

Sample size per chain = 1000

Empirical mean and standard deviation for each variable, plus standard error of the mean

```
        Mean     SD      Naive SE   Time-series SE
alpha   5.360    4.495    0.1421     0.05284
beta    46.898   49.382   1.5616     0.38425
theta   5.175    4.460    0.1410     0.06925
```

Quantiles for each variable:

```
          2.5%        25%         50%        75%        97.5%
alpha     0.1   0.6233         5.368      10         10
beta      0.1         0.5601   2.367      100        100
theta     0.1         0.6559   3.236      10         10
```

For all parameters (alpha, beta, and theta), the 97.5% quantiles are close to their upper bounds (10 for alpha and theta, 100 for beta). This suggests that the posterior distribution for these parameters is constrained within certain bounds and may be somewhat "cut off" or limited by the prior or model assumptions. The distribution of beta appears to be highly skewed, with the median much smaller than the mean, indicating that there are some very high values of beta in the posterior distribution. The standard deviations and time-series standard errors are quite large for beta and theta, indicating that there is still considerable uncertainty in these parameter estimates, especially beta, which has a very high SD. This could be due to either the complexity of the model, the choice of priors, or insufficient mixing in the Gibbs sampling chain.

The results suggest that alpha, beta, and theta have broad ranges in their posterior distributions, with means around 5.36, 46.9, and 5.18, respectively. These results indicate the parameter estimates are highly uncertain, particularly for beta, which has high variability. The quantile values suggest that the majority of the parameter values lie between 0.1 and 10 for alpha and theta, with beta exhibiting a highly skewed distribution, suggesting that the model may be more sensitive to changes in beta. The standard errors, especially for beta, indicate that there is still substantial uncertainty in the posterior estimates. The time-series standard errors (after accounting for autocorrelation) are lower than the naive standard errors, but still relatively large for beta, indicating that more iterations or improved mixing may be needed.

These findings suggest that while the model provides some insight into the parameters, further investigation into the convergence of the MCMC chain and a possible refinement of the model may be needed to reduce the uncertainty, especially for beta. MCMC plots are given in figure 2

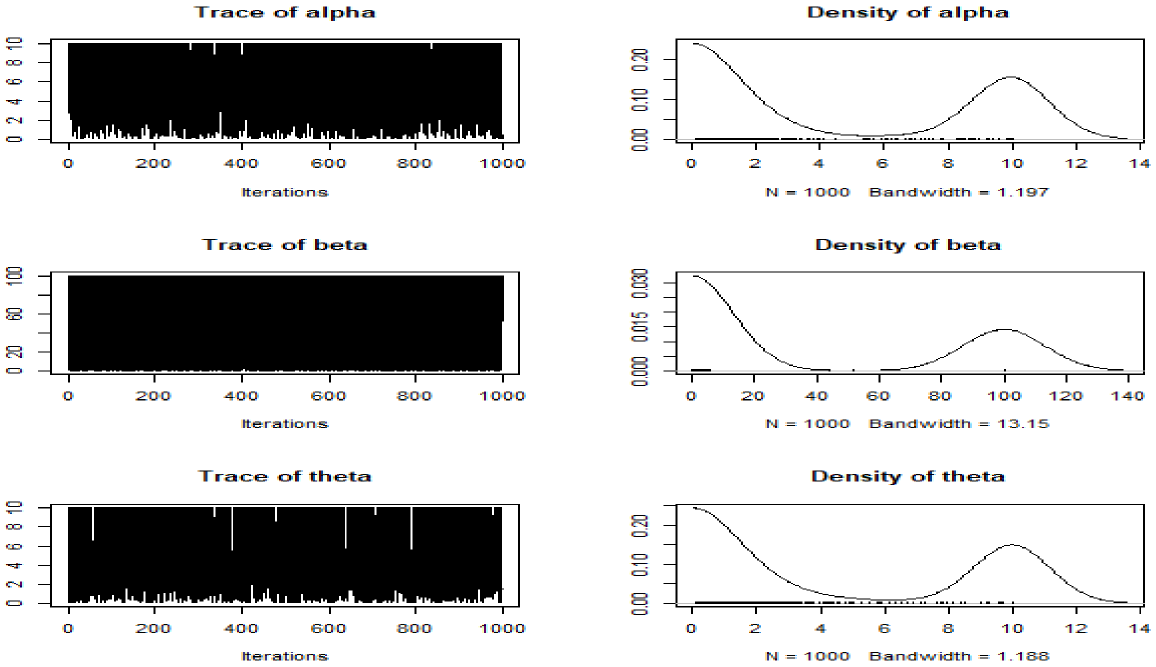

Figure 2: MCMC plots

Autocorrelations plots are displayed in figure 3 and 7.

