Exact Sampling Procedures for Hierarchical Bayesian Models *

Kjell Nygren †

*Keywords: Bayesian Statistics, Log-concave likelihood functions, Generalized Linear Models, Hierarchical Normal Regression, Hierarchical Poisson Regression, Likelihood Subgradient Densities

†Kjell Nygren is Senior Manager - Marketing Analytics, IMS Health, 960 Harvest Drive, Building A, Blue Bell, PA 19422 (email: knygren@us.imshealth.com)

Abstract

Preliminary and Incomplete - Do Not Quote

1 Introduction

Bayesian statistics is built around the concept of posterior probability. First proposed by Bayes (1763) and Laplace (1785, 1810) in the 18th century, calculations of posterior probabilities requires integration of complex high-dimensional integrals. Except for the simplest of models, there are no reliable analytical or numerical integration techniques available for such models.

Monte Carlo simulation (see e.g., Metropolis and Ulam (1949), Eckhardt (1987) and Fishman (1999)) offers a possible alternative integration technique. The simplest forms of Monte Carlo simulation involves generating independent and identically distributed draws from the densities of interest. Most of these exact sampling procedures make use of either the inverse transform method or von Neumann's (1951) accept-reject procedure. In Bayesian statistics, exact Monte Carlo sampling procedures have largely been limited to models with natural conjugate priors (Raiffa and Schlaifer, 1961). Models with natural conjugate priors have posterior densities of the same functional form as the prior density. The adaptive rejection sampling Algorithm (Gilks and Wild 1992, Gilks 1992) and Adaptive Rejection Metropolis Algorithms (Gilks, Best and Tan 1995) make efficient use of von Neumann's accept-reject procedure in the univariate case and has been implemented as part of Gibbs sampling procedures (Geman and Geman 1984, Gelfand and Smith 1990). These algorithms gradually construct an improved approximation to the full conditional posterior density by using tangents to the full conditional posterior density.

The innovative Gibbs sampling procedures considered by Gelfand and Smith (1990) in the context of a hierarchical random effects model have proven useful for large classes of models and is at the heart of the Bugs (Bayesian Inference Using Gibbs Sampling) Software. Two recent papers by Nygren (2003a, 2003b) derives efficient exact sampling procedures for a large class of generalized linear models with non-conjugate priors. In this paper, we extend most of the results in Nygren (2003b) to the corresponding hierarchical models. Our results are naturally divided into those dealing with models with normal data and models with log-concave likelihood functions.

The first model considered is a hierarchical random effects model with normal data. It includes the balanced hiearchical random effects model considered by Gelfand and Smith (1990) as a special case. We derive an exact sampling procedure for this model. The result for this basic model is then extended to a hierarchical random effects model with multivariate normal data and to a hierarchical Bayesian regression model. While the procedures for the normal and multivariate normal models work in general, the procedure for the hierarchical Bayesian regression model works only if certain full rank conditions are satisfied. The expected number of draws required before acceptance depends on the number of individuals modelled and the dimensionality of the hierarchical model. We compare our acceptance rates for the basic random effects model to the bounds obtained by Jones and Hobert (2004) for the rate of convergence of the Gibbs sampler for the same model.

We next consider hierarchical random effects models with log-concave likelihood functions and show that likelihood subgradient densities (Nygren 2003a) can be used in order to sample from the posterior density resulting from such models. Likelihood subgradient densities make use of subgradients for the negative of the likelihood function at some point $\overline{\beta}$. Our result requires that the subvector of the subgradient which corresponds to the random effects parameters is a vector of zeros. This identifiability condition essentially requires that a conditional maximum is obtained for the likelihood function. In other words, conditional on the other parameters, the random effects parameters must maximize the likelihood function. The derived procedure should prove useful for hierarchical Poisson, hierarchical Logit, and hierarchical survival models. The efficiency of the derived procedure can likely be improved further through the use of Mixture Generalized Likelihood subgradient densities as in Nygren (2003b).

The rest of this paper is organized as follows. In section 2, we derive the exact sampling procedures for hierarchical models with Normal and Multivariate Normal data. Section 3 derives the sampling procedure for the hierarchical random effects models with log-concave likelihood functions. In section 4, we provide explicit examples of how the procedure would be implemented for the hierarchical Poisson and hierarchical Logit models. Section 5, finally, contains some concluding discussion.

2 Generalized Linear Models with Normal and Multivariate Normal Data

2.1 A Hierarchical Normal Population Model

We first consider a Hierarchical Bayesian model of a normal population. Specifically, we consider the following:

$$\sigma^2_{Pop} \sim Inverse - Gamma(m_0/2, m_0 Var_{0,Pop}/2)$$

$$\mu \sim Normal(\nu, \sigma_0^2)$$

$$\beta_i \sim Normal(\mu, \sigma^2_{Pop}), i = 1, \ldots, m_1.$$

$$\sigma^2_{Data} \sim Inverse - Gamma(n_0/2, n_0 Var_{0,Data}/2)$$

and for all $i = 1, \ldots, m_1$,

$$y_{i,j} \sim Normal(\beta_i, \sigma^2_{Data}), j = 1, \ldots, n_i.$$

Here $\mathbf{y}$ represents an observed data vector. We note that this is a non-conjugate model. A balanced version of this model was considered by Gelfand and Smith (1990) in their seminal paper on the Gibbs Sampler. Geometric Ergodicity of the Gibbs sampler under certain assumptions has been established by Hobert and Geyer (1998). Jones and Hobert (2003) provide sufficient burn in iterations for the Gibbs Sampler under similar assumptions.

