Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
Vol. 28 (2025), No. 1, pp. 1–8
DOI : 10.47974/JIM-1766

Variational Bayes analysis of the normal-gamma-exponential prior

Noor Hassan Hussein *
Gorgees Shaheed Mohammad †
Department of Mathematics
College of Education
University of AL-Qadisiyah
AL-Qadisiyah 58002
Iraq

Abstract

The Variational Bayes (VB) method for the Normal-Gamma-Exponential (NGE) prior is derived. The relation between the VB method and the regular Gibbs sampler is shown. The factorized approximation tool is used to expand the VB method to the Mean Field Variational Bayes (MFVB) method. The predication ability and theoretical properties of this method with the NGE prior are demonstrated using simulated data and its result is compared with other methods showing that this method preforms reasonably well.

Subject Classification: *62N01, 62E20, 62B05.*

Keywords: *Variational bayes, Gamma-prior prior, Scale-mixture of normal.*

1. Introduction

The usual Monte Carlo Expectation Maximization (MCEM) method is one of a family of Maximum Likelihood (ML) methods. These methods have proven their usefulness and popularity in modern statistical analysis, especially since the concept behind this method is very natural and intuitive. This method uses iterative algorithm for ML estimation and while it has several advantages, it still suffers from some limitations as the complexity of the problem increases [1,2,4-10]. In recent years, there has being increasing effort to search for some alternative methods reduce some of the requirements that limit the ability of this method.

* *E-mail:* **noorhassn564@gmail.com (Corresponding Author)**
† *E-mail:* **gorgees.alsalamy@qu.edu.iq**

Recently, a new method has being introduced, which has been gaining popularity rapidly, known as "variational Bayesian inference" or "variational Bayes". Another name for this new method is the "*Variational Approximation*" since it can be used to approximate complex models in areas where the EM method doesn't work such as one the number of observations or the number of covaries becomes large. Interestingly, one can show that the EM method is a special case of this method.

In this paper, the following hierarchal model of the Normal-Gamma-Exponential (NGE) prior will be considered

$$\begin{aligned} \beta_i|\zeta_i^2, \eta_i^2, \sigma^2 &\sim N(0, \sigma^2\zeta_i^2\eta_i^2), \\ \zeta_i^2 &\sim \mathcal{G}(\alpha, 1), \\ \eta_i^2 &\sim \mathcal{IG}(1, \lambda), \\ \sigma^2 &\sim 1/\sigma^2 \end{aligned} \tag{1}$$

where we have written the Gamma-Exponential prior as the product of the Gamma and Inverse-Gamma distributions. Now, we can write the log-likelihood as

$$log\, p(y; \omega) = H(q, \omega) + KL(q||p) \tag{2}$$

with

$$H(q, \omega) = \int q(\beta, \eta^2, \zeta^2, \sigma^2)\, log\left(\frac{p(y, \beta, \eta^2, \zeta^2, \sigma^2; \omega)}{q(\beta, \eta^2, \zeta^2, \sigma^2)}\right) d(\beta, \eta^2, \zeta^2, \sigma^2) \tag{3}$$

and

$$KL(q||p) = -\int q(\beta, \eta^2, \zeta^2, \sigma^2)\, log\left(\frac{p(\beta, \eta^2, \zeta^2, \sigma^2|y; \omega)}{q(\beta, \eta^2, \zeta^2, \sigma^2)}\right) d(\beta, \eta^2, \zeta^2, \sigma^2) \tag{4}$$

where q is some probability distribution and $KL(q||p)$ is the Kullback-Leibler divergence [2] between $p(\beta, \eta^2, \zeta^2, \sigma^2|y; \omega)$ and $q(\beta, \eta^2, \zeta^2, \sigma^2)$ which can be thought of as a measure of the difference between the two distributions.

One of the properties of Kullback-Leibler divergence is that it is always positive and hence

$$KL(q||p) \geq 0. \tag{5}$$

Thus, we have

$$log\, p(y; \omega) \geq H(q, \omega). \tag{2}$$

From this, we can conclude that the term $H(q, \omega)$ is a lower bound of $log\, p(y; \omega)$. Additionally, when the Kullback-Leibler divergence vanishes, i.e. when there is no difference between q and p then $log\, p(y; \omega) = H(q, \omega)$.

