

Comparing Methods: Naïve Elastic Net against Bayesian Regression with a Horseshoe Prior

Sarah Masri
97415681

STAT 447C
The University of British Columbia
April 19, 2023

Abstract

Large-scale epigenetic analyses are an increasingly useful tool for prognostic prediction of disease and associated gene activation [1]. Problems involving high-dimensional gene expression data are often challenging by nature due to the number of predictors being much larger than the number of observations. Regularized mechanisms with built-in variable selection are often employed to overcome this. The objective of this project is to compare performance and advantages of some regularized linear regression approaches to Bayesian hierarchical generalized linear models in high-dimensional inference. Namely, to compare elastic net regression to Bayesian penalized regression with a horseshoe prior in the context of prediction and gene detection for high-dimensional molecular data.

A public repository containing implemented methodology can be found through this [this link](#).

Introduction

The host immune response is an incredibly complex system responsible for eliminating pathogens (e.g., viruses, bacteria) to protect the organism from infection or disease. Studying this functionality is crucial to understanding how an organism defends itself against foreign invaders, and helps incite progress over a variety of applications such as vaccine development, treatment for infectious disease, advancing cancer immunotherapy, and stimulating general improvements in healthcare and disease prevention [2].

The host immune response is made up of a myriad of cell types that vary in biological properties and physical location. Specifically, the immune system is comprised of two parts that work together but perform different tasks: the more general innate immune system, and the specialized adaptive immune system [3]. Innate immune cells such as eosinophils, neutrophils, basophils, monocytes, and mast cells, are a host's first line of defense against pathogens. Adaptive immune cells such as B lymphocytes and T lymphocytes are responsible for identifying different antigens. The innate and adaptive systems are linked through Natural Killer cells and antigen presenting cells, such as macrophages [4].

Recent technological advancements such as mass cytometry and single-cell RNA sequencing (scRNA-seq), have progressed in the pursuit of differentiating between immune cell subsets in transcriptomics data (RNA expression data) with increasing accuracy [5].

However, this kind of genomic data presents challenges due to its high-dimensional nature, often with a much larger number of predictors than samples as well as noisy and missing values. Machine learning methods are an accessible way to address these challenges by selecting gene subsets and estimating gene coefficients through regression models simultaneously. This can be done through a frequentist or Bayesian lens, using a variety of methods.

This report aims to compare the accuracy and efficiency of the frequentist naïve elastic net regression against the horseshoe prior for Bayesian regression for selecting gene subsets and estimating gene coefficients. A review of historical literature and methodology related to this comparison is provided, followed by a methodological description for each model, and finally a comparison carried out on real transcriptomic data.

Literature Review

In both research and industry, it is not uncommon to encounter high-dimensional datasets with a much larger number of predictors than samples, such as in gene expression data [6]. Including all covariates in such a dataset may lead to overfitting, and so determining which covariates are relevant to the outcome of interest becomes an increasingly useful problem. This can be accomplished through methods that perform model selection through either a frequentist or Bayesian framework. The comparison of frequentist against Bayesian methods is not a novel contrast, but

an important one. When formulating a problem, it is crucial to choose methods that achieve the greatest possible accuracy and computational efficiency.

This report restricts its methods to high-dimensional gene expression data. Although the application of the results is unique, the comparison of methods may be extended to other fields and applications that require similar high-dimensional datasets. Hence, this literature review includes the comparison of penalized linear regression models to comparable Bayesian approaches on high-dimensional data rather than the gene-expression data, although much of the literature cited is within the application of genomics.

A commonality between frequentist and Bayesian methods for gene selection and the estimation of gene coefficients is the application of shrinkage methods. Each of these methods pursue an optimal trade-off between accuracy and computational efficiency. These literature includes applications of the frequentist penalized regression methods assessed on their own, or compared directly against Bayesian priors with shrinkage properties. Some papers extend this comparison into novel approaches combining frequentist and Bayesian methods.