In order to develop an exact sampling procedure for this model, we introduce the following fictitious model (Model A):

$$\sigma^2_{Data} \sim Inverse - Gamma((n_0 + \sum_{i=1}^{m_1}(n_i - 1))/2, (n_0 Var_{0,Pop} + \sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j} - \overline{y}_i)^2)/2)$$

$$\beta_i \sim Normal(\overline{y}_i, \sigma^2_{Data}/n_i), i = 1, \ldots, m_1$$

$$\sigma^2_{Pop} \sim Inverse - Gamma((m_0 + (m_1 - 1))/2, (m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})^2)/2)$$

$$\mu \sim Normal(\overline{\beta}, \sigma^2_{Pop}/m_1)$$

and a log-likelihood function given by

$$LL(\beta, \mu) = -\frac{(m_0 + m_1 - 1)}{2} \ln(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})^2) - \frac{1}{2}(\nu - \mu)^2/\sigma_0^2$$

where $\overline{\beta} := \sum_{i=1}^{m_1} \beta_i/m_1$, $\overline{y}_i := \sum_{j=1}^{n_i} y_{i,j}/n_i$ for every $i = 1, \ldots, m_1$, and ν is now interpreted as the data. We now establish the following important relationship between the Bayesian Normal Population Model and Model A.

Claim 1. *Fictitious model A has the same posterior density as the Hierarchical Bayesian Normal Population Model.*

Proof of Claim 1: It suffices to show that the log-posterior density for the two models have the same terms involving σ^2_{Data}, β,σ^2_{Pop}, and μ. Starting with the log-posterior from the Hierarchical Bayesian Normal population model, we have (where K is a constant that does not depend on σ^2_{Data}, β,σ^2_{Pop}, or μ)

$$
\begin{aligned}
&[-(\tfrac{m_0}{2}-1)\ln(\sigma^2_{Pop})-\tfrac{1}{2}\tfrac{m_0 Var_{0,Pop}}{\sigma^2_{Pop}}] \\
&+[-\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}(\tfrac{(\beta_i-\mu)^2}{\sigma^2_{Pop}}+\ln(\sigma^2_{Pop}))] \\
&+[-(\tfrac{n_0}{2}-1)\ln(\sigma^2_{Data})-\tfrac{1}{2}\tfrac{n_0 Var_{0,Data}}{\sigma^2_{Data}}] \\
&+[-\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(\tfrac{(y_{i,j}-\beta_i)^2}{\sigma^2_{Data}}+\ln(\sigma^2_{Data}))] \\
&+[-\tfrac{1}{2}\tfrac{(\mu-\nu)^2}{\sigma_0^2}]+K \\
={}& [-(\tfrac{m_0+m_1}{2}-1)]\ln(\sigma^2_{Pop})+[-\tfrac{1}{2}\tfrac{m_0 Var_{0,Pop}}{\sigma^2_{Pop}}] \\
&+[-\tfrac{1}{2}\tfrac{\sum_{i=1}^{m_1}((\beta_i-\overline{\beta})-(\mu-\overline{\beta}))^2}{\sigma^2_{Pop}}] \\
&+[-(\tfrac{n_0+\sum_{i=1}^{m_1}n_i}{2}-1)]\ln(\sigma^2_{Data})+[-\tfrac{1}{2}\tfrac{n_0 Var_{0,Data}}{\sigma^2_{Data}}] \\
&+[-\tfrac{1}{2}\tfrac{\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}((y_{i,j}-\overline{y}_i)-(\beta_i-\overline{y}_i))^2}{\sigma^2_{Data}}] \\
&+[-\tfrac{1}{2}\tfrac{(\mu-\nu)^2}{\sigma_0^2}]+K \\
={}& [-(\tfrac{m_0+m_1}{2}-1)]\ln(\sigma^2_{Pop}) \\
&+[-\tfrac{1}{2}\tfrac{m_0 Var_{0,Pop}+\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})^2}{\sigma^2_{Pop}}] \\
&+[-\tfrac{1}{2}\tfrac{\sum_{i=1}^{m_1}((\mu-\overline{\beta})^2-2(\beta_i-\overline{\beta})(\mu-\overline{\beta}))}{\sigma^2_{Pop}}] \\
&+[-(\tfrac{n_0+\sum_{i=1}^{m_1}n_i}{2}-1)]\ln(\sigma^2_{Data}) \\
&+[-\tfrac{1}{2}\tfrac{n_0 Var_{0,Data}+\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j}-\overline{y}_i)^2}{\sigma^2_{Data}}] \\
&+[-\tfrac{1}{2}\tfrac{\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}((\beta_i-\overline{y}_i)^2-2(y_{i,j}-\overline{y}_i)(\beta_i-\overline{y}_i))}{\sigma^2_{Data}}] \\
&+[-\tfrac{1}{2}\tfrac{(\mu-\nu)^2}{\sigma_0^2}]+K \\
={}& [\tfrac{(m_0+m_1-1)}{2}\ln(m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})^2)] \\
&-[(\tfrac{m_0+m_1-1}{2}-1)]\ln(\sigma^2_{Pop}) \\
&-[\tfrac{1}{2}\tfrac{m_0 Var_{0,Pop}+\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})^2}{\sigma^2_{Pop}}] \\
&-[\tfrac{1}{2}\ln(\tfrac{\sigma^2_{Pop}}{m_1})+\tfrac{1}{2}(\mu-\overline{\beta})^2\tfrac{m_1}{\sigma^2_{Pop}}] \\
&+[-((\tfrac{n_0+\sum_{i=1}^{m_1}(n_i-1)}{2}-1)]\ln(\sigma^2_{Data}) \\
&-[\tfrac{1}{2}\tfrac{(n_0 Var_{0,Data}+\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j}-\overline{y}_i)^2)]}{\sigma^2_{Data}}] \\
&-\textstyle\sum_{i=1}^{m_1}[\tfrac{1}{2}\ln(\tfrac{\sigma^2_{Data}}{n_i})+\tfrac{1}{2}(\beta_i-\overline{y}_i)^2\tfrac{n_i}{\sigma^2_{Data}}] \\
&-[\tfrac{(m_0+m_1-1)}{2}\ln(m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})^2)] \\
&-[\tfrac{1}{2}\tfrac{(\mu-\nu)^2}{\sigma_0^2}] \\
&+K-(1/2)\ln(m_1)-\textstyle\sum_{i=1}^{m_1}(1/2)\ln(n_i)
\end{aligned}
$$

