مجلّه‌ی پژوهش‌های آماری ایران

J. Statist. Res. Iran **16** (2019): 559–568

دوره‌ی ۱۶، شماره‌ی ۲، پاییز و زمستان ۱۳۹۸، صص ۵۵۹–۵۶۸

DOI: 10.52547/jsri.16.2.559

DOR: 20.1001.1.25385771.2020.16.2.12.2

A Note on the Identifiability of General Bayesian Gaussian Models

Amir Hossein Ghatari[*,†], Ashkan Shabbak[‡] and Elham Tabrizi[⋆]

[†] Amirkabir University of Technology
[‡] Statistical Research and Training Center
[⋆] Kharazmi University

Received: 2021/07/08 Approved: 2021/30/11

Abstract. The main aim of this paper is to investigate the identifiability of Bayesian Gaussian regression model. The model is extensively implemented in the various Bayesian modeling concepts such as model fitting and model selection approaches. In accordance with the outcomes, the Bayesian Gaussian model is identifiable when the model's design matrix is full rank.

Keywords. Bayesian statistics; design matrix; Gaussian model; identifiability; Posterior distribution.

MSC 2010: 62J05, 62J12.

1 Introduction

Identifiability is a well-known concept in the statistical modeling and inference. It is a basic precondition for inference. In the absence of identifiability, statistical inference is practically meaningless(Tabrizi, et al. , 2020b). Martín and Quintana (2002) proposed the identifiability property and proved that the mathematical properties concerning the parameters estimators such as consistency, are meaningless for unidentifiable models. Hence, it is required

[*] Corresponding author

to check the identifiability of the model before using any model fitting and consequently model selection approach.
In the frequentist (non-Bayesian) statistics, the likelihood is applied to asses the identifiability property. Several researchers concentrated on this concept in their studies about different types of statistical models. For instance, Christensen (2011) proved that the multiple linear regression is not identifiable when the rank of design matrix is less than the number of the available covariates in the model. Bahrami Samani (2014) studied the identifiability of the covariance components in a mixed model for the continuous and ordinal responses. Also, Miao et al. (2016) studied the effect of identifiability in Gaussian, Gaussian mixture, and t mixture models under nonignorable missing data. In addition, Yu and Dong (2020) proposed identifiability conditions of the linear regression model for censored data. For more information about the importance and usage of the identifiability in the studies about statistical modeling in the various type of models, see Wang (2013), De Leon and Chough (2013), Tabrizi, et al. (2020a) and their cited references.
As mentioned, the likelihood of the statistical model is applied to asses the identifiability in the frequentist statistics. In the Bayesian framework, prior distributions are considered for the models parameters. Hence, they should be considered and studied as random variables instead of the unknown constant values. In this situation, there are various viewpoints. Lindley (1972) stated that the posterior distribution is a suitable substitute for the likelihood and it can be used to check the identifiability of model. Also, Gelfand and Sahu (1999) believed that the priors on parameters can be ignored and the likelihoods of the models are still working to check identifiability. Additionally, Dawid (1979) practically stated an idea about when a model is not identifiable based on conditional densities. In this paper, we follow the idea of Lindley (1972) and consider the posterior distribution to asses the identifiability of target model.
Bayesian Gaussian model for regression problems is the target model of this paper. The model is widely applicable in many types of Bayesian modeling approaches. For more information about the importance of Gaussian model in the Bayesian framework, see Vehtari and Ojanen (2012), Piironen and Vehtari (2017), Ghatari and Ganjali (2020) and their referred references. In the Bayesian statistics, inference for the parameters is based on the posterior distribution. Note that in any model selection method, a model is fitted for each step. Hence, it is necessary to check the identifiability before using any

method to fit or select models.
In this paper, to investigate the identifiability property for the Bayesian Gaussian, we proceed to prove the property by considering suitable prior distributions on the model parameters. In Section 2, the concept of identifiability is reviewed. In Section 3, the formation of the Bayesian Gaussian model is called as target model. In Section 4, the sufficient conditions for identifiability of the target model and other obtained results are provided.

2 Identifiability

As was mentioned in Section 1, identifiability is about the model parameters and it is not about the their estimators. Martín and Quintana (2002) defined the model identifiability as follows:

Definition 1. Let a statistical model be defined by a family of distributions for a random variable parameterized by the vector θ, P_θ , $\theta \in \Theta$, where Θ is the parameter space and P_θ denotes the distribution associated with θ. The model is identifiable on Θ if $P_{\theta_1} = P_{\theta_2}$ implies that $\theta_1 = \theta_2$ for all $\theta_1, \theta_2 \in \Theta$.