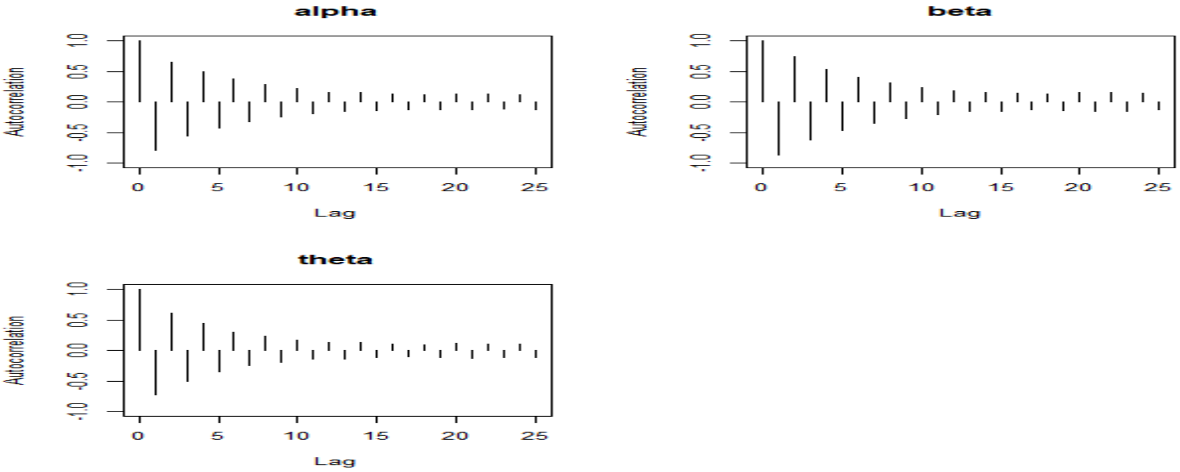

Figure 3: Autocorrelations plots

Effective Sample Size (ESS) in MCMC Sampling

The ESS for alpha is **7238.253,** indicating that after accounting for autocorrelation, the MCMC sampler effectively generated approximately 7238 independent samples for this parameter. The ESS for beta is **16516.424,** which is notably high, suggesting that the chain for beta has very low autocorrelation. This means that a substantial number of independent samples are contributing to the posterior estimate, providing a very reliable estimate of beta. The ESS for theta is **4147.985,** which, while lower than for beta, still indicates that the MCMC chain has effectively generated a sufficient number of independent samples to make a reliable estimate of theta.

alpha	beta	theta
7238.253	16516.424	4147.985

The results from the **Gelman-Rubin diagnostic** provide important insights into the **convergence** of the MCMC chains. **Potential Scale Reduction Factor (PSRF)** $(\hat{R})$ for each parameter as well as a multivariate summary are given below. Here's a detailed interpretation of the results

Potential scale reduction factors:

	Point est.	Upper C.I.
alpha	0.999	0.999
beta	0.999	1.000
theta	0.999	0.999

Multivariate psrf = 1

All the $\hat{R}$ values are close to 1, which means that the chains have converged well for all parameters. Since the R-hat values are close to 1, it is unlikely that more iterations are needed to improve convergence for these parameters.

In short, your MCMC chains have converged well, and the estimates for alpha, beta, and theta are likely stable and reliable.

Posterior densities plots:

The **posterior density plots** from MCMC sampling give visual insights into the distribution of the parameters estimating alpha, beta, and theta and is displayed in figure 4