Narisetty (2020) provides a comprehensive review and comparison of some of the most well-known methods for variable selection on high-dimensional data. Both frequentist and Bayesian methods are included, with an emphasis on Bayesian approaches with priors that induce shrinkage and sparsity. Frequentist methods include LASSO, elastic net, adaptive least-squares methods, and other regression methods with a nonconvex penalty. Bayesian frameworks for model selection are reviewed including a spike and slab prior, a horseshoe prior, global-local shrinkage priors, and a review of Gibbs sampling to compute the posterior distribution. Other samplings and approximation algorithms are reviewed (e.g., Laplace approximation, variational approximation). A variety of these widely used models are implemented on polymerase chain reaction (PCR) data, with results indicating Bayesian methods yielding better performance (smaller BIC) when compared to other frequentist methods. Namely, LASSO regression returns a model size of 10 and BIC of 144.22 whereas a Bayesian implementation with a horseshoe prior returns a model size of 7 with a BIC of 126.54.

More frequentist penalized regression methods are explored in Waldmann et al. (2013), Torang et al. (2019), and Ogutu et al. (2012). Linear regression methods with convex penalties including LASSO and elastic net are shown to overcome some challenges from high dimensional data through their shrinkage properties and selection operators. Cross validated results on simulated and real world data show opti-

mistic results for lasso and elastic net. Namely, these methods used as classifiers for genomic selection returned greater accuracy compared to ridge regression and non-regularized linear regression models.

Outside of the frequentist school of thought, Bayesian estimation using a horseshoe prior on high-dimensional data is proposed in Banerjee (2022). This choice is motivated by the ability of the horseshoe prior to induce a strong shrinkage effect on the mean parameters. The estimation performance is demonstrated on both simulated and real-life data, returning results with particularly high accuracy, precision, and sensitivity.

Münch et al. (2021) implements an adaptive group-regularized logistic elastic net regression method on both simulated and RNAseq data. This paper highlights the common criticism of elastic net, that feature selection is often unstable and returns different features for different instances of the data. Including prior information to the model seems to reduce this instability, and refers to a possible benefit of extending the method to a group-regularized horseshoe model. Other combined methods are proposed in Li and Lin (2010), and Makalic and Schmidt (2015). It is shown that a Bayesian elastic net behaves comparably with a non-Bayesian elastic net, but better in feature select [7]. Additionally, a Bayesian horseshoe linear regression outperforms other Bayesian methods in terms of computational efficiency for high-dimensional data.

Finally, Bhadra et al. (2017) contrast and surveys the Lasso and horseshoe regularization on high-dimensional data. The paper discusses the theoretical optimality for both Lasso and horseshoe estimators, and notes how horseshoe regression has not been studied at the same depth of penalized regression. It is addressed that the theoretical optimality of these methods rely on the underlying assumptions of the data. The paper concluded with optimism towards the Bayesian horseshoe estimator.

Although Bhadra et al. (2017) compared a penalized least-squares method with horseshoe regression, no papers found compared a naïve elastic net directly against the Bayesian horseshoe. This report aims to extend the result of this literature to the use of such methods on high-dimensional transcriptomics data.

Methods

Naïve Elastic Net Regression

Regularized methods for linear regression have the ability to control the complexity of a linear regression model by penalizing coefficients and effectively performing model selection. This is commonly used for

feature selection on high-dimensional data and can be particularly useful for genomic data. Elastic net regression, a method within the family of regularized regression methods, is a hybrid model that combines an L_1 and L_2 penalty to shrink coefficients of predictors. The emergence of this method was due to lasting critique of LASSO regression, that its feature selection is too dependant on the data itself which may lead to unstable results [8]. This combination of penalties helps to reduce the possibility of overfitting and adjust for collinearity, and has been shown to outperform other regularized methods such as LASSO and Ridge regression in very high dimensional data [4, 8].

Consider a data set with n observations and p predictors. Let $\mathbf{X} = (\mathbf{x}_1 | \dots | \mathbf{x}_p) \in \mathbb{R}^{n \times p}$ represent the design matrix and $\mathbf{y} \in \mathbb{R}^n$ be the response. Assume that the response is centered and predictors are standardized such that,

$$\sum_{i=1}^n y_i = 0, \quad \sum_{i=1}^n x_{ij} = 0, \quad \text{and,} \quad \sum_{i=1}^n x_{ij}^2 = 1, \quad (1)$$

for $j = 1, 2, \dots, p$. For some non-negative hyperparameters, the naïve elastic net (NEN) loss function is defined as

$$L(\lambda_1, \lambda_2, \beta) = \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2. \quad (2)$$