Equivalently, a model is not identifiable if there is $\theta_1 \neq \theta_2$ such that $P_{\theta_1} = P_{\theta_2}$. It means there is no reliable way to have real values of parameters even if the number of observations tends to infinity. Also, Martín and Quintana (2002) proposed that identifiability is a necessary condition for the existence of an asymptotically unbiased and also consistent estimators. Hence, statistical inference is pointless if the model is not identifiable (Tabrizi, et al. , 2020a). According to the importance of identifiability in inference, before fitting a statistical model in the statistics or any related researches, it is required to check the presence of this property. As stated in Section 1, we follow the stated idea of Lindley (1972) and consider the posterior distributions to check the identifiability of target model. We describe the Bayesian form of Gaussian model as the target model of the study.

3 Bayesian Gaussian Model

In this section, we illustrate the details of the model containing the prior distributions and hyper parameters.

We consider Bayesian Gaussian model as following structure:

$$\begin{aligned} &Y \mid X, \beta, \sigma^2 \sim N\left(X\beta, \sigma^2 I\right), \\ &\beta \mid \sigma^2 \sim N\left(0, \sigma^2 I\right), \\ &\sigma^2 \sim \mathrm{Inv-Gamma}\left(\frac{1}{2}, \frac{1}{2}\right), \end{aligned} \tag{1}$$

where $Y = (Y_1, \ldots, Y_n)'$ is a vector of independent and identically distributed (iid) samples from the response variable, $X_{n\times P}$ is a matrix of observations from P co-variates, $\beta_{P\times 1}$ is the regression coefficients vector and σ^2 is the variance of Y_i and $i = 1, \ldots, n$. This structure of Bayesian Gaussian model studied detailed by Fahrmeir et al. (2009). Note that we consider the case that the sample means of the columns of X and the vector of observed response y are zero (and if not we can make it zero by a transformation). Hence, the intercept is removed from β and it becomes $\beta = \left(\beta^1, \beta^2, \ldots, \beta^P\right)'$. The hyper-parameters for the prior distribution of σ^2 are suggested by Piironen and Vehtari (2017).
As was mentioned is Section 1, we should consider the model parameters as random variables. Now, let put only a prior distribution on β and do not consider the randomness assumption about σ^2. Hence, there is no prior distribution for σ^2 in the structure of posterior density. But, σ^2 is an unknown constant parameter and should be considered in the model. By using the expressions, we can consider a special case of model (1) as

$$\begin{aligned} &Y \mid X, \beta, \sigma^2 \sim N\left(X\beta, \sigma^2\right), \\ &\beta \mid \sigma^2 \sim N\left(0, \sigma^2 I\right), \end{aligned} \tag{2}$$

where σ^2 is just an unknown parameter in the model and not considered as a random variable.

4 Results

As mentioned in Section 1, the considered model is used in the Bayesian model selection approaches. Actually, it has a certain usage in each step of any selection method. In the concept of model selection, the main aim is choosing the best model among all candidate models. Gaussian model is a widely applicable procedure in the linear models context. It is necessary to check parameters identifiability, because the basic of most selection criteria are based on the statistical inference. Therefore, if the model is unidentifiable, the selection criteria become meaningless since the identifiability of model is a necessary condition to fit or select model. Now, we prove the identifiability of the Bayesian Gaussian model.

***Theorem* 1.** *Parameters of model* (1) *are identifiable if* $Rank(X) = P$.

Proof. In model (1), assume that $(\mathrm{Y} - \mathrm{X}\beta)'(\mathrm{Y} - \mathrm{X}\beta) = \|\mathrm{Y} - \mathrm{X}\beta\|_2^2$ and $\beta'\beta = \|\beta\|_2^2$, hence the joint posterior density of (β, σ^2) becomes:

$$\pi\left(\beta, \sigma^2 \mid \mathrm{Y}, \mathrm{X}\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\|\mathrm{Y} - \mathrm{X}\beta\|_2^2}{2\sigma^2}\right) \left(\sigma^2\right)^{-\frac{P}{2}} \exp\left(-\frac{\|\beta\|_2^2}{2\sigma^2}\right) \left(\sigma^2\right)^{-\frac{3}{2}} \exp\left(-\frac{1}{2\sigma^2}\right)$$

$$\propto \left(\sigma^2\right)^{-\frac{n+P+3}{2}} \exp\left(-\frac{\|\mathrm{Y} - \mathrm{X}\beta\|_2^2 + \|\beta\|_2^2 + 1}{2\sigma^2}\right).$$

