BIOINFORMATICS LETTER TO THE EDITOR Vol. 27 no. 15 2011, pages 2169–2170
doi:10.1093/bioinformatics/btr334

Gene expression Advance Access publication June 22, 2011

Response to "Comments on 'Bayesian variable selection for disease classification using gene expression data'"

Xin-Yuan Song* and Zhao-Hua Lu
Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Associate Editor: Alfonso Valencia

Contact: xysong@sta.cuhk.edu.hk

Received on April 6, 2011; revised on May 17, 2011; accepted on May 30, 2011

This letter is a response to the comments of Baragatti and Pommeret (2011) on Yang and Song (2010a) in *Bioinformatics*. Baragatti and Pommeret (2011) pointed out that in the case where the covariance matrix of the g-prior (Zellner, 1986) is singular, the computation of the posterior distributions proposed by Yang and Song (2010a) has a technical issue. In this letter, we provide a solution to address this issue.

Following notations in Yang and Song (2010a), Y is an n-dimensional vector of binary observed random variables, Z is an $n\times 1$ vector of the underlying continuous latent variables, $\mathbf{X}_\gamma$ is an $n\times p_\gamma$ matrix of selected covariates, α is an intercept, β_γ is a $p_\gamma\times 1$ vector of the corresponding regression coefficients and p_γ is the number of selected covariates. The g-prior (see Zellner, 1986) for β_γ is $N(0,c(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1})$, where c is a specified value. If $n<p_\gamma$, $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is not a full rank matrix and $(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}$ does not exist. Moreover, $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is nearly singular when highly collinear covariates exist. In these cases, Yang and Song (2010a) proposed a modified form of the g-prior, namely the generalized singular g-prior (gsg-prior), in which the matrix $(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}$ was replaced by the Moore–Penrose generalized inverse $(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{+}$. Under the model and prior specifications, the joint posterior distribution $p(Z,\alpha,\beta_\gamma,\gamma|Y,\mathbf{X})$ can be derived; see Equation (7) in Yang and Song (2010a). However, directly simulating observations from this joint posterior distribution is infeasible due to its complexity. To simplify the computation, Yang and Song (2010a) integrated α and β_γ out from $p(Z,\alpha,\beta_\gamma,\gamma|Y,\mathbf{X})$, and obtained the joint posterior distribution of (Z,γ) as follows:

$$p(Z,\gamma|Y,\mathbf{X})\propto \frac{1}{|\Sigma_\gamma|^{\frac{1}{2}}}\exp\left(-\frac{Z'\Sigma_\gamma^{-1}Z}{2}\right)\prod_{i=1}^{n}\mathrm{I}(A_i)\times\prod_{i=1}^{p}\pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}, \tag{1}$$

where $\Sigma_\gamma=\mathbf{I_n}+h11'+c\mathbf{X}_\gamma(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{+}\mathbf{X}_\gamma'$. When integrating out β_γ, the inverse of the matrix $A=\mathbf{X}_\gamma'\{(\mathbf{I}_n+h\mathbf{11}')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{X}_\gamma$ is used in their Equation (A7) (see Yang and Song, 2010b). The technical issue pointed out by Baragatti and Pommeret (2011) is about the infeasibility of taking inverse of the matrix A if $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is singular.

In the following, we show that A^{-1} is not necessary in the derivation of $p(Z,\gamma|Y,\mathbf{X})$, and thus $p(Z,\gamma|Y,\mathbf{X})$ in Yang and Song (2010a) is still valid when $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is singular.

Let $m_\gamma=rank(\mathbf{X}_\gamma)\leq\min(n,p_\gamma)$, $\mathbf{X}_\gamma'=\mathbf{QVD}$ be the singular value decomposition (SVD) of $\mathbf{X}_\gamma'$, where $\mathbf{Q}$ and $\mathbf{D}$ are $p_\gamma\times p_\gamma$ and $n\times n$ orthogonal matrices, respectively, and $\mathbf{V}$ is a $p_\gamma\times n$ diagonal matrix, the first m_γ diagonal elements of which are positive. The SVD can be expressed as

$$\mathbf{X}_\gamma'=\mathbf{QVD}=\mathbf{QR}=\left(\mathbf{Q}_1\ \mathbf{Q}_2\right)\begin{pmatrix}\mathbf{R}_1\\ \mathbf{0}\end{pmatrix}=\mathbf{Q}_1\mathbf{R}_1, \tag{2}$$