which is recognized as the log-posterior for the fictitious model. Q.E.D.

We note that the likelihood function for fictitious model A obtains its maximum when $\mu=\nu$ and $\beta_i=\overline{\beta}$ for every $i=1,\dots,m_1$. Hence it follows from Fact 1 in Nygren (2003) that the prior of the fictitious model can be used with an accept-reject procedure to generate a sample from the Posterior density. We thus have the following exact sampling procedure for the Hierarchical Bayesian Normal Population Model.

Repeat Until Acceptance:

(i) Generate $\sigma^2_{Data} \sim Inverse{-}Gamma((n_0+\sum_{i=1}^{m_1}(n_i-1))/2, (n_0 Var_{0,Data}+\sum_{i=1}^{m1}\sum_{j=1}^{n_i}(y_{i,j}-\overline{y}_i)^2)/2)$.
(ii) For $i = 1, \ldots, m_1$, generate $\beta_i \sim Normal(\overline{y}_i, \sigma^2_{Data}/n_i)$.
(iii) Generate $\sigma^2_{Pop} \sim Inverse-Gamma((m_0+(m_1-1))/2, (m_0 Var_{0,Pop}+\sum_{i=1}^{m1}(\beta_i-\overline{\beta})^2)/2)$.
(iv) Generate $\mu \sim Normal(\overline{\beta}, \sigma^2_{Pop}/m_1)$.
(v) Generate $U \sim Uniform(0, 1)$.
(vi) Accept if $ln(U) \leq -(1/2)(\mu-\nu)^2(1/\sigma_0^2) - \frac{(m_0+m_1-1)}{2}\ln(\frac{m_0 Var_{0,Pop}+\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})^2}{m_0 Var_{0,Pop}})$.

We note that the acceptance rate depends on two terms. The first term is of the same form as the term used in the accept-reject procedure for the non-hierarchical model in Nygren (2003(b)). It serves to shrink the mean parameter μ towards its prior mean ν. The impact of the shrinkage depends on the strength of the prior. For weak priors, the impact of the shrinkage is small. Conversely, it is large for a strong prior. The second term is new and serves to shrink individual estimates towards the sample mean. We note that the magnitude of this term increases with the total sample size $m_0 + m_1$ and the ratio

$$(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})^2)/(m_0 Var_{0,Pop}).$$

This ratio increases with the number of modelled individuals. Hence the acceptance rate will tend to decline as the number of modelled individuals increases. Exponentiating the acceptance conditions yields the following

$$U \leq e^{-(1/2)(\mu-\nu)^2(1/\sigma_0^2)}(\frac{m_0 Var_{0,Pop}}{m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})^2})^{\frac{(m_0+m_1-1)}{2}}$$

which reveals the dependence on the prior for the population variance parameter.

2.2 A Hierarchical Multivariate Normal Population Model

An important generalization of the Bayesian Normal Population model is to a multivariate normal population. We now consider the following Bayesian Multivariate-normal population model:

$$\mathbf{\Sigma_{Pop}} \sim Inverse - Wishart(m_0\mathbf{Var_{0,Pop}}, m_0)$$

$$\mu \sim Multivariate - Normal(\nu, \mathbf{\Sigma_0})$$

$$\beta_i \sim Multivariate - Normal(\mu, \Sigma_{Pop}), i = 1, \ldots, m_1.$$

$$\Sigma_{Data} \sim Inverse - Wishart(n_0 Var_{0,Data}, n_0)$$

and for all $i = 1, \ldots, m_1$,

$$\mathbf{y}_{i,j} \sim Multivariate-Normal(\beta_i, \Sigma_{Data}), j = 1, \ldots, n_i.$$

where again $\mathbf{y}$ represents an observed data vector.