In order to see connection between this method and the MCEM equations, we will maximize the lower bound $H(q, \omega)$ with respect to q with fixing the value of ω. Obviously, this happens when

$$p(\beta, \eta^2, \zeta^2, \sigma^2|y; \omega^{fixed}) = q(z) \tag{7}$$

which is when the Kullback-Leibler divergence is zero. If we insert the above equation in (**Error! Reference source not found.**), we have

$$H(q,\omega) = \int p(\beta,\eta^2,\zeta^2,\sigma^2|y;\omega^{fixed}) \times log\left(\frac{p(y,\beta,\eta^2,\zeta^2,\sigma^2;\omega)}{p(\beta,\eta^2,\zeta^2,\sigma^2|y;\omega^{fixed})}\right) d(\beta,\eta^2,\zeta^2,\sigma^2) \tag{8}$$

By using the logarithmic properties, we have

$$H(q,\omega) = \int p(\beta,\eta^2,\zeta^2,\sigma^2|y;\omega^{fixed})\, log\big(p(y,\beta,\eta^2,\zeta^2,\sigma^2;\omega)\big)\, d(\beta,\eta^2,\zeta^2,\sigma^2)$$
$$-\int p(\beta,\eta^2,\zeta^2,\sigma^2|y;\omega^{fixed})\, log\left(p(\beta,\eta^2,\zeta^2,\sigma^2|y;\omega^{fixed})\right) d(\beta,\eta^2,\zeta^2,\sigma^2) \tag{9}$$

By noticing that second term does not depend on ω and comparing this equation with MCEM formulation, we can see that

$$H(q,\omega) = Q(\omega,\omega^{fixed}) + constant \tag{10}$$

where the log of the complete-data likelihood $Q(\omega,\omega^{fixed})$ is given by

$$Q(\omega,\omega^{k-1}) = -p\, log\big(\Gamma(\alpha)\big) + \alpha \sum_{i=1}^{p} \mathbb{E}_{\alpha^{k-1}}[log\,(\zeta_i^2)\,|y] + p\, log(\lambda) - \lambda \sum_{i=1}^{p} \mathbb{E}_{\alpha^{k-1}}[\eta_i^{-2}|y] + \text{terms not containing } \alpha \text{ and } \lambda \tag{11}$$

with $\omega = (\alpha,\lambda)$. Therefore, it is clear to see the connection between $Q(\omega,\omega^{fixed})$ and $H(q,\omega)$ and thus the relationship between the variational Bayes and the MCEM methods.

One particular useful tool of this method is factorized approximation tool [2]. This works by assuming we can write q as

$$q(\beta,\eta^2,\zeta^2,\sigma^2) \approx q_1^*(\beta) q_2^*(\eta^2) q_3^*(\zeta^2) q_4^*(\sigma^2) \tag{12}$$

with

$$q_1^*(\beta) \sim N(\mu^*, V^*) \tag{13}$$

$$q_2^*(\eta^2) \sim GIG\left(\chi_1^*, 2, \alpha - \tfrac{1}{2}\right) \tag{14}$$

$$q_3^*(\zeta^2) \sim IG\left(\tfrac{3}{2}, \chi_2^*\right) \tag{15}$$

and

$$q_4^*(\sigma^2) \sim IG\left(\tfrac{n+p}{2}, \chi_3^*\right) \tag{16}$$

with

$$\mu^* = (X^TX + \boldsymbol{M}^*)^{-1}X^Ty \tag{17}$$

$$V^* = E_{q_4^*}(\sigma^2)(X^TX + \boldsymbol{M}^*)^{-1} \tag{18}$$

$$M^* = diag\left(E_{q_2^*}(\eta_1^{-2})E_{q_3^*}(\zeta_1^{-2}), \dots, E_{q_2^*}(\eta_p^{-2})E_{q_3^*}(\zeta_p^{-2})\right) \tag{19}$$

$$\chi_1^* = E_{q_1^*}(\beta^2)E_{q_4^*}(\sigma^{-2})\, E_{q_3^*}(\zeta_i^{-2}) \tag{20}$$

$$\chi_2^* = \tfrac{1}{2} E_{q_1^*}(\beta^2)E_{q_4^*}(\sigma^{-2})E_{q_2^*}(\eta_i^{-2}) + \lambda \tag{21}$$

and

$$\chi_3^* = \frac{E_{q_1^*}(\|y - X\beta\|_2^2) + E_{q^*}(\beta^T M^* \beta)}{2} \tag{22}$$

where we calculate the expectations

$$E_{q_1^*}(\beta_i^2) = \mu_i^{*2} + V_{ii}^*$$
$$E_{q_1^*}(\|y - X\beta\|_2^2) = \|y - X\mu^*\|_2^2 + tr(X^T X V^*) \tag{23}$$
$$E_{q^*}(X^T M^* X) = \sum_{i=1}^{p} \beta_i^{*2}\, E_{q_2^*}(\eta_i^{-2}) E_{q_3^*}(\zeta_i^{-2}) + tr(M^* V^*)$$