where $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are $p_\gamma\times m_\gamma$ and $p_\gamma\times(p_\gamma-m_\gamma)$ matrices, respectively, and $\mathbf{R}_1$ is an $m_\gamma\times n$ matrix with full row rank. When $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is singular, inspired by the Theorem 1 in Panagiotelis and Smith (2008), we modify the gsg-prior to make it proper:

$$\beta_\gamma\sim N(\mathbf{0},c(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{+})I(\mathbf{Q}_2'\beta_\gamma=0), \tag{3}$$

the density of which can be expressed as:

$$p(\beta_\gamma|\gamma)=(2\pi c)^{-\frac{m_\gamma}{2}}\prod_{i=1}^{m_\gamma}\lambda_i^{-\frac{1}{2}}\exp\left(-\frac{\beta_\gamma'\mathbf{X}_\gamma'\mathbf{X}_\gamma\beta_\gamma}{2c}\right)I(\mathbf{Q}_2'\beta_\gamma=0), \tag{4}$$

where $\lambda_1,\ldots,\lambda_{m_\gamma}$ are the non-zero eigenvalues of $(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{+}$. $\mathbf{Q}_2$ contains the eigenvectors corresponding to the null-space of $\mathbf{X}_\gamma'\mathbf{X}_\gamma$. By adding a point mass at zero in (3) or (4), we restrict β_γ on the reduced space over which the covariates set provides information. We will show that this restriction solves the problem of integrating out β_γ when $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is singular. We also show that this restriction is only for the purpose of integrating α and β_γ out from $p(Z,\alpha,\beta_\gamma,\gamma|Y,\mathbf{X})$, and it is not involved in the subsequent MCMC algorithm. Therefore, the restricted gsg-prior in (3) or (4) is a natural and realistic choice in the case of singularity of $\mathbf{X}_\gamma'\mathbf{X}_\gamma$. If $\mathbf{X}_\gamma'\mathbf{X}_\gamma$ is invertible, $\mathbf{Q}_2$ does not exist, the restriction in (3) or (4) disappears, leading to the gsg-prior proposed in Yang and Song (2010a). Let $\tilde{\beta}_\gamma=\mathbf{Q}_1'\beta_\gamma$, $\tilde{\tilde{\beta}}_\gamma=\mathbf{Q}_2'\beta_\gamma$, and $\beta_\gamma^*=(\tilde{\beta}_\gamma',\tilde{\tilde{\beta}}_\gamma')'=\mathbf{Q}'\beta_\gamma$. The prior (4) can be induced by the following prior for β_γ^*:

$$\begin{aligned} p(\beta_\gamma^*|\gamma)&=p(\tilde{\beta}_\gamma|\gamma)p(\tilde{\tilde{\beta}}_\gamma|\gamma) \\ &=\left[(2\pi c)^{-\frac{m_\gamma}{2}}\prod_{i=1}^{m_\gamma}\lambda_i^{-\frac{1}{2}}\exp\left(-\frac{\tilde{\beta}_\gamma'\mathbf{R}_1\mathbf{R}_1'\tilde{\beta}_\gamma}{2c}\right)\right]\times I(\tilde{\tilde{\beta}}_\gamma=0), \end{aligned} \tag{5}$$

where the distribution of $\tilde{\beta}_\gamma$ and $\tilde{\tilde{\beta}}_\gamma$, given γ are $N(0,c(\mathbf{R}_1\mathbf{R}_1')^{-1})$ and a degenerate distribution with a point mass density on $\tilde{\tilde{\beta}}_\gamma=0$, respectively (see Panagiotelis and Smith, 2008).

*To whom correspondence should be addressed.