As in the univariate case, we now introduce a fictitious model (Model B):

$$\Sigma_{Data} \sim Inverse-Wishart(n_0 Var_0 + \sum_{i=1}^{m_1} \sum_{j=1}^{n_i} (y_{i,j} - \overline{y}_i)(y_{i,j} - \overline{y}_i)^T, n_0 + \sum_{i=1}^{m_1} (n_i - 1))$$

$$\beta_i \sim Multivariate-Normal(\overline{y}_i, (1/n_i)\Sigma_{Data}), i = 1, \ldots, m_1$$

$$\Sigma_{Pop} \sim Inverse-Wishart(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1} (\beta_i - \overline{\beta})(\beta_i - \overline{\beta})^T, m_0 + (m_1 - 1))$$

$$\mu \sim Multivariate-Normal(\overline{\beta}, (1/m_1)\Sigma_{Pop})$$

and a log-likelihood function given by

$$LL(\beta, \mu) = -\frac{(m_0 + m_1 - 1)}{2} \ln(|m_0 Var_{0,Pop} + \sum_{i=1}^{m_1} (\beta_i - \overline{\beta})(\beta_i - \overline{\beta})^T|) - \frac{1}{2}(\nu - \mu)^T \Sigma_0^{-1} (\nu - \mu)$$

where $\overline{\beta} := (1/m_1) \sum_{i=1}^{m_1} \beta_i$, $\overline{y}_i := (1/n_i) \sum_{j=1}^{n_i} y_{i,j}$ for every $i = 1, \ldots, m_1$, and ν is now interpreted as the data.

We now show the following equivalence result.

Claim 2. *Fictitious model B has the same posterior density as the Hierarchical Bayesian Multivariate-Normal Population Model.*

Proof of Claim 2: It suffices to show that the log-posterior density for the two models have the same terms involving Σ_{Data}, β,Σ_{Pop}, and μ. Starting with the log-posterior from the p-dimensional Hierarchical Bayesian Multivariate-Normal population model, we have (where K is a constant that does not depend on Σ_{Data}, β,Σ_{Pop}, or μ)

$$
\begin{aligned}
&[-\tfrac{(m_0-p-1)}{2}\ln(|\Sigma_{Pop}|)] \\
&-[\tfrac{1}{2}trace(m_0 Var_{0,Pop}\Sigma_{Pop}^{-1})] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}((\beta_i-\mu)^T\Sigma_{Pop}^{-1}(\beta_i-\mu))] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}\ln(|\Sigma_{Pop}|)] \\
&+[-\tfrac{(n_0-p-1)}{2}\ln(|\Sigma_{Data}|)] \\
&-[\tfrac{1}{2}trace(n_0 Var_{0,Data}\Sigma_{Data}^{-1})] \\
&-[\tfrac{\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j}-\beta_i)^T\Sigma_{Data}^{-1}(y_{i,j}-\beta_i)}{2}] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}\ln(|\Sigma_{Data}|)] \\
&+[-\tfrac{1}{2}(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)]+K \quad = \quad [-\tfrac{(m_0+m_1-p-1)}{2}]\ln(|\Sigma_{Pop}|) \\
&\qquad -[\tfrac{1}{2}trace(m_0 Var_{0,Pop}\Sigma_{Pop}^{-1})] \\
&\qquad -[\tfrac{\sum_{i=1}^{m_1}((\beta_i-\overline{\beta})-(\mu-\overline{\beta}))^T\Sigma_{Pop}^{-1}((\beta_i-\overline{\beta})-(\mu-\overline{\beta}))}{2}] \\
&\qquad -[(\tfrac{n_0+(\sum_{i=1}^{m_1}n_i)-p-1}{2})]\ln(|\Sigma_{Data}|) \\
&\qquad -[\tfrac{1}{2}trace(n_0 Var_{0,Data}\Sigma_{Data}^{-1})] \\
&\qquad -[\tfrac{\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}((y_{i,j}-\overline{y}_i)-(\beta_i-\overline{y}_i))^T\Sigma_{Data}^{-1}((y_{i,j}-\overline{y}_i)-(\beta_i-\overline{y}_i))}{2}] \\
&\qquad +[-\tfrac{1}{2}(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)]+K \\
&= [-\tfrac{(m_0+m_1-p-1)}{2}]\ln(|\Sigma_{Pop}|) \\
&\qquad -[\tfrac{1}{2}trace((m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T)\Sigma_{Pop}^{-1})] \\
&\qquad -[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}(\mu-\overline{\beta})^T\Sigma_{Pop}^{-1}((\mu-\overline{\beta})-2(\beta_i-\overline{\beta}))] \\
&\qquad +[-(\tfrac{n_0+(\sum_{i=1}^{m_1}n_i)-p-1}{2})]\ln(|\Sigma_{Data}|) \\
&\qquad +[-\tfrac{trace((n_0 Var_{0,Data}+\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j}-\overline{y}_i)(y_{i,j}-\overline{y}_i)^T)\Sigma_{Data}^{-1})}{2}] \\
&\qquad +[-\tfrac{\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}((\beta_i-\overline{y}_i)^T\Sigma_{Data}^{-1}((\beta_i-\overline{y}_i)-2(y_{i,j}-\overline{y}_i))}{2}] \\
&\qquad +[-\tfrac{(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)}{2}]+K \\
&= [\tfrac{(m_0+m_1-1)}{2}\ln(|m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T|)] \\
&\qquad -[(\tfrac{m_0+(m_1-1)-p-1}{2})]\ln(|\Sigma_{Pop}|) \\
&\qquad -[\tfrac{1}{2}trace((m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T)\Sigma_{Pop}^{-1})] \\
&\qquad -[\tfrac{1}{2}\ln(|\tfrac{\Sigma_{Pop}}{m_1}|)+\tfrac{1}{2}(\mu-\overline{\beta})^T m_1\Sigma_{Pop}^{-1}(\mu-\overline{\beta})] \\
&\qquad +[-((\tfrac{n_0+(\sum_{i=1}^{m_1}(n_i-1))-p-1}{2})]\ln(|\Sigma_{Data}|) \\
&\qquad +[-\tfrac{trace((n_0 Var_{0,Data}+\sum_{i=1}^{m_1}\sum_{j=1}^{n_i}(y_{i,j}-\overline{y}_i)(y_{i,j}-\overline{y}_i)^T)\Sigma_{Data}^{-1})}{2}] \\
&\qquad -\textstyle\sum_{i=1}^{m_1}[\tfrac{1}{2}\ln(|\tfrac{\Sigma_{Data}}{n_i}|)+\tfrac{1}{2}(\beta_i-\overline{y}_i)^T n_i\Sigma_{Data}^{-1}(\beta_i-\overline{y}_i)] \\
&\qquad -[\tfrac{(m_0+m_1-1)}{2}\ln(|m_0 Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T|)] \\
&\qquad +[-\tfrac{(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)}{2}] \\
&\qquad +K-(p/2)\ln(m_1)-\textstyle\sum_{i=1}^{m_1}(p/2)\ln(n_i)
\end{aligned}
$$

