

# Bayesian Subset Selection for Inverse Gauss Regression Models

Yuanqing Zhao<sup>1,\*</sup>, Dengke Xu<sup>2</sup>, Liangqiong Jin<sup>3</sup> and Qingqiong Jiang<sup>3</sup>

<sup>1</sup>College of Mathematics and Information Science, Guiyang University, Guizhou, China

<sup>2</sup>Department of Statistics, Zhejiang Agriculture and Forestry University, Zhejiang, China

<sup>3</sup>College of Computer and Information Engineering, Guizhou Minzu University, Guizhou, China

\* Corresponding author

**Abstract**—Inspired by the idea of Kuo and Mallick, Bayesian subset selection for inverse Gauss regression models is studied by Gibbs sampler and Metropolis-Hastings algorithm in this paper. Simulation study and the aerobic fitness data example are employed to demonstrate the proposed methodology.

**Keywords**—Bayesian subset selection; Gibbs sampler; Metropolis-Hastings algorithm; Inverse Gauss regression models

## I. INTRODUCTION

The inverse Gaussian distribution, which is well known as a probability model that arises in a Brownian motion with positive drift [1], has been widely studied and applied in recent years due to the attractive probabilistic and statistical properties. For example, Tweedie [2,3] presented some statistical characteristics for the inverse Gauss distributions. The inverse Gaussian linear regression models were exhaustively studied by Davis [4] in her unpublished thesis. Seshadri [5] comprehensively and systematically summarized the theoretical properties of the inverse Gaussian distribution. Lin et al. [6] investigated the diagnostics for varying dispersion in the framework of inverse Gaussian linear models based on an adjusted score test. Xu et al. [7] developed a Bayesian procedure to analyze the double generalized linear regression models of the inverse Gaussian distribution. Wu and Li [8] employed penalized likelihood function approach to simultaneously select significant variables in mean and dispersion models of the inverse Gaussian distribution.

On the other hand, a considerable amount of work has been done by Bayesian variable selection during the last three decades. For instance, see more details in Schwarz [9] for Bayesian information criterion, in Kass [10] for Bayes factor, as well as in Geisser and Eddy [11] for pseudo-Bayes factor. Particularly, Mitchell and Beauchamp [12] have mainly focused on Bayesian subset selection of predictor variables under a linear regression model. In addition, Stochastic search variable selection procedure proposed by George and McCulloch [13] has become a useful tool for identifying promising subsets of predictors. Green [14] has employed the reversible jump Markov chain Monte Carlo approach to investigate model uncertainty. Nott and Leone [15] have studied Bayesian selection sampling schemes in the framework of generalized linear models. Recently, Bayesian adaptive lasso proposed by Leng et al. [16] has permitted a treatment for variable selection with flexible penalties. Etc. However, to our knowledge, there is little work done on

investigating Bayesian subset selection for inverse Gauss regression models. Inspired by the idea of Kuo and Mallick [17], the paper aims to propose a Bayesian procedure to select promising subset of explanatory variables under inverse Gauss regression models by using Markov chain Monte Carlo technology.

The structure of the paper is organized as follows. Inverse Gauss regression models which consists of model uncertainties are defined in section 2. In section 3, a Markov chain Monte Carlo procedure is proposed to identify the promising subsets of explanatory variables in the framework of inverse Gauss regression models; Meanwhile, we briefly derive the required conditional distributions in implementing the sampling-based approach. Simulation study and the aerobic fitness data example are employed to demonstrate our proposed methodology in section 4.