Now, if we put $a = -\frac{n+P+3}{2}$, then

$$\ln\left(\pi\left(\beta, \sigma^2 \mid \mathrm{Y}, \mathrm{X}\right)\right) \propto a \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\left(\|\mathrm{Y} - \mathrm{X}\beta\|_2^2 + \|\beta\|_2^2 + 1\right).$$

Therefore, for checking identifiability, we should consider

$$\pi\left(\beta_1, \sigma_1^2 \mid \mathrm{Y}, \mathrm{X}\right) = \pi\left(\beta_2, \sigma_2^2 \mid \mathrm{Y}, \mathrm{X}\right) \equiv \ln\left(\pi\left(\beta_1, \sigma_1^2 \mid \mathrm{Y}, \mathrm{X}\right)\right) = \ln\left(\pi\left(\beta_2, \sigma_2^2 \mid \mathrm{Y}, \mathrm{X}\right)\right), \quad (3)$$

and show $\left(\beta_1, \sigma_1^2\right) = \left(\beta_2, \sigma_2^2\right)$. For this purpose, according to (3),

$$a \ln\left(\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\left(\|\mathrm{Y} - \mathrm{X}\beta_1\|_2^2 + \|\beta_1\|_2^2 + 1\right) = a \ln\left(\sigma_2^2\right) - \frac{1}{2\sigma_2^2}\left(\|\mathrm{Y} - \mathrm{X}\beta_2\|_2^2 + \|\beta_2\|_2^2 + 1\right). \quad (4)$$

Since $Y \in \mathbb{R}^n$ is a random variable and $\mathrm{X}\beta_1 \in \mathbb{R}^n$, therefore we can put $\mathrm{Y} = \mathrm{X}\beta_1$ in (4). Hence,

$$a \ln\left(\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\left(\|\beta_1\|_2^2 + 1\right) = a \ln\left(\sigma_2^2\right) - \frac{1}{2\sigma_2^2}\left(\|\mathrm{X}\beta_1 - \mathrm{X}\beta_2\|_2^2 + \|\beta_2\|_2^2 + 1\right). \quad (5)$$

Similarly, we can put $\mathrm{Y} = \mathrm{X}\beta_2$. Hence,

$$a \ln\left(\sigma_1^2\right) - \frac{1}{2\sigma_1^2}\left(\|\mathrm{X}\beta_1 - \mathrm{X}\beta_2\|_2^2 + \|\beta_1\|_2^2 + 1\right) = a \ln\left(\sigma_2^2\right) - \frac{1}{2\sigma_2^2}\left(\|\beta_2\|_2^2 + 1\right). \quad (6)$$

By subtracting (5) from (6),

$$\left(\frac{1}{2\sigma_1^2}+\frac{1}{2\sigma_2^2}\right)\|\mathrm{X}\beta_1-\mathrm{X}\beta_2\|_2^2=0.$$

Then,

$$\|\mathrm{X}\beta_1-\mathrm{X}\beta_2\|_2^2=0,$$

and thus we can conclude that,

$$\mathrm{X}\beta_1=\mathrm{X}\beta_2.$$

Therefore, $\mathrm{X}\beta_1=\mathrm{X}\beta_2$. Since $(X'X)$ is invertible, we obtain that $\beta_1=\beta_2$. Now, for σ^2, we use the knowledge about $\beta_1=\beta_2$. We have:

$$\ln\left(\pi\left(\beta_1,\sigma_1^2\mid \mathrm{Y},\mathrm{X}\right)\right)=\ln\left(\pi\left(\beta_1,\sigma_2^2\mid \mathrm{Y},\mathrm{X}\right)\right). \tag{7}$$

Hence,

$$a\ln\left(\sigma_1^2\right)-\frac{1}{2\sigma_1^2}\left(\|\mathrm{Y}-\mathrm{X}\beta_1\|_2^2+\|\beta_1\|_2^2+1\right)=a\ln\left(\sigma_2^2\right)-\frac{1}{2\sigma_2^2}\left(\|\mathrm{Y}-\mathrm{X}\beta_1\|_2^2+\|\beta_1\|_2^2+1\right). \tag{8}$$

Assuming $b=\|\mathrm{Y}-\mathrm{X}\beta_1\|_2^2+\|\beta_1\|_2^2+1$ in (8),

$$a\ln\left(\sigma_1^2\right)-\frac{b}{2\sigma_1^2}=a\ln\left(\sigma_2^2\right)-\frac{b}{2\sigma_2^2}. \tag{9}$$

