Enhancing Computational Efficiency in High-Dimensional Bayesian Analysis: Applications to Cancer Genomics

Benjamin Osafo Agyare[1]

[1]Department of Statistics, University of Michigan, Ann Arbor.

Abstract

In this study, we present a comprehensive evaluation of the Two-Block Gibbs (2BG) sampler as a robust alternative to the traditional Three-Block Gibbs (3BG) sampler in Bayesian shrinkage models. Through extensive simulation studies, we demonstrate that the 2BG sampler exhibits superior computational efficiency and faster convergence rates, particularly in high-dimensional settings where the ratio of predictors to samples is large. We apply these findings to real-world data from the NCI-60 cancer cell panel, leveraging gene expression data to predict protein expression levels. Our analysis incorporates feature selection, identifying key genes that influence protein expression while shedding light on the underlying genetic mechanisms in cancer cells. The results indicate that the 2BG sampler not only produces more effective samples than the 3BG counterpart but also significantly reduces computational costs, thereby enhancing the applicability of Bayesian methods in high-dimensional data analysis. This contribution extends the understanding of shrinkage techniques in statistical modeling and offers valuable insights for cancer genomics research.

Keywords: *Bayesian Lasso; Spike-and-Slab; Gibbs sampler; High-Dimensional Statistics, MCMC Methods, Cancer Genomics*

1.   Introduction

Linear regression aims to model the relationship between a response variable, y, and a set of predictor variables, $x_1, x_2, \ldots, x_p$, using regression coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$. In situations where the number of predictors, p, far exceeds the number of observations, n, penalization techniques are often employed. These methods help to regularize the model by applying a penalty to the regression coefficients, effectively shrinking some β_j's towards zero. This leads to sparse estimates, which can improve model stability and prediction accuracy in high-dimensional settings. However, despite their effectiveness, penalization methods face the challenge of providing uncertainty quantification for the estimated parameters.

In the Bayesian paradigm, inference is achieved by incorporating prior distributions on the parameters, allowing for direct quantification of uncertainty through the posterior distribution. In recent years, Bayesian shrinkage models have gained popularity for addressing high-dimensional problems, with two main approaches: the use of *spike-and-slab priors*, and continuous shrinkage priors. Spike-and-slab models, such as Laplace-Zero mixtures (Johnstone and Silverman, 2004), combine a point mass at zero with a continuous distribution, providing a flexible framework for inducing sparsity. On the other hand, *purely continuous shrinkage priors*, such as the Horseshoe prior (Carvalho *et al.*, 2010), have emerged as effective alternatives for handling high-dimensional data while quantifying parameter uncertainties.

Additionally, the Bayesian lasso, introduced by Park and Casella (2008), extends the frequentist lasso by interpreting its objective as a Bayesian model. This approach places an independent Laplace prior on the regression coefficients, connecting it to the original lasso formulation developed by Tibshirani (1996), which has been widely used for regularization and variable selection in high-dimensional regression.

However, as with most Bayesian problems, there are no closed form expression for posterior quantification of the Bayesian Shrinkage models. An approach to approximating the posterior distribution is via Variational Bayes algorithms such as mean field variational Bayes (MFVB) methodology. While this is a fast algorithm, Neville *et al.* (2014) points that the MFVB algorithm can perform quite poorly due to posterior dependence among auxiliary variables. Markov chain Monte Carlo (MCMC), an indispensable tool in contemporary Bayesian inference provides us with alternative methods that can be used to approximate the posterior distribution and hence make inferences on the parameters, since the shrinkage priors of these models capitalize on hierarchical representations. Utilizing the power of MCMC, Park and Casella (2008) provided the three-step Gibbs sampler, otherwise known as the three-Block Gibbs sampler (3BG) algorithm to explore the posterior distribution for arbitrary n and p. While Khare and Hobert (2013) proved that the 3BG is geometrically ergodic for arbitrary values of n and p, it tends to suffer from slow convergence especially if p/n is large enough due to high correlation between components of the different blocks, especially that between the regression coefficients in one block and the variance parameters in another(Rajaratnam and Sparks, 2015). Owing to the deterioration in the convergence properties of the 3BG in high-dimesional settings, Rajaratnam *et al.* (2019) proposed an efficient version known as the two-Block Gibbs sampler (2BG) algorithm to overcome the sluggish convergence issues of the former. Theoretical convergence properties can be found in the aforementioned paper for interested readers.