## II. MODEL AND NOTATION

We take into account the inverse Gauss regression models specified by the following structure through out of the paper,

$$\begin{cases} y_i \sim IG(\mu_i, \lambda), \\ \mu_i = \sum_{j=1}^p x_{ij} \beta_j \gamma_j, \\ i=1, \dots, n, \end{cases} \quad (1)$$

in which  $y_i$  is the response for the  $i$ th individual,  $y_i \sim IG(\mu_i, \lambda)$  represents that  $y_i$  is distributed as the inverse Gauss distribution with mean parameter  $\mu_i (\mu_i > 0)$  and dispersion parameter  $\lambda (\lambda > 0)$ . Given mean parameter  $\mu_i$  and dispersion parameter  $\lambda$ , the probability density function of  $y_i$  can be expressed by

$$p(y_i | \mu_i, \lambda) = \left( \frac{\lambda}{2\pi y_i} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda (y_i - \mu_i)^2}{2 \mu_i^3 y_i} \right\}. \quad (2)$$

Here unknown parameter  $\beta = (\beta_1, \dots, \beta_p)^T$  is a  $p \times 1$  regression vector to be evaluated, the vector  $x_i = (x_{i1}, \dots, x_{ip})^T$

is a  $p \times 1$  explanatory variable, and  $\gamma = (\gamma_1, \dots, \gamma_p)^T$  is  $p \times 1$  indicator vector with the value of  $\gamma_j (j = 1, \dots, p)$  supported at two points 1 or 0. Particularly, for  $j = 1, \dots, p$ , if  $\gamma_j = 1$  then the  $j$ th explanatory variable  $x_{ij}$  are included in the mean model  $\mu_i = \sum_{j=1}^p x_{ij} \beta_j \gamma_j$ , and if  $\gamma_j = 0$  then we delete the  $j$ th explanatory

variable  $x_{ij}$  when building the mean model  $\mu_i = \sum_{j=1}^p x_{ij} \beta_j \gamma_j$ . The above models specified by (1) and (2) are termed as inverse Gauss regression models including model uncertainties. We mainly focus on deleting the unnecessary explanatory variables from the inverse Gauss regression models in this paper.

Denote  $\theta = (\beta^T, \lambda^T)^T$ . It is assumed that the prior distribution of unknown parameter  $\theta$  and  $\gamma$  is specified by the following structure

$$p(\theta, \gamma) = p(\theta) \prod_{j=1}^p p(\gamma_j) = p(\beta) p(\lambda) \prod_{j=1}^p p(\gamma_j). \quad (3)$$

Similar to Kuo and Mallick [17] and Zhao et al. [18], the following prior distributions are considered throughout this paper,

$$p(\beta) \sim N_p(\beta_0, \Sigma_\beta), p(\lambda) \sim \Gamma(\nu_0, \rho_0), p(\gamma_j) \sim B(1, p_j), \quad (4)$$

where  $N_p(\beta, \Sigma)$  denotes  $p$ -dimensional normal distribution with expectation  $\beta$  and covariance  $\Sigma$ ,  $\Gamma(\nu, \rho)$  represents the Gamma distribution with shape parameter  $\nu$  and scale parameter  $\rho$ ,  $B(1, q)$  is the Bernoulli distribution with parameter  $q$ .  $\beta_0, \Sigma_\beta, \nu_0, \rho_0, p_j$  are some known hyper-parameters.

### III. BAYESIAN ANALYSIS OF THE MODELS

In this section, we present a Markov chain Monte Carlo approach that can identify the promising subset of explanatory variables in the framework of the inverse Gauss regression models. Denote  $y = \{y_1, \dots, y_n\}, x = \{x_1, \dots, x_n\}$ . The Bayesian statistical analysis to parameters  $\theta$  and  $\gamma$  is made by the joints posterior distribution  $p(\theta, \gamma | y, x) \propto p(y | x, \theta, \gamma) p(\theta, \gamma)$ , in which  $p(y | x, \theta, \gamma)$  is the likelihood function of the observed data  $y$  and  $x$ , which is proportion to

$$\prod_{i=1}^n p(y_i | \mu_i, \lambda) \propto \lambda^{\frac{n}{2}} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i} \right\}, \quad (5)$$