which is recognized as the log-posterior for the fictitious model. Q.E.D.

We note that the likelihood function for fictitious model B obtains its maximum where vector $\mu = \nu$ and $\beta^i = \overline{\beta}$ for $i = 1, \ldots, m_1$. Combining Claim 2 with Fact 1 in Nygren (2003a), we thus have the following exact sampling procedure for the Bayesian Multivariate Normal Population Model.

Repeat Until Acceptance:

(i) Generate $\Sigma_{Data} \sim Inverse - Wishart(n_0 Var_0 + \sum_{i=1}^{m_1} \sum_{j=1}^{n_i} (y_{i,j} - \overline{y}_i)(y_{i,j} - \overline{y}_i)^T, n_0 + \sum_{i=1}^{m_1}(n_i - 1))$.
(ii) For $i = 1, \ldots, m_1$, generate $\beta_i \sim Multivariate - Normal(\overline{y}_i, (1/n_i)\Sigma_{Data})$.
(iii) Generate $\Sigma_{Pop} \sim Inverse-Wishart(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})(\beta_i - \overline{\beta})^T, m_0 + (m_1 - 1))$.
(iv) Generate $\mu \sim Multivariate - Normal(\overline{\beta}, (1/m_1)\Sigma_{Pop})$.
(v) Generate $U \sim Uniform(0, 1)$.
(vi) Accept if $ln(U) \leq -\frac{(m_0+m_1-1)}{2} \ln(\frac{|m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta_i - \overline{\beta})(\beta_i - \overline{\beta})^T|}{|m_0 Var_{0,Pop}|}) - \frac{1}{2}(\mu - \nu)^T \Sigma_0^{-1}(\mu - \nu)$.

The acceptance conditions is essentially just a multivariate generalization of the acceptance condition in the univariate case. To gain insight into the impact of the dimensionality, it is worthwhile to consider the acceptance rate in the case where

$$\sum_{i=1}^{m_1}(\beta_i - \overline{\beta})(\beta_i - \overline{\beta})^T = (m_1 - 1)Var_{0,Pop}.$$

In this case, the exponentiated acceptance condition reduces to

$$U \leq (\frac{m_0}{m_0 + m_1 - 1})^{\frac{(m_0+m_1-1)p}{2}} e^{-\frac{1}{2}(\mu-\nu)^T \Sigma_0^{-1}(\mu-\nu)}$$

which reveals the dependence on the dimensionality p.

2.3 Hierarchical Bayesian Regression with Normal Error terms

The hierarchical Bayesian regression model is an important generalization of the Bayesian regression model. In this section, we derive exact sampling procedures for a Hierarchical Bayesian regression model. More specifically, we consider the following model

$$\mu \sim Multivariate - Normal(\nu, \mathbf{\Sigma_0})$$

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$$

$$\beta 1^i \sim Multivariate - Normal(\mu, \Sigma_{Pop}), i = 1, \ldots, m_1$$

$$\beta 2 \sim Multivariate - Normal(\gamma, \tau_{\mathbf{0}})$$

$$\sigma^2_{Data} \sim Inverse - Gamma(n_0/2, n_0 Var_{0,Data}/2)$$

$$y^i_j \sim Normal(\mathbf{X1^i_j}\beta 1^i + \mathbf{X2^i_j}\beta 2, \sigma^2_{Data}), i = 1, \ldots, m_1, j = 1, \ldots, n_i.$$

Here

$$\mathbf{y} := \begin{bmatrix} \mathbf{y}^1 \\ \mathbf{y}^2 \\ . \\ . \\ \mathbf{y^{m_1}} \end{bmatrix}$$

represents an observed data vector. Let us now introduce some additional notation. Let

$$\tilde{\mathbf{X1}} := \begin{bmatrix} \mathbf{X1}^1 & \mathbf{0} & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{X1}^2 & \mathbf{0} & . & . & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X1}^3 & . & . & \mathbf{0} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & . & . & \mathbf{X1}^{m_1} \end{bmatrix},$$

$$\tilde{\mathbf{X2}} := \begin{bmatrix} \mathbf{X2}^1 \\ \mathbf{X2}^2 \\ . \\ . \\ \mathbf{X2}^{m_1} \end{bmatrix},$$

$$\beta := \begin{bmatrix} \beta 1^1 \\ \beta 1^2 \\ . \\ . \\ \beta 1^{m_1} \\ \beta 2 \end{bmatrix},$$

and $\tilde{\mathbf{X}} := \tilde{\mathbf{X1}} + \tilde{\mathbf{X2}}$.