In implementation of the MCMC algorithm, Yang and Song (2010a) first integrated α and β_γ out from the joint posterior distribution $p(Z,\alpha,\beta_\gamma,\gamma|Y,\mathbf{X})$, and then sampled from the marginal posterior distribution $p(Z,\gamma|Y,\mathbf{X})$ with Gibbs sampler. In the following, we show that the same marginal posterior distribution $p(Z,\gamma|Y,\mathbf{X})$ can be derived when $\mathbf{X}'_\gamma\mathbf{X}_\gamma$ is singular. To do so, we transform β_γ to β^*_γ as follows:

$$
\begin{aligned}
p(Z,\gamma|Y,\mathbf{X}) &= \iint p(Z,\alpha,\beta_\gamma,\gamma|Y,\mathbf{X})d\alpha d\beta_\gamma \\
&= \iint p(Z,\alpha,\beta^*_\gamma,\gamma|Y,\mathbf{X})\left|\frac{d\beta_\gamma}{d\beta^*_\gamma}\right|d\alpha d\beta^*_\gamma \quad \left(\left|\frac{d\beta_\gamma}{d\beta^*_\gamma}\right|=|\mathbf{Q}|=1\right) \\
&= \iiint p(Z,\alpha,\tilde{\beta}_\gamma,\tilde{\tilde{\beta}}_\gamma,\gamma|Y,\mathbf{X})d\alpha d\tilde{\beta}_\gamma d\tilde{\tilde{\beta}}_\gamma, \qquad (6)
\end{aligned}
$$

where

$$
\begin{aligned}
&p(Z,\alpha,\tilde{\beta}_\gamma,\tilde{\tilde{\beta}}_\gamma,\gamma|Y,\mathbf{X}) \\
&\propto \left[\exp\left\{-\frac{(Z-\alpha 1-\mathbf{R}'_1\tilde{\beta}_\gamma)'(Z-\alpha 1-\mathbf{R}'_1\tilde{\beta}_\gamma)}{2}\right\}\prod_{i=1}^{n} I(A_i)\right] \\
&\times \exp\left(-\frac{\alpha^2}{2h}\right)\times\left[\exp\left(-\frac{\tilde{\beta}'_\gamma\mathbf{R}_1\mathbf{R}'_1\tilde{\beta}_\gamma}{2c}\right)I(\tilde{\tilde{\beta}}_\gamma=0)\prod_{i=1}^{m_\gamma}\lambda_i^{-1/2}\right] \\
&\times\prod_{i=1}^{p}\pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i}. \qquad (7)
\end{aligned}
$$

From (7), $\tilde{\tilde{\beta}}_\gamma$ is integrated out immediately. According to the same reasonings given in Yang and Song (2010a), α can be integrated out by forming a Gaussian probability density of α in (7), and the remaining exponentiated term in (7) can then be rewritten as

$$
\begin{aligned}
&-\frac{\tilde{\beta}'_\gamma\tilde{A}\tilde{\beta}_\gamma-2\tilde{\beta}'_\gamma\tilde{B}}{2}-\frac{Z'(\mathbf{I}_n+h11')Z}{2} \\
&= -\frac{(\tilde{\beta}_\gamma-\tilde{A}^{-1}\tilde{B})'\tilde{A}(\tilde{\beta}_\gamma-\tilde{A}^{-1}\tilde{B})}{2}- \\
&\quad\frac{Z'(\mathbf{I}_n+h11')Z-\tilde{B}'\tilde{A}^{-1}\tilde{B}}{2}, \qquad (8)
\end{aligned}
$$

where $\tilde{A}=\mathbf{R}_1\{(\mathbf{I}_n+h11')^{-1}+c^{-1}\mathbf{I}_n\}\mathbf{R}'_1$, and $\tilde{B}=\mathbf{R}_1(\mathbf{I}_n+h11')^{-1}Z$. The above expression (8) is similar to (A7) in Yang and Song (2010b) but here $\tilde{A}$ is invertible. The first term in (8) is a completed quadratic form of $\tilde{\beta}_\gamma$, which forms a gaussian probability density and can be integrated out. The second term in (8) forms the kernel of a posterior probability density of $Z|\mathbf{X},\gamma$ as $-Z'\tilde{\Sigma}_\gamma^{-1}Z/2$, where $\tilde{\Sigma}_\gamma^{-1}=(\mathbf{I}+h11')^{-1}-(\mathbf{I}_n+h11')^{-1}\mathbf{R}'_1\tilde{A}^{-1}\mathbf{R}_1(\mathbf{I}_n+h11')^{-1}$. Similar to the derivation in Yang and Song (2010b), we can prove that

$$
\begin{aligned}
&p(Z|\mathbf{X},\gamma)\sim N(0,\tilde{\Sigma}_\gamma), \quad \text{and} \\
&p(Z,\gamma|Y,\mathbf{X})\propto\frac{1}{|\tilde{\Sigma}_\gamma|^{1/2}}\exp\left(-\frac{Z'\tilde{\Sigma}_\gamma^{-1}Z}{2}\right)\prod_{i=1}^{n}I(A_i) \\
&\qquad\times\prod_{i=1}^{p}\pi_i^{\gamma_i}(1-\pi_i)^{1-\gamma_i},
\end{aligned}
$$

where $\tilde{\Sigma}_\gamma=\mathbf{I}_n+h11'+c\mathbf{R}'_1(\mathbf{R}_1\mathbf{R}'_1)^+\mathbf{R}_1$.