where  $\mu_i$  is given in equation (1). It is easy to find that there is not an analytical solution to  $p(\theta, \gamma | y, x)$  due to complexity of the considered model and the uncertain indicator vector  $\gamma$  involved. Therefore, it is very challenging to deal directly with this posterior distribution. Motivated by the idea of data augmentation procedure, the Gibbs sampler [19], a useful tool in statistical calculation, is employed to obtain a sequence of samples from the following joint posterior distribution  $p(\theta, \gamma | y, x) \propto$

$$\lambda^{\frac{\nu_0 + \frac{n}{2} - 1}{2}} \exp \left\{ -\frac{1}{2} (\beta - \beta_0)^T \Sigma_\beta^{-1} (\beta - \beta_0) - \frac{\lambda}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i} - \lambda \rho_0 \right\} \prod_{j=1}^p p(\gamma_j). \quad (6)$$

The key to this algorithm is that random samples  $\{\beta, \lambda, \gamma\}$  are iteratively generated from the following three conditional distribution functions:  $p(\beta | y, x, \lambda, \gamma)$ ,  $p(\lambda | y, x, \beta, \gamma)$  and  $p(\gamma | y, x, \theta, r)$ .

#### A. Conditional Distributions

The conditional distributions are briefly derived in this subsection, which are required in performing the Gibbs sampler [19].

For the distribution  $p(\beta | y, x, \lambda, \gamma)$ , it follows from (6) that

$$p(\beta | y, x, \lambda, \gamma) \propto \exp \left\{ -\frac{1}{2} (\beta - \beta_0)^T \Sigma_\beta^{-1} (\beta - \beta_0) - \frac{\lambda}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i} \right\}. \quad (7)$$

For the distribution  $p(\lambda | y, x, \beta, \gamma)$ , after some calculations, it can be shown that

$$p(\lambda | y, x, \beta, \gamma) \propto \lambda^{\frac{\nu_0 + \frac{n}{2} - 1}{2}} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i} - \lambda \rho_0 \right\},$$

which leads to

$$p(\lambda | y, x, \beta, \gamma) \sim \Gamma(\nu, \rho), \quad (8)$$

where  $\nu = \nu_0 + \frac{n}{2}$  and  $\rho = \rho_0 + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i}$ .

For the distribution  $p(\gamma | y, x, \theta, r)$ , we have

$$p(\gamma | y, x, \theta, r) \propto \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\mu_i^2 y_i} \right\} \prod_{j=1}^p p(\gamma_j). \quad (9)$$

#### B. Implementation

Next, we describe how to obtain the sample  $\gamma$  from (9) firstly. For  $j = 1, \dots, p$ , let  $\gamma_{-j} = (\gamma_1, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_p)^T$ , the conditional distribution of  $\gamma_j$  given  $\gamma_{-j}, y, x$  and  $\theta$  is the Bernoulli distribution

$$B(1, \frac{c_j}{c_j + d_j}), \quad (10)$$

which  $c_j = p_j \prod_{i=1}^n p(y_i | \mu_i, \lambda, \gamma^*)$ ,  $d_j = (1 - p_j) \prod_{i=1}^n p(y_i | \mu_i, \lambda, \gamma^{**})$ ,  $\gamma^*$  and  $\gamma^{**}$

denote  $\gamma$  with the  $j$ th element  $\gamma_j$  replaced by 1 and 0, respectively. It is easy to find that conditional distributions from (8) and (10) are some standard and familiar distributions. Therefore, generating samples from these conditional distributions are direct and fast. However, the corresponding conditional distribution given in (7) is complicated and non-standard distributions. As a result, to directly generate samples from the conditional distribution is rather difficult. The MH algorithm [20,21] is one of the most effective ways to generate samples from those corresponding conditional distributions with the help of proposal distributions. Similar to Tang and Zhao [22] and Zhao and Tang [23], we choose the following

normal distribution  $N_p(0, \sigma_{\beta}^2 \Omega_{\beta})$  as the proposal distribution for sampling  $\beta$  from (7), in which

$$\Omega_{\beta} = (V + \Sigma_{\beta}^{-1})^{-1} \quad \text{and} \quad V = E \left\{ -\frac{\partial^2 \log p(y|x, \theta, \gamma)}{\partial \beta \partial \beta^T} \right\}_{\beta=0}.$$