This function can also be generalized to

$$L(\lambda, \alpha, \beta) = \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda(\alpha \|\beta\|_1 + (1 - \alpha) \|\beta\|_2^2), \quad (3)$$

for some $\alpha \in [0, 1]$. If $\alpha = 0$, the penalty term is reduced to that of LASSO regression, and if $\alpha = 1$ it is reduced to the penalty of ridge regression. The NEN estimator $\hat{\beta}$ is the minimize of its loss function defined by,

$$\hat{\beta} = \arg \min_{\beta} L(\lambda_1, \lambda_2, \beta). \quad (4)$$

In short, NEN can be considered as a penalized least squares problem. The solution to the NEN problem can be shown through simple algebra by transforming the problem into an equivalent LASSO problem by augmenting the data [8]. Given \mathbf{y} and \mathbf{X} , define $\mathbf{y}^* \in \mathbb{R}^{n+p}$, $\mathbf{X}^* \in \mathbb{R}^{(n+p) \times p}$ such that

$$\mathbf{y}^* = \begin{pmatrix} \mathbf{y} \\ \mathbf{0}_p \end{pmatrix}, \quad \mathbf{X}^* = (1 + \lambda_2)^{-1/2} \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2} \mathbf{I}_{p \times p} \end{pmatrix} \quad (5)$$

Let $\gamma = \frac{\lambda_1}{\sqrt{1 + \lambda_2}}$ and $\beta^* = \sqrt{1 + \lambda_2} \beta$. We can write the NEN loss function as

$$L(\gamma, \beta^*) = \|\mathbf{y}^* - \mathbf{X}^* \beta^*\|_2^2 + \gamma \|\beta^*\|_1. \quad (6)$$

Further, let

$$\hat{\beta}^* = \arg \min_{\beta^*} L(\gamma, \beta^*). \quad (7)$$

Finally, we get

$$\hat{\beta} = \frac{1}{\sqrt{1 + \lambda_2}} \beta^*. \quad (8)$$

The tuning parameters (λ, α) control the strength and trade-off between the L_1 and L_2 penalties. A smaller value of λ leads to stronger regularization and more sparse coefficient estimates. The L_1 -ratio parameter α determines the ration of L_1 and L_2 penalties. Cross-validation methods like grid search can help optimize these hyperparameters for model performance, and adjust uniquely to the dataset being used to build the model.

Horseshoe Prior

The horseshoe prior is a Bayesian prior distribution with strong theoretical guarantees in the context of regression analysis. The motivation behind the choice of this prior is largely due to its shrinkage properties, the ability to identify true non-zero coefficients, which is particularly useful for datasets with a large number of predictors and small number of samples. The tails of this distribution is rather wide and decays like a second-order polynomial. It follows that its penalty functions exhibits logarithmic behaviour, and is non-convex [9, 10].

The horseshoe prior for regression coefficients $\beta \in \mathbb{R}^p$, with some global shrinkage parameter τ , is given by the normal means model integrated into the following hierarchical model,

$$\begin{aligned} \mathbf{y} | \mathbf{X}, \beta &\sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n), \\ \beta_i | \lambda_i^2, \tau^2, \sigma^2 &\sim \mathcal{N}(0, \lambda_i^2 \tau^2 \sigma^2), \\ \sigma^2 &\sim \sigma^{-2} d\sigma^2, \\ \lambda_i^2 | v_i &\sim \mathcal{IG}(1/2, 1/v_i), \\ \tau^2 | \zeta &\sim \mathcal{IG}(1/2, 1/\zeta), \\ v_1, \dots, v_p, \zeta &\sim \mathcal{IG}(1/2, 1) \end{aligned}$$

for $i = 1, 2, \dots, p$ [9]. Note that the above makes use of the scale mixture representation of the half Cauchy distribution [10]. Specifically, for some half Cauchy random variable $X \sim C^+(0, \psi)$, we can write

$$X^2 | \phi \sim \mathcal{IG}(1/2, 1/\phi), \quad \phi | \psi \sim \mathcal{IG}(1/2, 1/\psi^2) \quad (9)$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse gamma distribution.