If $\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$ has full rank p, then we can introduce the following fictitious model (Model C):

$$\sigma^2_{Data} \sim Inverse - Gamma((n_0 + (\sum_{i=1}^{m_1} n_i) - p)/2, (n_0 Var_{0,Data} + (y - \tilde{\mathbf{X}}\tilde{\beta})^T (y - \tilde{\mathbf{X}}\tilde{\beta}))/2)$$

$$\beta \sim Multivariate - Normal(\tilde{\beta}, \sigma^2_{Data}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1})$$

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1} (\beta 1^i - \overline{\beta})(\beta 1^i - \overline{\beta})^T, m_0 + (m_1 - 1))$$

$$\mu \sim Multivariate - Normal(\overline{\beta}, (1/m_1)\Sigma_{Pop})$$

and a log-likelihood function given by

$$\begin{aligned} LL(\beta, \mu) \quad = \quad & -\tfrac{(m_0+m_1-1)}{2} \ln(|m_0 Var_{0,Pop} + \textstyle\sum_{i=1}^{m_1} (\beta 1^i - \overline{\beta})(\beta 1^i - \overline{\beta})^T|) \\ & -\tfrac{1}{2}(\nu - \mu)^T \Sigma_0^{-1} (\nu - \mu) - \tfrac{1}{2}(\beta 2 - \gamma)^T \tau_0^{-1} (\beta 2 - \gamma) \end{aligned}$$

where $\tilde{\beta} := (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\mathbf{y}$, $\overline{\beta} := (1/m_1)\sum_{i=1}^{m_1} \beta 1^i$, and ν is now interpreted as the data. In the case of full rank, we can now show the following equivalence result.

Claim 3. *If $\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$ has full rank, then fictitious model C has the same posterior density as the Normal Regression Model.*

Proof of Claim 3: It suffices to show that the log-posterior density for the two models have the same terms involving Σ_{Data}, β,Σ^2_{Pop}, and μ. Starting with the log-posterior from the p-dimensional Hierarchical Bayesian Multivariate-Normal population model, we have (where K is a constant that does not depend on Σ_{Data}, β,Σ_{Pop}, or μ)

$$\begin{aligned}
&[-\tfrac{(m_0-p-1)}{2}\ln(|\Sigma_{Pop}|)] \\
&-[\tfrac{1}{2}trace(m_0Var_{0,Pop}\Sigma^{-1}_{Pop})] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}((\beta_i-\mu)^T\Sigma^{-1}_{Pop}(\beta_i-\mu))] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}\ln(|\Sigma_{Pop}|)] \\
&-[(\tfrac{(n_0-2)}{2})\ln(\sigma^2_{Data})] \\
&-[(1/2)n_0Var_{0,Data}/\sigma^2_{Data}] \\
&-[(1/2\sigma^2_{Data})(\mathbf{y}-\tilde{\mathbf{X}}\beta)^T(\mathbf{y}-\tilde{\mathbf{X}}\beta)] \\
&-[(1/2)(\textstyle\sum_{i=1}^{m_1}n_i)\ln(\sigma^2_{Data})] \\
&+[-\tfrac{1}{2}(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)] \\
&-\tfrac{1}{2}(\beta2-\gamma)^T\tau_0^{-1}(\beta2-\gamma)+K \\
= \; &[-\tfrac{(m_0+m_1-p-1)}{2}]\ln(|\Sigma_{Pop}|) \\
&-[\tfrac{1}{2}trace(m_0Var_{0,Pop}\Sigma^{-1}_{Pop})] \\
&-[\tfrac{\sum_{i=1}^{m_1}((\beta_i-\overline{\beta})-(\mu-\overline{\beta}))^T\Sigma^{-1}_{Pop}((\beta_i-\overline{\beta})-(\mu-\overline{\beta}))}{2}] \\
&[-(((n_0+\textstyle\sum_{i=1}^{m_1}n_i)/2)-1)\ln(\sigma^2_{Data})] \\
&-[(1/2)n_0Var_{0,Data}/\sigma^2_{Data}] \\
&-[(1/2\sigma^2_{Data})((\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta})-(\tilde{\mathbf{X}}\beta-\tilde{\mathbf{X}}\tilde{\beta}))^T((\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta})-(\tilde{\mathbf{X}}\beta-\tilde{\mathbf{X}}\tilde{\beta}))] \\
&+[-\tfrac{1}{2}(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)] \\
&-\tfrac{1}{2}(\beta2-\gamma)^T\tau_0^{-1}(\beta2-\gamma)+K \\
= \; &[-\tfrac{(m_0+m_1-p-1)}{2}]\ln(|\Sigma_{Pop}|) \\
&-[\tfrac{1}{2}trace((m_0Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T)\Sigma^{-1}_{Pop})] \\
&-[\tfrac{1}{2}\textstyle\sum_{i=1}^{m_1}(\mu-\overline{\beta})^T\Sigma^{-1}_{Pop}((\mu-\overline{\beta})-2(\beta_i-\overline{\beta}))] \\
&[-(((n_0+\textstyle\sum_{i=1}^{m_1}n_i)/2)-1)\ln(\sigma^2_{Data})] \\
&-[(1/2)n_0Var_{0,Data}/\sigma^2_{Data}] \\
&-[(1/2\sigma^2_{Data})(\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta})^T(\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta})] \\
&-[(1/2\sigma^2_{Data})(\beta-\tilde{\beta})^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}(\beta-\tilde{\beta})] \\
&+[-\tfrac{(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)}{2}] \\
&-\tfrac{1}{2}(\beta2-\gamma)^T\tau_0^{-1}(\beta2-\gamma)+K \\
= \; &[\tfrac{(m_0+m_1-1)}{2}\ln(|m_0Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T|)] \\
&-[(\tfrac{m_0+(m_1-1)-p-1}{2})]\ln(|\Sigma_{Pop}|) \\
&-[\tfrac{1}{2}trace((m_0Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T)\Sigma^{-1}_{Pop})] \\
&-[\tfrac{1}{2}\ln(|\tfrac{\Sigma_{Pop}}{m_1}|)+\tfrac{1}{2}(\mu-\overline{\beta})^Tm_1\Sigma^{-1}_{Pop}(\mu-\overline{\beta})] \\
&[-(((n_0+(\textstyle\sum_{i=1}^{m_1}n_i)-p)/2)-1)\ln(\sigma^2_{Data})] \\
&-[(1/2)(n_0Var_{0,Data}+(\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta})^T(\mathbf{y}-\tilde{\mathbf{X}}\tilde{\beta}))/\sigma^2_{Data}] \\
&-[(1/2)\ln(|\sigma^2_{Data}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}|+(1/2\sigma^2_{Data})(\beta-\tilde{\beta})^T\tilde{\mathbf{X}}^T\tilde{\mathbf{X}}(\beta-\tilde{\beta})] \\
&-[\tfrac{(m_0+m_1-1)}{2}\ln(|m_0Var_{0,Pop}+\textstyle\sum_{i=1}^{m_1}(\beta_i-\overline{\beta})(\beta_i-\overline{\beta})^T|)] \\
&+[-\tfrac{(\mu-\nu)^T\Sigma_0^{-1}(\mu-\nu)}{2}]-\tfrac{1}{2}(\beta2-\gamma)^T\tau_0^{-1}(\beta2-\gamma) \\
&+K-(p/2)\ln(m_1)+[(1/2)\ln(|(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}|)]
\end{aligned}$$