The coefficient  $\sigma_{\beta}^2$  is selected to make sure the average acceptance rate is about 0.25. The MH algorithm [20,21] for generating observation  $\beta$  from its corresponding conditional distribution (8) is performed as follows. At the  $m$ th iteration with current observation  $\beta^{(m)}$ , new candidate  $\beta^*$  is obtained from the normal distribution  $N_p(\beta^{(m)}, \sigma_{\beta}^2 \Omega_{\beta})$ . We accept the

candidate  $\beta^{(m)}$  with the probability  $\min \left\{ 1, \frac{p(\beta^* | y, x, z, b, \gamma)}{p(\beta^{(m)} | y, x, z, b, \gamma)} \right\}$ .

### C. Bayesian Subset Selection

The joint Bayesian estimates of parameters  $\beta, \lambda$  and  $\gamma$  are expressed by

$$\hat{\beta} = \frac{1}{M} \sum_{m=1}^M \beta^{(m)}, \hat{\lambda} = \frac{1}{M} \sum_{m=1}^M \lambda^{(m)}, \hat{\gamma} = \frac{1}{M} \sum_{m=1}^M \gamma^{(m)}. \quad (11)$$

Where  $\{(\beta^{(m)}, \lambda^{(m)}, \gamma^{(m)}) : m = 1, \dots, M\}$  are the samples of  $\{\beta, \lambda, \gamma\}$  generated from the joint posterior distribution  $p(\theta, \gamma | y, x)$  via the previous Gibbs sampler procedure. Following Geyer [22], these joint Bayesian estimates and sample covariance matrices are consistent with their corresponding posterior expectations and posterior covariance matrices. Similar to Kuo and Mallick [17] and Zhao et al. [18], the promising subsets of explanatory variables can be identified by high posterior probabilities of  $\gamma$

## IV. NUMERICAL EXAMPLES

In this section, several numerical examples are employed to demonstrate the preceding proposed Bayesian subset selection method.

### A. Simulation Study

For  $i = 1, \dots, n$ , it is assumed that the response  $y_i$  is distributed as the inverse Gauss distribution with mean parameter  $\mu_i$  and dispersion parameter  $\lambda$ , i.e.,

$$y_i \sim IG(\mu_i, \lambda). \quad (12)$$

The mean parameter  $\mu_i$  is expressed by

$$\mu_i = x_{i1}\beta_1\gamma_1 + x_{i2}\beta_2\gamma_2 + x_{i3}\beta_3\gamma_3 + x_{i4}\beta_4\gamma_4 + x_{i5}\beta_5\gamma_5. \quad (13)$$

We assume the explanatory variable  $x_{ij} (j = 1, \dots, 5)$  is independently simulated from the uniform distribution on the closed interval  $[0, 1]$ . In addition, the true values of parameters are set by  $(\beta_1\gamma_1, \beta_2\gamma_2, \beta_3\gamma_3, \beta_4\gamma_4, \beta_5\gamma_5)^T = (0, 0, 0, 1, 1)^T$  and  $\lambda = 1$ . Then (13) becomes to

$$\mu_i = x_{i4} + x_{i5}. \quad (14)$$

Data sets are generated from the above models (12) and (14) with different sample sizes  $n=75$  and  $n=150$  in the simulation study. In this paper, the following hyper-parameters in equation (4) are specified in all numerical examples which may stand for non-information priors,

$$\beta_0 = 0, \Sigma_{\beta} = 100I_p, \nu_0 = \rho_0 = 0.01 \quad \text{and} \quad p_j = 0.5 (j = 1, \dots, p).$$

Where  $I_p$  is the  $p \times p$  identity matrix. The preceding proposed Bayesian procedure is employed to make Bayesian subset selection on the assumption of the above model settings and prior inputs. We show the four highest frequency models in table I on the basis of 20000 iterations in Gibbs sampler. Examinations of table I, the following observations are indicated (i) Our approach can precisely identify the mean

model is  $x_{i4} + x_{i5}$  as expected; (ii) Bayesian subset selection method becomes better as the sample size  $n$  increases. All these observations clearly demonstrate that our proposed method is plausible and effective.