In this study, an extensive simulation analysis is conducted to compare the computational efficiencies of the 3BG and the 2BG using both the *spike-and-slab* priors and Bayesian lasso model following Rajaratnam *et al.* (2019). Here, computing efficiency is measured as the effective sample size per second (N_{eff}/T). Further, the mixing rates are evaluated by examining lag-one autocorrelations ρ_1, with values closer to zero indicating improved mixing performance (Rajaratnam and Sparks, 2015). We consequently apply this method on the well-known NCI-60 cancer cell panel from the National cancer Institute.

Jin and Tan (2021) introduced the 2BG for Bayesian group lasso, sparse group lasso, and fused lasso models, comparing its performance to the 3BG and Hamiltonian Monte Carlo (HMC). While their work focused on group-structured priors, this study takes a different direction by evaluating the performance of the 2BG relative to the 3BG for two classic Bayesian shrinkage models: the spike-and-slab prior and the Bayesian lasso. Furthermore, this paper extends the application to real-world data from the NCI-60 cancer cell line panel, using gene expression data to predict protein expression levels. By incorporating feature selection, the analysis not only identifies the most influential genes but also provides insights into the genetic mechanisms driving protein expression in cancer cells. This contribution deepens the understanding of shrinkage methods in high-dimensional settings while offering novel applications in cancer genomics.

The rest of the article is organized as follows: Section 2 presents an overview of the Bayesian shrinkage framework for regression, including a review of the spike-and-slab priors, the Bayesian lasso, and the 2BG algorithms for exploring posterior distributions. Sections 3 and 4 detail the simulation studies and real data analyses conducted to empirically compare the two algorithms. The article concludes with a discussion of the findings in Section 5. The codes for this project are found at https://github.com/bosafoagyare/2-BGS.

2.   Methods

2.1.   Bayesian Shrinkage Models

Consider a dataset $\mathcal{D}_n = \{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\} = (\mathcal{X}_n, \mathcal{Y}_n)$, where $\mathcal{Y}_n \in \mathbb{R}^n$ represents the response vector, and $\mathcal{X}_n \in \mathbb{R}^{n \times p}$ is an $n \times p$ design matrix of standardized covariates. Let $\boldsymbol{\beta} \in \mathbb{R}^p$ denote the vector of regression coefficients, $\sigma^2 > 0$ the residual variance, and $\mu \in \mathbb{R}$ an unknown

intercept. The regression model is expressed as:

$$\boldsymbol{Y} \mid \boldsymbol{\beta}, \sigma^2 \sim \mathrm{N}_n \left(\mu \boldsymbol{1}_n + \boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n \right) \tag{1}$$

The focus here is on high-dimensional regression problems where the number of covariates, p, greatly exceeds the sample size, n. In a Bayesian framework where sparsity is desired, the use of *spike-and-slab* priors—combining a spike at zero with a normal density and another normal density that is flat near zero—has been explored as a method to shrink regression coefficients toward zero. Alternatively, *purely continuous* shrinkage priors have also been used to achieve this effect. In many Bayesian high-dimensional regression settings, the prior is typically specified as:

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau} \sim \mathrm{N}_p \left(\boldsymbol{0}_p, \sigma^2 \boldsymbol{D}_{\boldsymbol{\tau}} \right), \quad \boldsymbol{\tau} \sim \pi(\boldsymbol{\tau}), \tag{2}$$

where $\pi(\boldsymbol{\tau})$ represents the prior distribution on $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$. Additionally, the prior for σ^2 and μ is assumed to be the improper prior $\pi(\sigma^2, \mu) = 1/\sigma^2$, which is independent of $\pi(\boldsymbol{\tau})$. By integrating out μ and combining Equations 1 and 2, the following conditional distributions are obtained:

$$
\begin{aligned}
&\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{Y} \sim \pi \left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{Y} \right), \\
&\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{Y} \sim \text{ Inverse-Gamma} \\
&\left[(n+p-1)/2, \|\tilde{\boldsymbol{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2/2 + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{\tau}}^{-1} \boldsymbol{\beta}/2 \right], \\
&\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \boldsymbol{Y} \sim \mathrm{N}_p \left(\boldsymbol{A}_{\boldsymbol{\tau}}^{-1} \boldsymbol{X}^{\mathrm{T}} \tilde{\boldsymbol{Y}}, \sigma^2 \boldsymbol{A}_{\boldsymbol{\tau}}^{-1} \right),
\end{aligned}
\tag{3}
$$

where $\boldsymbol{A}_{\boldsymbol{\tau}} = \boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + \boldsymbol{D}_{\boldsymbol{\tau}}^{-1}$ and $\boldsymbol{D}_{\boldsymbol{\tau}} = \mathrm{Diag}\left(\tau_1, \tau_2, \ldots, \tau_p \right)$.