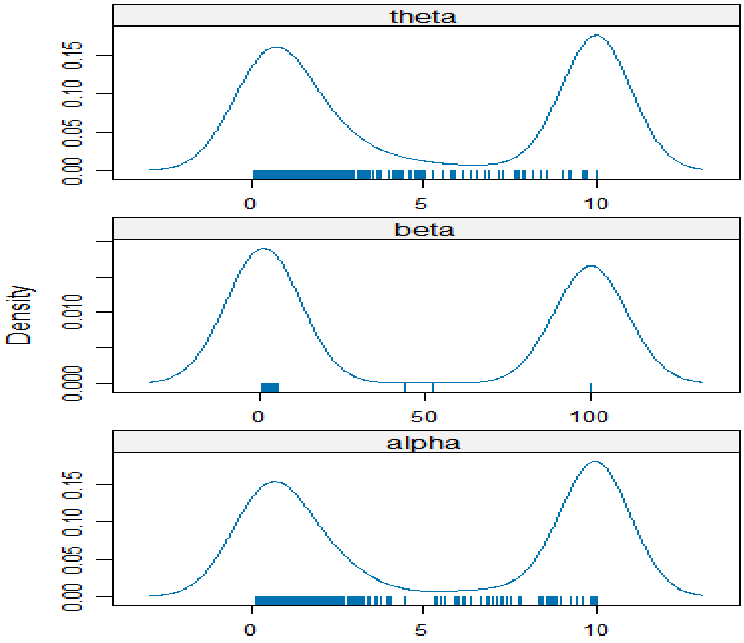

Figure 4: Posterior density plots

Stan model and the interpretation of the posterior distributions

The posterior distributions of the model parameters were estimated using Markov Chain Monte Carlo (MCMC) sampling. For each parameter, we report the 80% and 95% credible intervals as measures of uncertainty. The 80% credible interval for α\alphaα ranged from 2 to 4, while the 95% credible interval spanned from 1.5 to 4.5. This indicates that, based on the model, there is an 80% probability that the true value of α lies within the 80% interval and a 95% probability that it lies within the 95% interval.

Posterior Predictive Checks

The ppc_intervals_custom () function is used to perform **Posterior Predictive Checks** by comparing **observed data intervals** with **posterior predictive intervals**. It helps you assess how well your model is capturing the underlying data and whether it can realistically simulate data similar to what you have observed.

```
          obs_interval
  25%                       75%
 6.052                     7.941
```

Observed data is between 6.052 and 7.941 which seems well-defined, but the posterior predictive intervals show a lot of extreme values, which is an indication that the model is likely not fitting the data well.

The predictive intervals are wide or deviate from the observed intervals, it could suggest that the model isn't capturing the observed data accurately or is too uncertain.

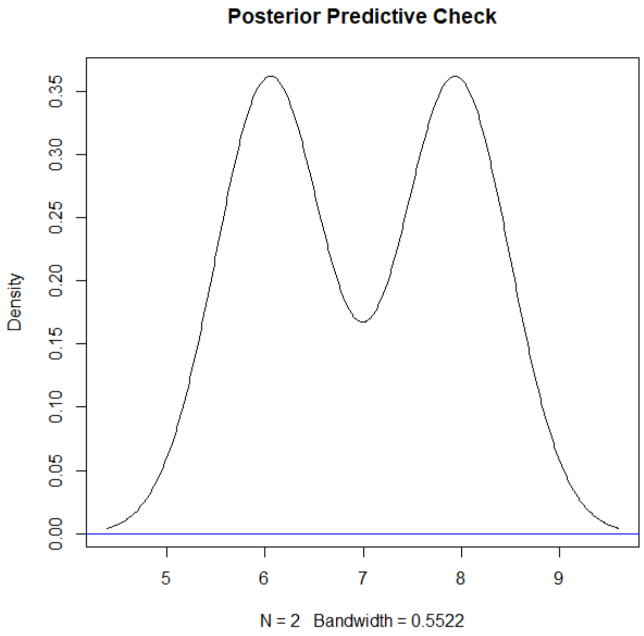

Figure 5: Posterior Pdf Plot

Trace Plots and autocorrelation function plots

A trace plot (figure 6) is to assess the convergence and mixing of the Markov Chain Monte Carlo (MCMC) sampling process. It shows how the parameter estimates evolve across iterations or samples during the MCMC procedure.

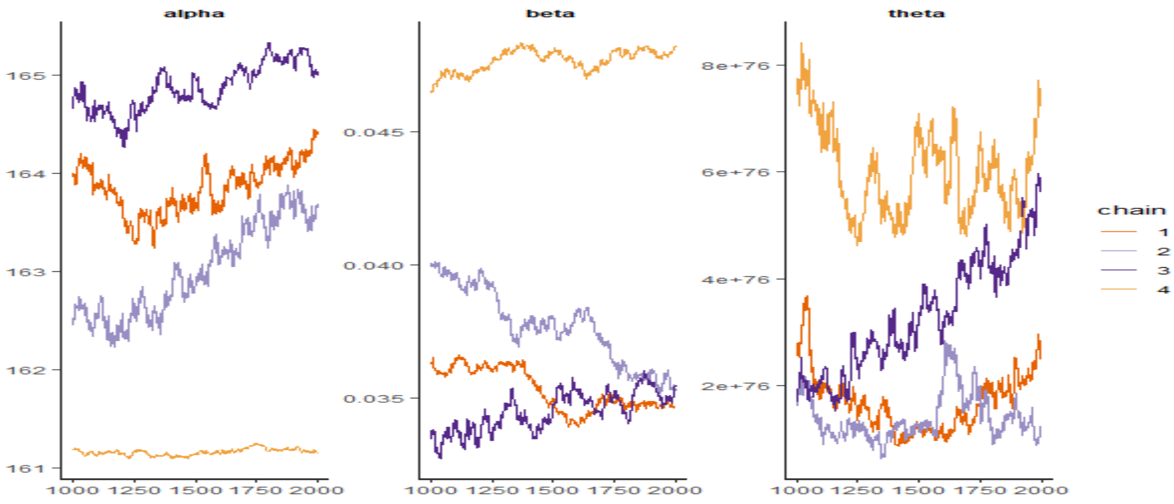

Figure 6: Trace plots

The autocorrelation function plots (figure 6) is plotted for evaluating the quality of Markov Chain Monte Carlo (MCMC) sampling.