Thus, we can calculate the lower bound as

$$L = E_q\, log\big(p(y, \beta, \eta^2, \zeta^2, \sigma^2)\big) - E_q\, log\big(q(\beta, \eta^2, \zeta^2, \sigma^2)\big) \tag{24}$$

Substituting the joint density in the last equation, we get

$$L = p\, log\, \lambda - \frac{n}{2} log\, 2\pi - \frac{n+p}{2} log\, \sigma^2 - \sum_{i=1}^{p} log \left[\frac{(\chi_1^*/2)^{\frac{2\alpha-1}{4}}}{K_{\alpha-\frac{1}{2}}(2\chi_1^*)} \right]$$
$$- \frac{3}{2} \sum_{i=1}^{p} log\, \chi_{2i}^* + \frac{1}{2} log\, |V^*| + p\, log \frac{3}{2} - p\, log\, \alpha + \sum_{i=1}^{p} E_{q_2^*}(\zeta_i^{-2})$$
$$+ \sum_{i=1}^{p} \left(\frac{\chi_{1i}^*}{2} - 1 \right) E_{q_2^*}(\zeta_i^2) + \sum_{i=1}^{p} (\chi_{2i}^* - \lambda) E_{q_3^*}(\eta_i^{-2}) -$$
$$\frac{p+n}{2} log\, \chi_3^* + log \frac{p+n}{2} + \chi_3^* E_{q_4^*}(\sigma^{-2}) + \frac{p}{2} \tag{25}$$

where $K_x'(\cdot)$ is the modified Bessel function of the second kind of the Generalized inverse Gaussian distribution. Finally, we can incorporate the EM algorithm to the variational inference method using the Mean Field Variational Bayes (MFVB) with

$$E_{q_2^{*(k-1)}, \alpha^{(k-1)}}[log(\zeta_i^2)] = \frac{1}{2} log \left(\frac{\chi_1^{*(k-1)}}{2} \right) + \frac{\partial log\left(K_{\alpha^{k-1}-1/2}\left(\sqrt{2\chi_1^{*(k-1)}} \right) \right)}{\partial \alpha^{k-1}} \tag{26}$$

And

$$E_{q_3^{*(k-1)}, \lambda^{(k-1)}}[log(\eta_i^2)] = log\, \chi_2^{*(k-1)} - \psi\left(\frac{3}{2}\right) \tag{27}$$

We can now use this to simultaneously evaluate our variables while at the same time update our hyperparamters.

2. Experimental Studies

In this section, we will use simulated data to test how our model preforms compared to other models such as the least absolute shrinkage and selection operator (LASSO, [3,10]), the elastic net penalty (ENet, [13]), the horseshoe estimator (HS, [8]), the minimax concave penalty (MCP, [12]) and smoothly clipped absolute deviation (SCAD, [9,11,12,14-23]). In addition, we will use three parameters of comparison such as the mean squared error (MSE), standard deviation (SD) and the false positive rate FPR with 10000 repetitions with 1400 burnins.

Firstly, we will analyze the performance of the MFVB method in comparison with different sizes of data by setting $\beta = (7.5,0,0,2,0,4,0,1)$ and $\sigma^2 = 1$. The results are shown in Table 1. We notice that the MFVB algorithm preforms better with larger data sizes by given smaller MSE and MSEsd parameters.

Secondly, in order to give a more deeper study of the NGE prior using both sparse and dense models by setting $\beta = (5,0,0,0,0,3,0,0)$ and $\beta = (3,0,4,1,7,0,1,2)$ respectively. The results are shown in Tables 2 and 3. It is very obvious from these tables that none of the methods considered performs better than our model.

Table 1
Simulation results for the MFVB method for different sizes of data.