Given the ability to sample from $\pi(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{Y})$, these conditional distributions can be used to construct a three-Block Gibbs sampler (3BG) for drawing from the joint posterior $\pi(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{Y})$. The one-step transition density $\hat{k}$ with respect to Lebesgue measure on $\mathbb{R}^p \times \mathbb{R}_+$ is given by:

$$
\begin{aligned}
\hat{k} \left[\left(\boldsymbol{\beta}_0, \sigma_0^2 \right), \left(\boldsymbol{\beta}_1, \sigma_1^2 \right) \right] = \int_{\mathbb{R}_+^p} &\pi \left(\sigma_1^2 \mid \boldsymbol{\beta}_1, \boldsymbol{\tau}, \boldsymbol{Y} \right) \pi \left(\boldsymbol{\beta}_1 \mid \boldsymbol{\tau}, \sigma_0^2, \boldsymbol{Y} \right) \\
&\times \pi \left(\boldsymbol{\tau} \mid \boldsymbol{\beta}_0, \sigma_0^2, \boldsymbol{Y} \right) d\boldsymbol{\tau}.
\end{aligned}
\tag{4}
$$

2.1.1.   *The Spike-and-Slab* Prior

Under the *spike-and-slab* prior framework, each τ_j is assigned an independent discrete prior that places probability w_j on the value $\kappa_j \zeta_j$ and probability $1 - w_j$ on the value ζ_j, where $\zeta_j > 0$ is a small value, $\kappa_j > 0$ is large, and $w_j \in (0, 1)$. This setup leads to a conditional posterior distribution for $\boldsymbol{\tau} \mid (\boldsymbol{\beta}, \sigma^2, \boldsymbol{Y})$, which is a product of independent discrete distributions. Each distribution assigns probability $\tilde{w}_j$ to the point $\kappa_j \zeta_j$ and probability $1 - \tilde{w}_j$ to the point ζ_j (Rajaratnam *et al.*, 2019). The value of $\tilde{w}_j$ is given by:

$$\tilde{w}_j = \left\{ 1 + \frac{(1 - w_j)\sqrt{\kappa_j}}{w_j} \exp\left[-\frac{\beta_j^2}{2\sigma^2} \left(\frac{\kappa_j - 1}{\kappa_j \zeta_j} \right) \right] \right\}^{-1}. \tag{5}$$

In this study, $\tilde{w}_j$, κ_j, and ζ_j are treated as constants, with values chosen to satisfy the constraints of the subsequent analyses.

2.1.2.   The Bayesian Lasso

In the Bayesian lasso framework, each τ_j is assigned an independent Exponential $\left(\lambda^2/2 \right)$ prior, where $\lambda^2/2$ is the rate parameter of the exponential distribution. This setup implies that the

marginal prior of $\boldsymbol{\beta}$ (given σ^2) follows independent Laplace distributions for each component, effectively shrinking the regression coefficients. Consequently, the conditional posterior distribution of $\boldsymbol{\tau} \mid (\boldsymbol{\beta}, \sigma^2, \boldsymbol{Y})$ assigns independent inverse Gaussian distributions to each $1/\tau_j$, which allows for efficient sampling. This full framework is extensively discussed by Park and Casella (2008) within the context of the 3BG algorithm.

2.2.   The Two-Block Gibbs Sampler for Bayesian Shrinkage Models

Rajaratnam *et al.* (2019) introduce a lemma that forms the foundation for a two-block Gibbs sampler, which addresses the slow convergence properties of the three-block Gibbs sampler .

Lemma 1. For the Bayesian model specified in equation 2, the posterior distribution of $\sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{Y}$ follows an inverse gamma distribution with shape parameter $(n-1)/2$ and scale parameter $\tilde{\boldsymbol{Y}}^{\mathrm{T}} \left(\boldsymbol{I}_n - \boldsymbol{X} \boldsymbol{A}_{\boldsymbol{\tau}}^{-1} \boldsymbol{X}^{\mathrm{T}} \right) \tilde{\boldsymbol{Y}}/2$.