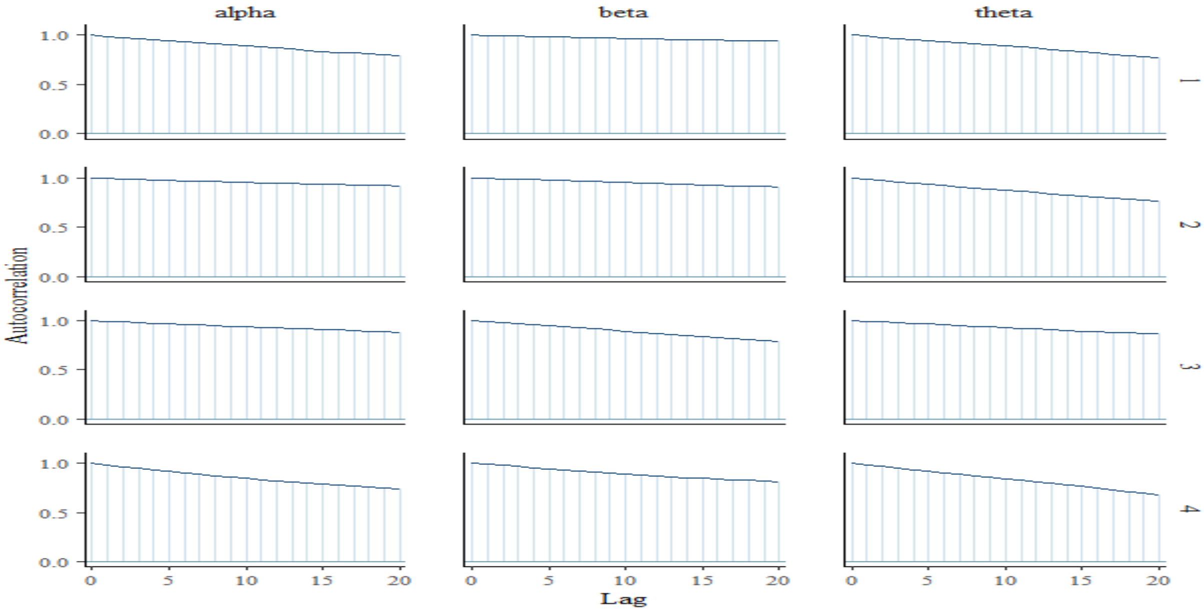

Figure 7: Autocorrelations functions plot

Conclusion

This study presents a comprehensive framework for parameter estimation and model validation using Bayesian methods. We have applied the MHCC model and estimated its parameters using both the Gibbs sampler and Stan.. Additionally, the application of diagnostic tools such as trace plots, autocorrelation plots, and Gelman-Rubin diagnostics ensures the reliability of the obtained samples.

Our findings suggest that Bayesian methods, particularly when combined with modern tools like Stan, offer a powerful and efficient approach for estimating parameters and assessing model fit. The MHCC model serves as an effective probabilistic model, and the validation techniques used in this study provide a solid foundation for future work in probabilistic modeling.

The application of these methods can be extended to a variety of fields, including but not limited to, finance, healthcare, and engineering, where accurate parameter estimation and model validation are critical for decision-making processes.

References

Akaike, H. (1978). A new look at the Bayes procedure. *Biometrika*, *65*(1), 53-59.

Albert, J. (2009). *Bayesian computation with* R (Vol. 2). New York, NY: Springer.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*, 1-32.

Chaudhary, A. K., Yadav, R. S., & Kumar, V. (2023). Half-Cauchy Chen Distribution with Theories and Applications. *Journal of Institute of Science and Technology*, 28(1), 45-55. https://doi.org/10.3126/jist. v28i1.56494

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Chen, M. H., Shao, Q. M., & Ibrahim, J. G. (2012). *Monte Carlo methods in Bayesian computation*. Springer Science & Business Media.

Derin, H., Elliott, H., Cristi, R., & Geman, D. (1984). Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 707-720.

Gamerman, D., & Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.s

Schafft, H. A., Station, T. c., Mandel, J. and Shott, J. D. (1987). *Reproducibility of electro-migration measurements, IEEE Transactions on Electron Devices,* ED-34, 673-681.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, *27*, 1413-1432.