By using a logarithmic transformation,

$$\ln\frac{\ln\left(\sigma_1^2\right)}{\ln\left(\sigma_2^2\right)}+\ln\left(\frac{\sigma_1^2}{\sigma_2^2}\right)=0. \tag{10}$$

Now, let $\sigma^2\geq 1$, hence, there are three cases. $\sigma_1^2\geq\sigma_2^2$, $\sigma_1^2\leq\sigma_2^2$ or $\sigma_1^2=\sigma_2^2$. For the first case, if $\sigma_1^2>\sigma_2^2\geq 1$, then

$$\frac{\ln\left(\sigma_1^2\right)}{\ln\left(\sigma_2^2\right)}>1 \quad \& \quad \ln\left(\frac{\sigma_1^2}{\sigma_2^2}\right)>0,$$

Consequently,

$$\ln\frac{\ln\left(\sigma_1^2\right)}{\ln\left(\sigma_2^2\right)}+\ln\frac{\sigma_1^2}{\sigma_2^2}>0.$$

and it is a contradiction.
Similarly, for the second one, if $\sigma_2^2 > \sigma_1^2 \geq 1$, then

$$\ln \frac{\ln\left(\sigma_1^2\right)}{\ln\left(\sigma_2^2\right)} < 0, \ln \frac{\sigma_1^2}{\sigma_2^2} < 0.$$

Therefore,

$$\ln \frac{\ln\left(\sigma_1^2\right)}{\ln\left(\sigma_2^2\right)} + \ln \frac{\sigma_1^2}{\sigma_2^2} < 0,$$

and it is a contradiction too. Hence, we conclude that $\sigma_1^2 = \sigma_2^2 \geq 1$. On the other hand, let $\sigma^2 \leq 1$, we obtain the same results for the three aforementioned cases and we imply thT $\sigma_1^2 = \sigma_2^2 \leq 1$. Therefore, we derive that $\sigma_1^2 = \sigma_2^2$. Thus, the parameters identifiability of the Gaussian model is shown. □

Theorem 1 proved the identifiability property for a general form of Bayesian Gaussian model. If we consider model (2), then σ^2 is just an unknown constant parameter in the model and we only have the assumption of randomness and prior distribution for β in the Bayesian Gaussian model. In continue, we prove that model (2) has the identifiability property as the special case of model (1).

***Corollary* 1.** *Parameters of model* (2) *are identifiable if Rank*$(X) = P$.

Proof. For model (2), the posterior density of β become as

$$\begin{aligned}\pi\left(\beta \mid \mathrm{Y}, \mathrm{X}, \sigma^2\right) &\propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\|\mathrm{Y} - \mathrm{X}\beta\|_2^2}{2\sigma^2}\right)\left(\sigma^2\right)^{-\frac{P}{2}} \exp\left(-\frac{\|\beta\|_2^2}{2\sigma^2}\right) \\ &\propto \left(\sigma^2\right)^{-\frac{n+P}{2}} \exp\left(-\frac{\|\mathrm{Y} - \mathrm{X}\beta\|_2^2 + \|\beta\|_2^2}{2\sigma^2}\right).\end{aligned}$$

To check identifiability we have

$$\ln\left(\pi\left(\beta_1 \mid \mathrm{Y}, \mathrm{X}, \sigma_1^2\right)\right) = \ln\left(\pi\left(\beta_2 \mid \mathrm{Y}, \mathrm{X}, \sigma_2^2\right)\right).$$

Similar to the proof of Theorem 1, by putting $\mathrm{Y} = \mathrm{X}\beta_1$, $\mathrm{Y} = \mathrm{X}\beta_2$ and using the fact that $(X'X)$ is invertible, it is concluded that $\beta_1 = \beta_2$. Also, for σ^2, we obtain (10). Hence, $\sigma_1^2 = \sigma_2^2$. □

Corollary 1 implies the Gaussian model still has identifiability property even when there is no assumption about randomness of σ^2. Now, we want to

illustrate that the identifiability of Gaussian model in frequentist viewpoint can be concluded as a special case of model (1).
Kass and Raftery (1995) explained that if we consider flat priors on the model parameters, the inference based on the posterior distribution is equivalent with results via the likelihood. Now, if we put $\pi(\sigma^2) = \pi(\beta) = 1$, the posterior density of Gaussian model becomes:

$$\pi\left(\beta, \sigma^2 \mid \mathrm{Y}, \mathrm{X}\right) \propto \left(\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{\|\mathrm{Y} - \mathrm{X}\beta\|_2^2}{2\sigma^2}\right),$$

and it means that the posterior is equivalent with the likelihood of the observation. Also, similar to Theorem 1 and Corollary 1, it can be easily shown that the identifiability property is still established even if $\pi(\sigma^2) = \pi(\beta) = 1$. Therefore, we can derive that the frequentist form of Gaussian model has the identifiability property.