Methods	n	MSE (sd)	FPR (FPRsd)	FNR (FNRsd)
NGE (MFVB)	25	0.3375 (0.1703)	7.0000 (0.0000)	0.0000 (0.0000)
NGE (MFVB)	50	0.2333 (0.0794)	4.0000 (0.0000)	0.0000 (0.0000)
NGE (MFVB)	100	0.6756 (0.1952)	3.0000 (0.0000)	0.0000 (0.0000)
NGE (MFVB)	200	0.0591 (0.0329)	0.0000 (0.0000)	0.0000 (0.0000)
NGE (MCEM)	250	0.0350 (0.0142)	0.0000 (0.0000)	0.0000 (0.0000)

Table 2
Results for model with few active variables

Methods	MSE (sd)	FPR (FPRsd)	FNR (FNRsd)
NGE (MFVB)	0.0812 (0.0771)	0.0000 (0.0000)	0.0110 (0.0210)
LASSO	0.0820 (0.0960)	0.2000 (0.6325)	0.0000 (0.0000)
ENet	0.1105 (0.1116)	0.6000 (1.2649)	0.0000 (0.0000)
HS	0.2277 (0.1048)	0.7000 (1.0593)	0.0000 (0.0000)
MCP	0.1122 (0.0855)	0.0000 (0.0000)	0.0000 (0.0000)
SCAD	0.0967 (0.0757)	1.4000 (1.5055)	0.0000 (0.0000)

Table 3
Results for model with most variables are active.

Methods	MSE (sd)	FPR (FPRsd)	FNR (FNRsd)
NGE (MFVB)	0.2599 (0.2050)	0.5000 (0.7780)	0.0000 (0.0000)
LASSO	0.2780 (0.1658)	0.4000 (0.8433)	0.0000 (0.0000)
ENet	0.2661 (0.1702)	0.3000 (0.6749)	0.0000 (0.0000)
HS	0.5276 (0.2126)	0.9000 (0.7379)	0.0000 (0.0000)
MCP	0.2956 (0.1641)	0.0000 (0.0000)	0.0000 (0.0000)
SCAD	0.3271 (0.1707)	1.3000 (0.4830)	0.0000 (0.0000)

3. Conclusions

We have provided an analysis of how the NGE prior compares with others. To do this, we have provided and alternative algorithm from the familiar Expectation-Maximaztion (EM) methods. We have used the factorized approximation to construct the Mean Field Variational Bayes (MFVB) method which is easier to deal with this type of model. By comparing our model with other models, we have showed how in several scenarios our model has better predication result.

References

[1] A. Alhamzawi and G. S. Mohammad, "Statistical inference with normal-compound gamma priors in regression models," arXiv preprint arXiv:2212.13144 (2022). [Online]. Available: https://doi.org/10.48550/arXiv.2212.13144.

[2] R. Bai and M. Ghosh, "On the Beta prime prior for scale parameters in high-dimensional Bayesian regression models," *Statistica Sinica,* vol. 31, no. 2, pp. 843-865 (2021). [Online]. Available: https://doi.org/10.5705/ss.202019.0037.

[3] C. Faes, J. T. Ormerod, and M. P. Wand, "Variational Bayesian inference for parametric and nonparametric regression with missing data," *Journal of the American Statistical Association,* vol. 106, no. 495, pp. 959-971 (2011). [Online]. Available: https://doi.org/10.1198/jasa.2011.tm10301.

[4] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348-1360 (2001). [Online]. Available: https://doi.org/10.1198/016214501753382273.

[5] H. Jeffreys, "An invariant form for the prior probability in estimation problems," Proceedings of the Royal Society of London. Series A. *Mathematical and Physical Sciences,* vol. 186, no. 1007, pp. 453-461 (1946). [Online]. Available: https://doi.org/10.1098/rspa.1946.0056.

[6] R. A. Levine and G. Casella, "Implementations of the Monte Carlo EM algorithm," *Journal of Computational and Graphical Statistics,* vol. 10, no. 3, pp. 422-439 (2001). [Online]. Available: https://doi.org/10.1198/106186001317115045.

[7] H. Mallick, R. Alhamzawi, E. Paul, and V. Svetnik, "The reciprocal Bayesian lasso," *Statistics in Medicine,* vol. 40, no. 22, pp. 4830-4849 (2021). [Online]. Available: https://doi.org/10.1002/sim.9098.