The hierarchical models allows for the convenient use of a Gibbs sampling scheme to sample the posterior distribution [11]. Makalic and Schmidt (2016) give posterior distributions of the regression coefficients β , σ^2 , the local and global hyper-variances λ_j^2 and τ^2 , and the auxiliary variables v_j and ζ . As seen in equations (A.1) - (A.6), the conditional posterior distributions for all parameters except β follow an inverse-Gamma distributions.

Data

The elastic net method compared against a Bayesian regression model with a horseshoe prior is performed on real-world transcriptomics data. This data include 175 immune cells sequenced from 11 patients diagnosed with breast cancer [12]. The data is taken from the Gene Expression Omnibus, which is a public data repository for array and sequence-based genetic data [13]. Data cleaning, pre-processing, and normalization protocols are sourced from Torang et al., (2019).

The data consists of RNA sequenced data from 15 different immune cell types. These cell types are inherently divided into two groups: immune cells and adaptive cells. If these cell types were to be analyzed together, then any method applied may unnecessarily eliminate genes due to gene overlap between groups [4]. To avoid this issue, only immune cells are analysed in this report. Finally, there are 15453 total genes (covariates) observed in the dataset.

Analysis

The primary goal of the analysis is to train a gene selection model on immune cell types. Genetic data often requires extensive pre-processing that is motivated primarily by the underlying biology and not the statistical methods intended for use. The pre-processing used for this data is borrowed from Torang et al. (2019) and is not unique to the methods used in this report. First, the data is adjusted using a multistep normalization process before training the models. Then, insignificant genes with low expression counts are removed from the dataset. Finally, missing values are imputed.

The dataset is split into training and testing sets of size 108 and 77 respectively. A cross-validated elastic net model to perform gene selection and coefficient estimation is fit on the training set using `cv.glmnet` with `nfold=10` from the `glmnet` package in R. This minimizes the loss defined by equation 3 for each $\lambda = 0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, 0.0001$.

Additionally, Bayesian regression with a horseshoe prior as described above is implemented on the training set to select genes and estimate gene coefficients. The probabilistic programming language is employed using the `RStan` package to fit the specified model to the data using Markov Chain Monte Carlo (MCMC) sampling with 2000 iterations and 4 chains. This process allows for Bayesian estimation of regression coefficients while incorporating the horseshoe prior, which promotes sparsity in the coefficient estimates.

Note that Torang et al., (2019) simulated supplementary noisy data using the limited transcriptomics data available. This was not possible for the purposes of this report, since estimating models from

this much larger dataset is too computationally burdensome to compute on the machine used during this analysis. The limited number of genetic samples are analysed, with the expectation of low performance.

Results

Simulation

Since the breast cancer data is not expected to demonstrate the methods described to their full potential, a much simpler dataset is simulated to compare methods. Synthetic data is simulated following a linear regression model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (10)$$

where $\mathbf{y} \in \mathbb{R}^n$ is the response, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the design matrix, β is the vector of true regression coefficients, and ε is the random error term assumed to follow a Normal distribution. The number of samples are fixed to be 1000, and the number of predictors 100. To generate \mathbf{X} , values are randomly drawn from a standard Normal distribution. Predictors are also normalized to have mean zero and unit variance to ensure they exist on the same scale. True regression coefficients β are represented by a sparse vector, with 7 non-zero coefficients sampled uniformly at random in the interval $(-5, 5)$. This simulates the circumstance where only a few predictors influence the response variable. Finally, \mathbf{y} is computed using equation 10.

Naïve elastic net is fitted using a logarithmic grid of 100 proposed values of λ in the interval $(0.001, 1000)$. An approximate optimal value of $\lambda = 0.0376$ is selected by minimizing the cross-validated mean square error over. A mixing parameter of 0.05 is selected to establish an equal balance between L_1 and L_2 penalties. A visualization of this choice can be observed in Figure 2.

Bayesian regression with a horseshoe prior is subsequently. As discussed previously, this introduces a hierarchical model where the regression coefficients β are assumed to follow a sparse prior distribution. This methodology encourages sparsity by allowing some coefficients to shrink to zero, which is indeed observed in the results. Since only 2000 iterations per chain are run, variable selection is performed with a threshold of 0.001. In other words, regression coefficients less than 0.001 are assumed to be shrunk to zero, and not needed in the regression model.