This lemma enables the development of a 2BG sampler to generate samples from the joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2)$, which retains the computational tractability of the original 3BG sampler. The key difference lies in the way σ^2 is sampled: instead of drawing $\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{Y}$ as done in 3BG (refer to equation 3), the 2BG sampler uses the draw $\sigma^2 \mid \boldsymbol{\tau}$, as derived in Lemma 1. The algorithm alternates between sampling $(\boldsymbol{\beta}, \sigma^2) \mid \boldsymbol{\tau}, \boldsymbol{Y}$ and $\boldsymbol{\tau} \mid (\boldsymbol{\beta}, \sigma^2, \boldsymbol{Y})$. Specifically, $(\boldsymbol{\beta}, \sigma^2)$ is sampled by first drawing $\sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{Y}$, followed by drawing $\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \boldsymbol{Y}$.

The cyclical steps of the 2BG sampler can be summarized as follows:

$$\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{Y} \sim \pi \left(\boldsymbol{\tau} \mid \boldsymbol{\beta}, \sigma^2, \boldsymbol{Y} \right),$$

$$(\boldsymbol{\beta}, \sigma^2) \mid \boldsymbol{\tau}, \boldsymbol{Y} \sim \begin{cases} \sigma^2 \mid \boldsymbol{\tau}, \boldsymbol{Y} \sim \text{Inverse-Gamma}\left[\frac{n-1}{2}, \frac{\tilde{\boldsymbol{Y}}^{\mathrm{T}} (\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A}_{\boldsymbol{\tau}}^{-1}\boldsymbol{X}^{\mathrm{T}}) \tilde{\boldsymbol{Y}}}{2} \right], \\ \boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\tau}, \boldsymbol{Y} \sim \mathrm{N}_p \left(\boldsymbol{A}_{\boldsymbol{\tau}}^{-1} \boldsymbol{X}^{\mathrm{T}} \tilde{\boldsymbol{Y}}, \sigma^2 \boldsymbol{A}_{\boldsymbol{\tau}}^{-1} \right), \end{cases}$$

where $\boldsymbol{A}_{\boldsymbol{\tau}} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{D}_{\boldsymbol{\tau}}^{-1}$, and $\boldsymbol{D}_{\boldsymbol{\tau}}$ is the diagonal matrix formed by the τ_j's.

3.   Simulation studies

We assess the computational efficiency of the 2BG and 3BG samplers through a series of simulation studies. All simulations were conducted on a Windows 11 Pro PC with 16.0 GB RAM, 8 cores, and an Intel(R) Core(TM) i7-8650U @ 1.90GHz processor.

3.1.   Data Generation

The data is simulated from the following linear model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_* + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^{n \times 1}$ is a vector of standard normal random variables, and $\boldsymbol{\beta}_* \in \mathbb{R}^{p \times 1}$ represents the true regression coefficients. We conduct simulations for two distinct models:

1. *Spike-and-Slab Model*: Two sets of simulations are performed with $n = \{50, 100\}$.

2. *Bayesian Lasso Model*: A single set of simulations is performed with $n = 75$.

For both models, the dimension p is chosen such that $\frac{p}{n} \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5\}$. This results in 10 datasets for each (n, p) combination for the Spike-and-Slab model using the 2-block Gibbs sampler and 1-10 block simulations for the Bayesian Lasso model.

For each dataset, the design matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is generated by drawing rows independently from the p-dimensional standard multivariate normal distribution. The columns of $\boldsymbol{X}$ are then standardized to have a mean of zero and a squared Euclidean norm of n. For each p, the first $p/5$ elements of $\boldsymbol{\beta}_*$ are nonzero, drawn independently from the t_2 distribution.

3.2. Simulation Setup

We use the *foreach* package in R to run 6 parallel Markov chains, each with 15,000 iterations. The first 10% of iterations are discarded as burn-in. All chains are initialized with $\boldsymbol{\beta} = \mathbf{1}_p$ and $\sigma_0^2 = 0$. The regularization parameter for the Bayesian Lasso model is set to $\lambda = 1$.

For the Spike-and-Slab model, the hyperparameters are consistently set as $\tilde{w}_j = 1/2$, $\kappa_j = 100$, and $\zeta_j = 1/100$.