Conclusion

Identifiability is an essential prerequisite in statistical inference. Statistical estimation and inference are very challenging in the lack of identifiability. In this paper, some necessary conditions for identifiability in Bayesian Gaussian model and its special cases have been proposed. In Theorem 4.1 and Corollary 4.2, it is proved and elaborated that having full rank design matrix is a vital condition for identifiability. We hope that the methods used here to prove Theorem 4.1 and Corollary 4.2 will shed light on the identifiability problem of the other Bayesian statistical models. For the future studies, assessing the identifiability of the generalized linear models in a Bayesian framework is suggested.

Acknowledgment

Authors would like to thank anonymous referees for their helpful comments and hints.

References

Aldrich, J.H. and Nelson, F.D. (1984). *Linear Probability, Logit, and Probit Models*. Sage, United States.

Bahrami Samani, E. (2014). Sensitivity Analysis for the Identifiability with Application to Latent Random Effect Model for the Mixed Data. *Journal of Applied Statistics*, **41**, 2761-2776.

Christensen, R. (2011). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer Science & Business Media, USA.

Dawid, A.P. (1979). Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, **41**, 1-15.

De Leon, A.R., and Chough, K.C. (2013). *Analysis of Mixed Data: Methods & Applications*. CRC Press, United Kingdom.

Fahrmeir, L., Kneib, T., and Lang, S. (2009). *Lineare Regressionsmodelle*. Springer.

Gelfand, A.E. and Sahu, S.K. (1999). Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, **94**, 247-253.

Ghatari, A.H., and Ganjali, M. (2020). An Analysis on Covariates Selection Problem for Gaussian Model by Maximum a Posteriori Criterion Using Frequentist and Bayesian Approaches. *Journal of Advanced Mathematical Modeling*, **10**, 245-266.

Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795.

Lindley, D.V. (1972). *BayesianStatistics:AReview*. SIAM, UnitedStates.

Martín, E.S., and Quintana, F. (2002). Consistency and Identifiability Revisited, *Brazilian Journal of Probability and Statistics*, **16**, 99-106.

Miao, W., Ding, P., and Geng, Z. (2016). Identifiability of Normal and Normal Mixture Models with Nonignorable Missing Data. *Journal of the American Statistical Association*, **111**, 1673-1683.

Piironen, J., and Vehtari, A. (2017). Comparison of Bayesian Predictive Methods for Model Selection. *Statistics and Computing*, **27**, 711-735.

Tabrizi, E., Bahrami Samani, E., and Ganjali, M. (2020a). Identifiability of Parameters in Longitudinal Correlated Poisson and Inflated Beta Regression Model with Non-ignorable Missing Mechanism. *Statistics*, **54**, 524-543.

Tabrizi, E., Bahrami Samani, E., and Ganjali, M. (2020b). A Note on the Identifiability of Latent Variable Models for Mixed Longitudinal data. *Statistics & Probability Letters*, **167**, 1-5.

Vehtari, A., and Ojanen, J. (2012). A Survey of Bayesian Predictive Methods for Model Assessment, Selection and Comparison. *Statistics Surveys*, **6**, 142-228.

Wang, W. (2013). Identifiability of Linear Mixed Effects Models. *Institute of Mathematical Statistics and Bernoulli Society*, **7**, 244-263.

Yu, Q., and Dong, J. (2020). Identifiability Conditions for the Linear Regression Model Under Right Censoring, *Communications in Statistics-Theory and Methods*, 1-19.

Amir Hossein Ghatari
Department of Statistics,
Faculty of Mathematics and Computer Science,
Amirkabir University of Technology,
Tehran, Iran.
email: *a.h.ghatari@aut.ac.ir*

Ashkan Shabbak
Department of Data Processing,
Statistical Research and Training Center,
Tehran, Iran.
email: *a.shabbak@gmail.com*

Elham Tabrizi
Department of Mathematics,
Faculty of Mathematics and Computer Science,
Kharazmi University,
Tehran, Iran.
email: *tabrizi.elham@gmail.com*