Results from the simulated data show that the NEN and Bayesian regression model with a horseshoe prior perform comparably, the NEN reporting a lower and more desirable mean square error (MSE),

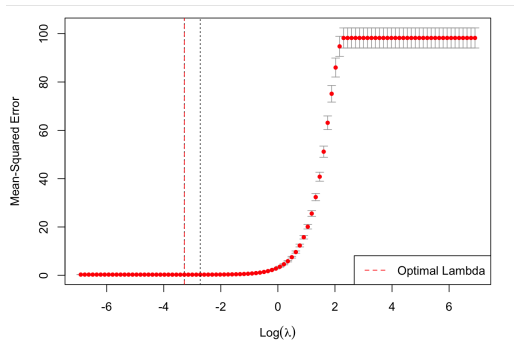


Figure 1: Simulation Optimal Lambda

Cross-validated errors over different values of λ computed in the naïve elastic net model. Vertical dashed lines indicate the most optimal values selected via cross validation.

Table 1: Fitted Model Results

	NEN	Horseshoe
MSE	0.236	0.240
Number of Selected Variables	34	28
Number of Correctly Selected Variables	2	3

Table reporting results from fitted NEN and horseshoe

but the horseshoe model return more sparse coefficients, closer to the ground truth.

Breast Cancer Data

The NEN and horseshoe methods were each applied on the high-dimensional breast cancer data. Cell types are classified through each model, and non-zero gene coefficients are returned. It is important to note that despite the limited data, the NEN method took approximately 40 seconds to run and the Bayesian regression with a horseshoe prior took over 12 hours. The latter computational length might be explained by the high dimensional of the dataset, and the complexity of the horseshoe prior. It is possible that a more simple prior that involves fewer hyperparameters may have been less computationally demanding.

As expected, results from NEN reflect a poor fit. With an optimal $\lambda = 0.0001$, the total cell type classification accuracy is reported to be only 0.365. The generated breast cancer data used in Torang et al. (2019)'s study reflect classification of approximately 0.876 indicating that our poor performance is likely due to the extremely small number of samples compared to the much larger number of predictors. It is important to note that the focus of this report is not to classify cell type, but to examine the feature selection

capabilities between the two methods since our stan code is not adapted to perform classification.

NEN selects a total of 110 genes to be used in the regression model out of a total 1946 genes, whereas Bayesian regression with a horseshoe prior selects 1946 that are far from zero. The use of the term "far from zero" reflects a gene coefficient threshold of 0.1. In other words, since only 2000 iterations were used for each chain in the MCMC algorithm, coefficients with absolute value less than or equal to 0.1 are assumed to be close enough to 0 to exclude from the regression model. It is notable that each model captured a unique different subset of genes, with no intersection. It is not possible to validate the accuracy these coefficients since determining the true gene expression scores with absolute certainty is challenging biologically, and no ground truths are provided in the dataset.

The stark differences in the selected genes may provide insight to the biological factors underlying this data. NEN tends to favour features that demonstrate strong linear relationships with the response, implying that its selected genes exhibit strong linear dependencies in the dataset. Alternately, Bayesian regression with a horseshoe prior assumes the gene coefficients to be close to zero, but allows for a small subset of coefficients to be large. Instead of capturing linear relationships, this method instead reflects significant non-linear or sparse associations within the dataset. Overall, these difference in selected genes may be explained by the complicated nature and diversity of genomics data, and highlights the importance of employing multiple statistical techniques.

Conclusion

Although the results from the breast cancer data is not as informative as one would hope in a formal study, the completion return of the implementation offer optimism in utilizing Bayesian regression with a horseshoe prior on this kind of genomic data. The differences in performance and results of these methods underscores the importance of applying diverse methodology within this field. Both NEN and Bayesian regression with a horseshoe prior extract meaningful insights into transcriptomics, their methods are built on incomparable foundations, with NEN emphasizing linear relationships, and Bayesian regression prioritizing sparse signal detection and non-linear associations. By leveraging the strengths and differences between these kinds of frequentist against Bayesian methods towards feature selection and coefficient estimation, researchers can gain a broader, more comprehensive understanding of the underlying biological processes.