3.3. Performance Metrics

To evaluate computational efficiency, we estimate the average lag-1 autocorrelation ρ_1 of the σ^2 marginal after burn-in. This metric is chosen because Rajaratnam and Sparks (2015) suggest that σ^2 offers better insight into chain mixing than $\boldsymbol{\beta}$. Additionally, we estimate the effective sample size per second, N_{eff}/T, where N_{eff} is calculated using the R package *coda* (Plummer *et al.*, 2006) as:

$$N_{\text{eff}} = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho_k},$$

with N denoting the total number of iterations and ρ_k the lag-k autocorrelation.

3.4. Results for the *Spike-and-slab* model

The left panel of Figure 1 presents the empirical average lag-one autocorrelation of the σ^2 component across the six MCMC chains for the *spike-and-slab* model at $n = 50$. It is evident that the new 2BG sampler consistently exhibits smaller average autocorrelations compared to the 3BG sampler across all (n, p) combinations. This suggests that the 2BG sampler has a better mixing rate, and thus, is computationally more efficient than the 3BG algorithm.

Moreover, as the dimensionality of the covariates p increases relative to the sample size n, the performance gap between the 2BG and 3BG samplers widens, further emphasizing the efficiency of the 2BG sampler in high-dimensional settings.

The right panel of Figure 1 shows the average effective sample size per second, N_{eff}/T, on a base-10 log scale for each sampler. The results demonstrate that the 2BG sampler consistently produces more effective samples than the 3BG sampler. As p grows larger relative to n, the computational advantage of the 2BG sampler becomes more pronounced. This suggests that the 2BG sampler incurs a smaller computational cost per chain, making it a more suitable algorithm for high-dimensional problems.

A similar trend is observed when scaling up the sample size to $n = 100$, as depicted in Figure 2. The 2BG sampler continues to outperform the 3BG in terms of both autocorrelation and effective sample size.

3.5. Results for the Bayesian lasso model

A similar analysis as in Section 3.4 is performed under the Bayesian Lasso model. The left panel of Figure 3 presents the empirical average lag-one autocorrelation of the σ^2 component across six MCMC chains at $n = 75$. Consistent with the findings for the *spike-and-slab* model, the 2BG sampler exhibits smaller average autocorrelations than the 3BG sampler across all (n, p) combinations. This trend holds under both the "n-small, p-large" and "n-large, p-small" regimes.

Interestingly, the closest average lag-one autocorrelation between the two algorithms occurs when $n = p = 100$, suggesting comparable performance in this specific scenario. However, as p increases relative to n, the 2BG sampler consistently outperforms the 3BG sampler in terms of mixing efficiency.

The right panel of Figure 3 shows the average effective sample size per second, N_{eff}/T, also on a base-10 log scale. Unsurprisingly, the 2BG sampler generates significantly more effective samples

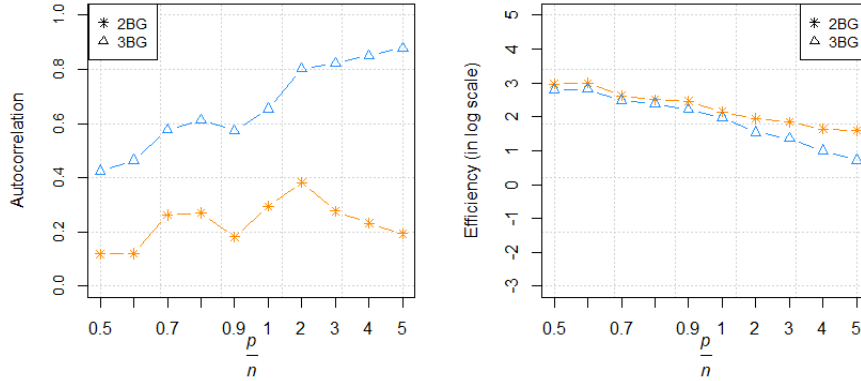**Figure 1.** Empirical lag-one autocorrelation at $n = 50$ (left) and the average effective sample size per second in base-10 log (right) of the σ^2 component of the four MCMC chains for the *spike-and-slab* model.

than the 3BG counterpart, further demonstrating its computational superiority. This indicates that the 2BG sampler achieves a higher degree of efficiency, especially in high-dimensional settings, making it a more suitable choice for Bayesian Lasso models.