[8] T. Park and G. Casella, "The Bayesian lasso," *Journal of the American Statistical Association,* vol. 103, no. 482, pp. 681-686 (2008). [Online]. Available: https://doi.org/10.1198/016214508000000337.

[9] N. G. Polson, J. G. Scott, and J. Windle, “The Bayesian bridge,” *Journal of the Royal Statistical Society: Statistical Methodology,* vol. 76, no. 4, pp. 713-733 (2014). [Online]. Available: https://doi.org/10.1111/rssb.12042.

[10] S. M. Kadham, “Acute interstitial pneumonia image enhancement using fuzzy partial transforms,” *Applied Geomatics,* vol. 16, no. 1, pp. 35-39 (2024).

[11] Q. Song and F. Liang, “Nearly optimal Bayesian shrinkage for high-dimensional regression,” *Science China Mathematics,* vol. 66, no. 2, pp. 409-442 (2023). [Online]. Available: https://doi.org/10.48550/arXiv.1712.08964.

[12] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society,* Series B, vol. 58, pp. 267-288 (1996).

[13] H. Woods, H. H. Steinour, and H. R. Starke, “Effect of composition of Portland cement on heat evolved during hardening,” *Industrial & Engineering Chemistry*, vol. 24, no. 11, pp. 1207-1214 (1932). [Online]. Available: https://doi.org/10.1021/ie50275a002.

[14] A. M. Darweesh, R. H. Hasan and S. M. Kadham, Results of fourth-order differential superordination and subordination for univalent functions defined by integral operator. *Mathematical Modelling of Engineering Problems*, Vol. 11, No. 2, pp. 385-392 (2024). https://doi.org/10.18280/mmep.110210

[15] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society,* Series B, vol. 67, pp. 301-320 (2005). [Online]. Available: https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[16] W. Gao, L. Shi, and M. R. Farahani, “Szeged Related Indices of TUAC6[p, q],” *J. Discrete Math. Sci. Cryptography,* vol. 20, no. 2, pp. 553-563 (2017). [Online]. Available: https://doi.org/10.1080/09720529.2016.1228312.

[17] S. M. Kadham and A. N. Alkiffai, “Generalization of Fuzzy SHA-Transform with Medical Application,” *Mathematical Modelling of Engineering Problems*, vol. 9, no. 5, pp. 1206-1214 (2022).

[18] F. Afzal, S. Rukh, D. Afzal, M. R. Farahani, M. Cancan, and S. Ediz, "A study of non-commutative Von-Neumann regular rings," *J. Inform. Optim. Sci.,* vol. 42, no. 7, pp. 1425-1436 (2021). [Online]. Available: https://doi.org/10.1080/02522667.2021.1896650.

[19] M. S. Sardar, M. Alaeiyan, M. R. Farahani, M. Cancan, and S. Ediz, "Resistance distance in some classes of rooted product graphs obtained by Laplacian generalized inverse method," *J. Inform. Optim. Sci.,* vol. 42, no. 7, pp. 1447-1467 (2021). [Online]. Available: https://doi.org/10.1080/02522667.2021.1899210.

[20] M. Nadeem, S. Ahmad, M. K. Siddiqui, M. A. Ali, M. R. Farahani, and A. J. M. Khalaf, "On some applications related with algebraic structures through different well-known graphs," *J. Discrete Math. Sci. Cryptography,* vol. 24, no. 2, pp. 451-471 (2021). [Online]. Available: https://doi.org/10.1080/09720529.2021.1885806.

[21] S. M. Kadham, "Acute interstitial pneumonia image enhancement using fuzzy partial transforms," *Applied Geomatics*, vol. 16, no. 1, pp. 35–39 (May 2023), doi: https://doi.org/10.1007/s12518-023-00509-8.

[22] A.M. Darweesh, R.H. Hasan, and S.M. Kadham, "Results of Fourth-Order Differential Superordination and Subordination for Univalent Functions Defined by Integral Operator," *Mathematical Modelling and Engineering Problems*, vol. 11, no. 2, pp. 385–392 (Feb. 2024), doi: https://doi.org/10.18280/mmep.110210.

[22] S. M. Kadham and A. N. Alkiffai, "Generalization of Fuzzy SHA-Transform with Medical Application," *Mathematical Modelling of Engineering Problems*, vol. 9, no. 5, pp. 1378–1384 (Dec. 2022), doi: https://doi.org/10.18280/mmep.090528.

Received June, 2024