References

- [1] Tang, Zaixiang et al. “The spike-and-slab lasso generalized linear models for prediction and associated genes detection.” en. In: *Genetics* 205.1 (Jan. 2017), pp. 77–88.
- [2] *Host immune response: Enabling faster diagnoses.* <https://biomeme.com/our-vision/host-response>. Accessed: 2024-4-17.
- [3] *The innate and adaptive immune systems.* Institute for Quality and Efficiency in Health Care (IQWiG), July 2020.
- [4] Torang, Arezo, Gupta, Paraag, and Klinke 2nd, David J. “An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets.” en. In: *BMC Bioinformatics* 20.1 (Aug. 2019), p. 433.
- [5] Lin, Wenjia et al. “Mass cytometry and single-cell RNA sequencing reveal immune cell characteristics of active and inactive phases of Crohn’s disease.” en. In: *Front. Med. (Lausanne)* 9 (2022), p. 1064106.
- [6] Narisetty, Naveen Naidu. “Bayesian model selection for high-dimensional data.” In: *Handbook of statistics* (2020), pp. 207–248.
- [7] Li, Qing and Lin, Nan. “The Bayesian elastic net.” In: *Bayesian Anal.* 5.1 (Mar. 2010), pp. 151–170.
- [8] Zou, Hui and Hastie, Trevor. “Regularization and variable selection via the elastic net.” en. In: *J. R. Stat. Soc. Series B Stat. Methodol.* 67.2 (Apr. 2005), pp. 301–320.
- [9] Bhadra, Anindya et al. “Lasso meets horseshoe : A survey.” In: (June 2017). arXiv: [1706.10179](https://arxiv.org/abs/1706.10179) [stat.ME].
- [10] Banerjee, Sayantan. “Horseshoe shrinkage methods for Bayesian fusion estimation.” en. In: *Comput. Stat. Data Anal.* 174.107450 (Oct. 2022), p. 107450.
- [11] Makalic, Enes and Schmidt, Daniel F. “A Simple Sampler for the Horseshoe Estimator.” In: *IEEE Signal Processing Letters* 23.1 (Jan. 2016), pp. 179–182. ISSN: 1558-2361. DOI: [10.1109/lsp.2015.2503725](https://doi.org/10.1109/lsp.2015.2503725). URL: <http://dx.doi.org/10.1109/LSP.2015.2503725>.
- [12] Chung, Woosung et al. “Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer.” en. In: *Nat. Commun.* 8.1 (May 2017), p. 15081.
- [13] Edgar, Ron, Domrachev, Michael, and Lash, Alex E. “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.” en. In: *Nucleic Acids Res.* 30.1 (Jan. 2002), pp. 207–210.
- [14] Allerbo, Oskar, Jonasson, Johan, and Jörnsten, Rebecka. “Elastic Gradient Descent, an Iterative Optimization Method Approximating the Solution Paths of the Elastic Net.” In: (Feb. 2022).
- [15] Jennings, Elizabeth M et al. “Bayesian methods for expression-based integration of various types of genomics data.” en. In: *EURASIP J. Bioinform. Syst. Biol.* 2013.1 (Sept. 2013), p. 13.
- [16] Münch, Magnus M et al. “Adaptive group-regularized logistic elastic net regression.” en. In: *Biostatistics* 22.4 (Oct. 2021), pp. 723–737.
- [17] Waldmann, Patrik et al. “Evaluation of the lasso and the elastic net in genome-wide association studies.” en. In: *Front. Genet.* 4 (Dec. 2013), p. 270.
- [18] Ogutu, Joseph O, Schulz-Streeck, Torben, and Piepho, Hans-Peter. “Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions.” en. In: *BMC Proc.* 6 Suppl 2.S2 (May 2012), S10.

Appendix A: Horseshoe Posterior

The conditional posterior distribution of β is

$$\beta | \cdot \sim \mathcal{N}(\mathbf{A}^{-1} \mathbf{X}^T \mathbf{y}, \sigma^2 \mathbf{A}^{-1}), \quad (\text{A.1})$$

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X} + (\tau^2 \mathbf{\Lambda})^{-1}),$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1^2, \dots, \lambda_p^2)$.