4. Application to Real Data

We now apply both the Bayesian lasso and the *spike-and-slab* models to a real dataset. We use the well-known NCI-60 cancer cell panel from the National cancer Institute. This dataset comprises protein expressions for a specific protein selected as the response variable, and the gene expressions of the 100 genes that have the highest (robustly estimated) correlations with the response variable which are screened as candidate predictors. This pre-processing is done using the R *robustHD* package (Alfons, 2021). Hence, we have $n = 59$ samples and $p = 100$ covariates. Each covariate was further standardized to have mean zero and squared Euclidean norm n. We set the hyperparameters of the priors $\tilde{w}_j = 1/2, \kappa = 100$ and $\zeta = 1/200$ with $\lambda = 0.5$. For both shrinkage models, we run 18,000 chains and set the first 10% aside as burn-in.

As illustrated in Figures 4 and 5, the chains for both the *spike-and-slab* and Bayesian lasso models have attained stationarity, indicating that they provide sufficient samples for posterior estimation. Further, from Table 1, we observe that the lag-one autocorrelation for the 2BG model for both the *spike-and-slab* and Bayesian lasso models (0.387 and 0.161 respectively) are much smaller than the 3BG counterparts, indicating better mixing.

Moreover, the 2BG sampler generates about twice as many effective samples as the 3BG sampler in the *spike-and-slab* model, and approximately five times as many for the Bayesian Lasso model. This demonstrates that the 2BG sampler is not only computationally more efficient but also highly effective, particularly in high-dimensional scenarios.

Dataset	n	p	Autocorrelation		N_{eff}	
			2GB	3GB	2BG	3BG
Gene (*spike-and-slab*)	59	100	0.387	0.774	3,263	1,639
Gene (*Bayesian lasso*)	59	100	0.161	0.700	10,921	2,856

Table 1. Lag-one autocorrelation and effective sample size per second for the σ^2-component of 2BG and 3BG of the *spike-and-slab* and Bayesian Lasso models as applied to the protein gene dataset

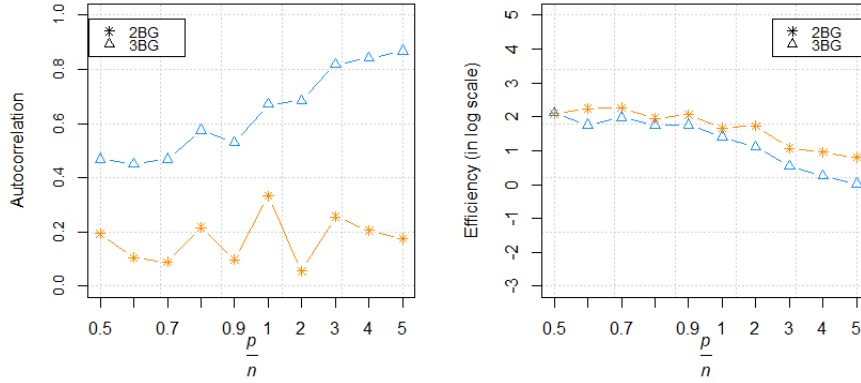**Figure 2.** Empirical lag-one autocorrelation at $n = 100$ (left) and the average effective sample size per second in base-10 log (right) of the σ^2 component of the four MCMC chains for the *spike-and-slab* model

5.   Discussion

In this study, we have used both simulation studies and real data analysis to demonstrate the computational prowess of the two-block Gibbs (2BG) sampler algorithm, specifically within the context of the *spike-and-slab* and Bayesian Lasso models. Our results consistently show that the 2BG is faster and more efficient relative to the three-block Gibbs (3BG) sampler. This efficiency is particularly evident in its superior mixing rates, as evidenced by lower lag-one autocorrelations, and its ability to produce substantially higher effective sample sizes per unit of time. These computational advantages make the 2BG sampler a highly attractive choice for practitioners working in high-dimensional settings, where computational complexity can often be a bottleneck.