TABLE I HIGH FREQUENCY MODELS IN THE SIMULATION STUDY

n=75		n=150	
Explanatory variables	Frequency	explanatory variables	Frequency
$x_{i4} \ x_{i5}$	83.85%	$x_{i4} \ x_{i5}$	96.46%
$x_{i1} \ x_{i4} \ x_{i5}$	5.04%	$x_{i2} \ x_{i4} \ x_{i5}$	1.54%
$x_{i3} \ x_{i4} \ x_{i5}$	4.72%	$x_{i3} \ x_{i4} \ x_{i5}$	1.10%
$x_{i2} \ x_{i4} \ x_{i5}$	4.35%	$x_{i1} \ x_{i4} \ x_{i5}$	0.85%

### B. A Real Data Example

In order to further illustrate our proposed Bayesian approach, we think about a real data example from SAS Use's Guide [24], which is referred to as aerobic fitness data. The aerobic fitness data includes  $p=6$  independent explanatory variables and  $n=31$  responses. For  $i = 1, \dots, n$ , the response is

aerobic fitness, denoted as  $y_i$ , explanatory variables are weight (kg), age (years), time to run 1.5 miles in minutes (runtime), heart rate while running (rnpulse), heart rate while resting (restpulse) and maximum heart rate records while running

(maxpulse), and denoted as  $x_{i1}, \dots, x_{i6}$ , respectively. In order to employ an equation to predict fitness on the basis of the excise tests rather than on expensive cumbersome oxygen consumption measurement, the main goal is to investigate the effect of  $x_{i1}, \dots, x_{i6}$  to  $y_i$ , SAS's results [21] indicate that all explanatory variables except  $x_{i1}$  and  $x_{i5}$  are significant. Inverse Gauss model (12) and the following mean model

$$\mu_i = x_{i1}\beta_1\gamma_1 + x_{i2}\beta_2\gamma_2 + x_{i3}\beta_3\gamma_3 + x_{i4}\beta_4\gamma_4 + x_{i5}\beta_5\gamma_5 + x_{i6}\beta_6\gamma_6, \quad (15)$$

Together with our proposed Bayesian subset selection procedure are employed to analyze the data in the paper. Based on 20000 samples which are collected by Gibbs samplers, we present the higher frequency models in table II.

From table II, it is easy to find that  $x_{i2}, x_{i3}, x_{i4}$  and  $x_{i6}$  make a significant contribution to  $y_i$ , which indicates that the unnecessary explanatory variables  $x_{i1}$  and  $x_{i5}$  we delete are consistent with SAS.

TABLE II HIGH FREQUENCY MODELS IN THE REAL DATA ANALYSIS

Explanatory variables	Frequency
$x_{i2} \ x_{i3} \ x_{i4} \ x_{i6}$	54.92%
$x_{i2} \ x_{i3} \ x_{i4} \ x_{i5} \ x_{i6}$	19.25%
$x_{i1} \ x_{i2} \ x_{i3} \ x_{i4} \ x_{i6}$	14.65%

### ACKNOWLEDGMENT

Zhao's work is fully supported by grants from the National Natural Science Foundation of China (11761016), Guizhou Provincial Science and Technology Foundation of China

(LH[2014]7164), the Natural Science Foundation of the Education Department of Guizhou province in China (KY[2014]249) and the Curriculum development project of Guiyang university in Guizhou of China (2014). Xu's work is partly supported by grants from National Statistical Science Research Project (2015LY70, 2016LZ06), Zhejiang Provincial Natural Science Foundation of China (LY17A010026).