In the following, we further show that $\tilde{\Sigma}_\gamma=\Sigma_\gamma$. Thus, the marginal posterior distribution $p(Z,\gamma|Y,\mathbf{X})$ in Yang and Song (2010a) is unchanged even if $\mathbf{X}'_\gamma\mathbf{X}_\gamma$ is singular.

Proposition 1. *If $\mathbf{S}$ is a symmetric matrix and $\mathbf{Q}$ is an orthogonal matrix of appropriate dimensions, then* $(\mathbf{QSQ}')^+=\mathbf{QS}^+\mathbf{Q}'$.

Proposition 2.

$$
(\mathbf{RR}')^+=\begin{pmatrix}(\mathbf{R}_1\mathbf{R}'_1)^+ & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}.
$$

These propositions can be verified easily by using the definition of the Moore–Penrose inverse.

Theorem 1. $\mathbf{X}_\gamma(\mathbf{X}'_\gamma\mathbf{X}_\gamma)^+\mathbf{X}'_\gamma=\mathbf{R}'_1(\mathbf{R}_1\mathbf{R}'_1)^+\mathbf{R}_1$.

Proof.

$$
\begin{aligned}
\mathbf{X}_\gamma(\mathbf{X}'_\gamma\mathbf{X}_\gamma)^+\mathbf{X}'_\gamma &= \mathbf{R}'\mathbf{Q}'(\mathbf{QRR}'\mathbf{Q}')^+\mathbf{QR} \\
&= \mathbf{R}'(\mathbf{RR}')^+\mathbf{R} \\
&= \mathbf{R}'_1(\mathbf{R}_1\mathbf{R}'_1)^+\mathbf{R}_1.
\end{aligned}
$$

Based on Theorem 1,

$$
\begin{aligned}
\tilde{\Sigma}_\gamma &= \mathbf{I}_n+h11'+c\mathbf{R}'_1(\mathbf{R}_1\mathbf{R}'_1)^+\mathbf{R}_1 \\
&= \mathbf{I}_n+h11'+c\mathbf{X}'_\gamma(\mathbf{X}_\gamma\mathbf{X}'_\gamma)^+\mathbf{X}_\gamma=\Sigma_\gamma.
\end{aligned}
$$

We are very thankful to Baragatti and Pommeret for pointing out the technical issue when integrating out β_γ in the case of singularity of $\mathbf{X}'_\gamma\mathbf{X}_\gamma$. In this response, we have shown that the technical issue can be addressed by slightly modifying the gsg-prior in Yang and Song (2010a), and that this minor modification does not change the marginal posterior distribution $p(Z,\gamma|Y,\mathbf{X})$. Therefore, the MCMC algorithm proposed in Yang and Song (2010a) encounters no problem regardless of singularity of $\mathbf{X}'_\gamma\mathbf{X}_\gamma$.

Funding: This research was fully supported by grants (GRF 446609 and 403109) from the Research Grant Council of the Hong Kong Special Administration Region.

Conflict of Interest: none declared.

REFERENCES

Baragatti,M. and Pommeret,D. (2011) Comments on ‘Bayesian variable selection for disease classification using gene expression data’. *Bioinformatics*, **27**, 1194.

Panagiotelis,A. and Smith,M. (2008) Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *J. Econometrics*, **143**, 291–316.

Yang,A.J. and Song,X.Y. (2010a) Bayesian variable selection for disease classification using gene expression data. *Bioinformatics*, **26**, 215–222.

Yang,A.J. and Song,X.Y. (2010b) Supplementary material to Bayesian variable selection for disease classification using gene expression data. *Bioinformatics*, **26**.

Zellner,A. (1986) On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Geol,P.K. and Zellner,A. (eds) *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*. Elsevier Science, North Holland, Amsterdam, pp. 233–243.