Furthermore, this study extends the application to real-world data from the NCI-60 cancer cell line panel, using gene expression data to predict protein expression levels. By incorporating feature selection, the analysis not only identifies the most influential genes but also provides insights into the genetic mechanisms driving protein expression in cancer cells. This contribution deepens the understanding of shrinkage methods in high-dimensional settings while offering novel applications in cancer genomics. The results of our real data application show that the 2BG sampler can handle real-world, high-dimensional data efficiently, providing both robust estimates and computational scalability.
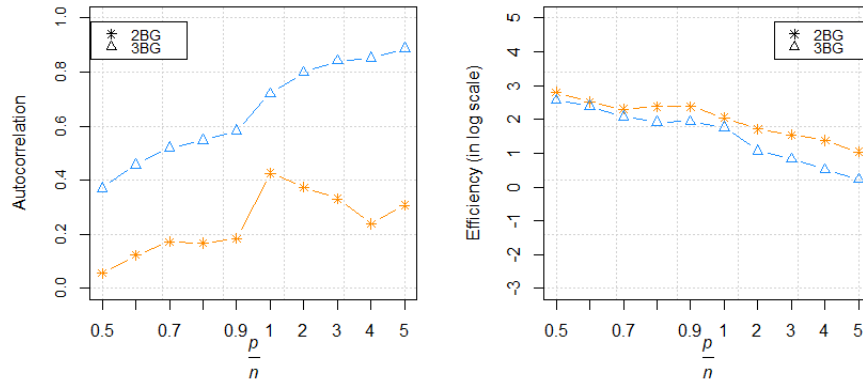
Figure 3. Empirical lag-one autocorrelation at $n = 75$ (left) and the average effective sample size per second in base-10 log (right) of the σ^2 component of the four MCMC chains for the Bayesian lasso model
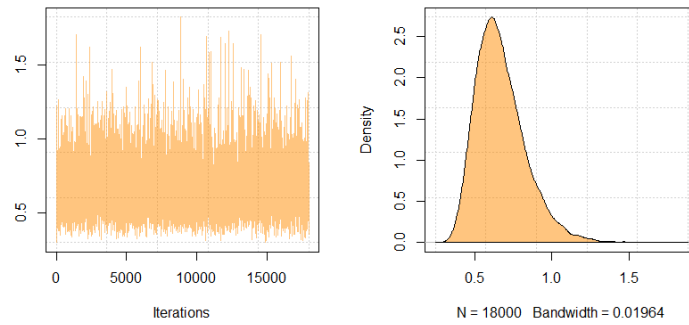

Figure 4. Trace-plot and density plot of the 2BG *spike-and-slab* model for the proteins data
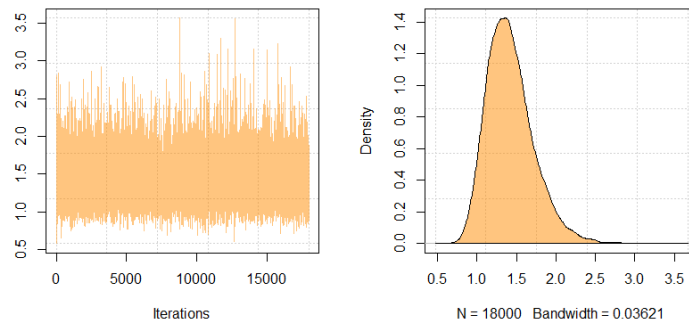

Figure 5. Trace-plot and density plot of the 2BG Bayesian lasso model for the proteins data

References

Alfons, A. (2021). robusthd: An r package for robust regression with high-dimensional data. *Journal of Open Source Software*, **6**(67), 3786.

Carvalho, C. M. *et al.* (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**(2), 465–480.

Jin, R. and Tan, A. (2021). Fast markov chain monte carlo for high-dimensional bayesian regression models with shrinkage priors. *Journal of Computational and Graphical Statistics*, **30**(3), 632–646.

Johnstone, I. M. and Silverman, B. W. (2004). Needles and straw in haystacks: Empirical bayes estimates of possibly sparse sequences.

Khare, K. and Hobert, J. P. (2013). Geometric ergodicity of the bayesian lasso.

Neville, S. E. *et al.* (2014). Mean field variational bayes for continuous sparse signal shrinkage: pitfalls and remedies.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, **103**(482), 681–686.

Plummer, M. *et al.* (2006). Coda: convergence diagnosis and output analysis for mcmc. *R news*, **6**(1), 7–11.

Rajaratnam, B. and Sparks, D. (2015). Mcmc-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. *arXiv preprint arXiv:1508.00947*.

Rajaratnam, B. *et al.* (2019). Uncertainty quantification for modern high-dimensional regression via scalable bayesian methods. *Journal of Computational and Graphical Statistics*, **28**(1), 174–184.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1), 267–288.