which is recognized as the log-posterior for the fictitious model. Q.E.D.

We note that the likelihood function for fictitious model C obtains its maximum at vector μ. Combining Claim 3 with Fact 1 in Nygren (2003), we thus have the following exact sampling procedure for the p-dimensional Hierarchical Bayesian Regression model with normal error terms.

Repeat Until Acceptance:

(i) Generate

$$\sigma^2_{Data} \sim Inverse - Gamma((n_0 + (\sum_{i=1}^{m_1} n_i) - p)/2, (n_0 Var_{0,Data} + (y - \tilde{\mathbf{X}}\tilde{\beta})^T (y - \tilde{\mathbf{X}}\tilde{\beta}))/2).$$

(ii) Generate $\beta \sim Multivariate - Normal(\tilde{\beta}, \sigma^2_{Data}(\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1})$.
(iii) Generate

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta 1^i - \overline{\beta})(\beta 1^i - \overline{\beta})^T, m_0 + (m_1 - 1)).$$

(iv) Generate $\mu \sim Multivariate - Normal(\overline{\beta}, (1/m_1)\Sigma_{Pop})$.
(v) Generate $U \sim Uniform(0, 1)$.
(vi) Accept if

$$\begin{aligned} ln(U) \quad \leq \quad & -\tfrac{(m_0+m_1-1)}{2} \ln\left(\tfrac{|m_0 Var_{0,Pop} + \sum_{i=1}^{m_1}(\beta 1^i - \overline{\beta})(\beta 1^i - \overline{\beta})^T|}{|m_0 Var_{0,Pop}|}\right) \\ & -\tfrac{1}{2}(\nu - \mu)^T \Sigma_0^{-1}(\nu - \mu) - \tfrac{1}{2}(\beta 2 - \gamma)^T \tau_0^{-1}(\beta 2 - \gamma). \end{aligned}$$

We note that this procedure again should have reasonable acceptance rates as long as the prior is relatively weak and centered not too far from the center of the support of the fictitious prior.

3 Hierarchical Generalized Linear Models

In this section, we consider the following hierarchical model:

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$$

$$\beta 2 \sim Multivariate - Normal(\mu, \mathbf{\Sigma_0})$$

for $i = 1, \ldots, m_1$,

$$\beta 1^i \sim Multivariate - Normal(\mathbf{0}, \Sigma_{Pop}), i = 1, \ldots, m_1$$

and a likelihood function

$$f(y|\beta 1, \beta 2).$$

Let

$$\beta := \begin{bmatrix} \beta 1^1 \\ \beta 1^2 \\ . \\ . \\ \beta 1^{m_1} \\ \beta 2 \end{bmatrix}$$

and denote by

$$c(\overline{\beta}) := \begin{bmatrix} c1^1(\overline{\beta}) \\ c1^2(\overline{\beta}) \\ . \\ . \\ c1^{m_1}(\overline{\beta}) \\ c2(\overline{\beta}) \end{bmatrix}$$

a sub-gradient for the negative of the log-likelihood at the point $\overline{\beta}$.