### REFERENCES

- [1] M. C. K. Tweedie, "Inverse statistical variates," Nature, vol. 55, pp. 453, April 1945.
- [2] M.C.K. Tweedie, "Statistical properties of inverse Gaussian distributions. I," Ann. Math. Stat, vol. 28 No. 2, pp. 362–377, June 1957.
- [3] M.C.K. Tweedie, "Statistical properties of inverse Gaussian distributions. II," Ann. Math. Stat, vol. 28 No. 3, pp. 696–705, September 1957.
- [4] A. S. Davis, "Linear statistical inference as related to the inverse Gaussian distribution," Unpublished.
- [5] V. Seshadri, The Inverse Gaussian Distribution, Oxford: Clarendon Press, 1993.
- [6] J. G. Lin, B. C. Wei, N. S. Zhang, "Varying dispersion diagnostics for inverse Gaussian regression models," J. Appl. Stat, vol. 31 No. 10, pp. 1157–1170, December 2004.
- [7] D. Xu, Y. Zhao, W. Chen, "Bayesian inference for double generalized linear regression models of the inverse gaussian distribution," Int. J. Appl. Math. Stat, vol. 52 No. 4, pp.121-131, April 2014.
- [8] L. Wu, and H. Li, "Variable selection for joint mean and dispersion models of the inverse Gaussian distribution," Metrika, vol. 75 No. 6, pp. 795–808, August 2012.
- [9] G. Schwarz, "Estimating the Dimension of a Model," Ann. Stat, vol. 6 No. 2, pp. 461–464, March 1978.
- [10] R. E. Kass, A. E. Raftery, "Bayes factors," J. Am. Stat. Assoc, vol. 90 No. 430, pp. 773–795, June 1995.
- [11] S. Geisser, and W. F. Eddy, "A predictive approach to model selection," J. Am. Stat. Assoc, vol. 74 No. 365, pp. 153–160, March 1979.
- [12] T. J. Mitchell, and J. J. Beauchamp, "Bayesian variable selection in linear regression," J. Am. Stat. Assoc, vol. 83 No. 404, pp. 1023–1032, December 1988.
- [13] E. I. George, and R. E. McCulloch, "Variable selection via Gibbs sampling," J. Am. Stat. Assoc, vol. 88 No. 423, pp. 881–889, September 1993.
- [14] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model dertermination," Biometrika, vol. 82 No. 4, pp. 711–732, December 1995.
- [15] D. J. Nott, and D. Leonte, "Sampling schemes for Bayesian selection in generalized linear models," J. Comput. Graph. Stat, vol. 13 No. 2, pp. 362–382, June 2004.
- [16] C. Leng, M. N. Tran, and D. Nott, "Bayesian Adaptive Lasso," Ann. I. Stat. Math, vol. 66 No. 2, pp. 221–244, April 2014.
- [17] L. Kuo, and B. Mallick, "Variable selection for regression models," Sankhya, vol. B60 No. 1, pp. 65–81, April 1998.
- [18] Y. Zhao, D. Xue, X. Duan, Y. Pang, "Bayesian subset selection for reproductive dispersion linear models," J. Sys. Sci. & Info, vol. 2 No.1, pp. 77–85, February 2014.
- [19] S. Geman, and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," IEEE. T. Pattern. Anal, vol. 6 No. 6, pp. 721–741, November 1984.
- [20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machine," J. Chem. Phys, vol. 21 No. 6, pp. 1087–1092, June 1953.
- [21] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their application," Biometrika, vol. 57 No. 1, pp. 97–109, April 1970.
- [22] N. S. Tang, and Y. Y. Zhao, "Semiparametric Bayesian analysis of nonlinear reproductive dispersion mixed models for longitudinal data," J. Multivariate. Anal, vol. 115, pp. 68–83, March 2013.

- [23] Y. Y. Zhao, and N. S. Tang, "Maximum likelihood estimation and influence analysis in multivariate skew-normal reproductive dispersion mixed models for longitudinal data," *Statistics*, vol. 49 No. 6, pp. 1348–1365, November 2015.
- [24] SAS Institute, *SAS User's Guide: Statistics ver.5*, SAS Institute. Inc, 1985