The condition posterior distribution of σ^2 is given by

$$\sigma^2 | \cdot \sim \mathcal{IG}((n+p)/2, (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)/2 + \beta^T (\tau^2 \mathbf{\Lambda})^{-1} \beta/2). \quad (\text{A.2})$$

Conditional posterior distributions for λ_i^2 and τ^2 are given by

$$\lambda_i^2 | \cdot \sim \mathcal{IG}\left(1, \frac{1}{v_i} + \frac{\beta_i^2}{2\tau^2\sigma^2}\right), \quad (\text{A.3})$$

$$\tau^2 | \cdot \sim \mathcal{IG}\left(\frac{p+1}{2}, \frac{1}{\zeta} + \frac{1}{2\sigma^2} \sum_{i=1}^p \frac{\beta_i^2}{\lambda_i^2}\right). \quad (\text{A.4})$$

The conditional posterior distribution for auxiliary variables are,

$$v_i | \cdot \sim \mathcal{IG}\left(1, 1 + \frac{1}{\lambda_i^2}\right) \quad (\text{A.5})$$

$$\zeta | \cdot \sim \mathcal{IG}\left(1, 1 + \frac{1}{\tau^2}\right) \quad (\text{A.6})$$

Appendix B: Code

Relevant stan code is included in this appendix. All other code use can be viewed using [this link](#).

```
data {
  int<lower=1> n; // Number of samples
  int<lower=1> p; // Number of predictors
  matrix[n,p] X; // Design matrix
  real y[n];      // n-dimensional response
                  // vector
}

parameters {
  real beta0;           // Intercept
  vector[p] beta;       // Coefficients
  vector<lower=0>[p] lambda; // Global shrinkage
                        // parameter
  real<lower=0> tau;     // Local shrinkage
                        // parameter
}

transformed parameters {
  vector[n] theta ;
  theta = X * beta + beta0;
}

model {
  // Priors
  beta0 ~ normal(0, 10);
  beta ~ normal(0, 10);
  lambda ~ cauchy(0, 1);
  tau ~ cauchy(0, 1);

  // Horseshoe prior
  beta ~ normal(0, 1 * tau * lambda);

  // Likelihood
  y ~ normal(theta, 1); // Gaussian likelihood
}

generated quantities {
  real y_pred[n];
  y_pred = normal_rng(theta, 1); // posterior
                                   // predictive
}
```

Appendix C: Figures

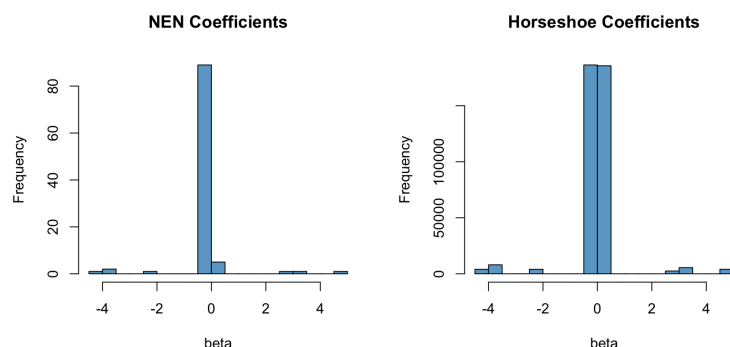


Figure 2: Estimated Coefficients
Histograms displaying the estimated regression coefficients from NEN and Bayesian regression with a Horseshoe prior on simulated data.

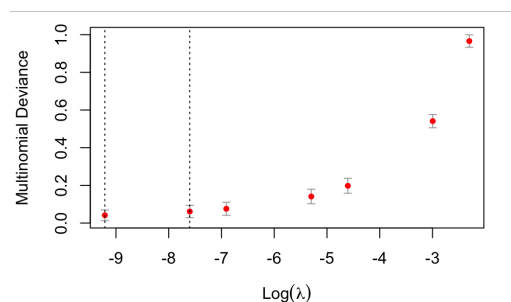


Figure 3: Breast Cancer Optimal Lambda
Cross-validated errors over different values of λ in the breast cancer data. Vertical lines indicate the most optimal values selected via cross validation.