We now claim the following:

Claim 4. *If $c1(\overline{\beta}) = \mathbf{0}$, then the following distribution is a likelihood subgradient density for the hierarchical Bayesian model at $\overline{\beta}$.*

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$$

$$\beta 2 \sim Multivariate - Normal(\mu - \mathbf{\Sigma_0} c2(\overline{\beta}), \mathbf{\Sigma_0})$$

for $i = 1, \ldots, m_1$,

$$\beta 1^i \sim Multivariate - Normal(\mathbf{0}, \Sigma_{Pop}), i = 1, \ldots, m_1$$

Proof of Claim: Denote by $g_{\Sigma_{Pop}}(.)$ the prior distributions for Σ_{Pop}, by $g_{\beta 2}(.)$ the prior distribution for $\beta 2$, and for every $i = 1, \ldots, m_1$ denote by $g_{\beta 1^i}(.|\Sigma_{Pop})$ the prior distribution for $\beta 1^i$. Define a new distribution $h(.)$ by

$$h(\beta 2, \Sigma_{Pop}, \beta 1) := exp(-c2(\overline{\beta})^T \beta 2) g_{\beta 2}(\beta 2) g_{\Sigma_{Pop}}(\Sigma_{Pop}) (\prod_{i=1}^{m_1} g_{\beta 1^i}(\beta 1^i | \Sigma_{Pop})) / MGF(c(\overline{\beta}))$$

where

$$\begin{aligned} MGF(c(\overline{\beta})) &:= \int_{\beta 2} \int_{\Sigma_{Pop}} \int_{\beta 1} e^{-c2(\overline{\beta})^T \beta 2} g_{\beta 2}(\beta 2) g_{\Sigma_{Pop}}(\Sigma_{Pop}) (\textstyle\prod_{i=1}^{m_1} g_{\beta 1^i}(\beta 1^i | \Sigma_{Pop})) d\beta 1 d\Sigma_{Pop} d\beta 2 \\ &= \int_{\beta 2} e^{-c2(\overline{\beta})^T \beta 2} g_{\beta 2}(\beta 2) \int_{\Sigma_{Pop}} g_{\Sigma_{Pop}}(\Sigma_{Pop}) \int_{\beta 1} (\textstyle\prod_{i=1}^{m_1} g_{\beta 1^i}(\beta 1^i | \Sigma_{Pop})) d\beta 1 d\Sigma_{Pop} d\beta 2 \\ &= \int_{\beta 2} e^{-c2(\overline{\beta})^T \beta 2} g_{\beta 2}(\beta 2) \int_{\Sigma_{Pop}} g_{\Sigma_{Pop}}(\Sigma_{Pop}) d\Sigma_{Pop} d\beta 2 \\ &= \int_{\beta 2} e^{-c2(\overline{\beta})^T \beta 2} g_{\beta 2}(\beta 2) d\beta 2 \\ &= exp(-c2(\overline{\beta})^T \mu + \tfrac{1}{2} c2(\overline{\beta})^T \Sigma_0 c2(\overline{\beta})). \end{aligned}$$

Here the last equality follows from Fact 4 in Nygren (2003a). Given the definition of a likelihood subgradient density function in Nygren (2003a) it suffices to show that the probability density function in our theorem corresponds to density function $h(.)$.

Define a new density $h_{\beta 2}(.)$ by

$$h_{\beta 2}(\beta 2) = exp(-c2(\overline{\beta})\beta 2)g_{\beta 2}(\beta 2)/MGF(c(\overline{\beta}))$$

Given the above formula for $MGF(c(\overline{\beta}))$, it is straightforward to verify that $h_{\beta 2}(\beta 2)$ is a multivariate normal density function with mean vector $\mu - \Sigma_0 c2(\overline{\beta})$ and variance-covariance matrix Σ_0.

It then follows that

$$h(\beta 2, \Sigma_{Pop}, \beta 1) := h_{\beta 2}(\beta 2) g_{\Sigma_{Pop}}(\Sigma_{Pop})(\prod_{i=1}^{m_1} g_{\beta 1^i}(\beta 1^i|\Sigma_{Pop}))$$

which is recognized as the density in our Theorem. Q.E.D.

Combining this result with Theorem 3 in Nygren (2003a) yields the following exact sampling procedure for the hierarchical model.

Repeat Until Acceptance:

(i) Generate $\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$.
(ii) Generate $\beta 2 \sim Multivariate - Normal(\mu - \Sigma_0 c2(\overline{\beta}), \mathbf{\Sigma_0})$.
(iii) for $i = 1, \ldots, m_1$, generate
$\beta 1^i \sim Multivariate - Normal(\mathbf{0}, \Sigma_{Pop}), i = 1, \ldots, m_1$.
(iv) Generate $U \sim Uniform(0, 1)$.
(v) Accept if $U \leq exp(-c2(\overline{\beta})^T \overline{\beta 2}) f(\mathbf{y}|\beta)/(exp(-c2(\overline{\beta})^T \beta 2) f(\mathbf{y}|\overline{\beta}))$

The expected number of draws required before acceptance when using this procedure is given explicitly by

$$a := \frac{f(\mathbf{y}|\overline{\beta})}{f(\mathbf{y})} \frac{MGF(c(\overline{\beta}))}{exp(-c2(\overline{\beta})^T \overline{\beta 2})}$$

The expected number of draws required can be reduced through the use of mixture Generalized Likelihood-Subgradient densities as in Nygren (2003b).

A special case of the above model occurs when the likelihood function can be factored as follows:

$$f(y|\beta 1, \beta 2) = \prod_{i=1}^{m_1} f_i(y_i|X1_i \beta 1_i + X2_i \beta 2).$$

Another common formulation of the latter type of hierarchical model is the following

$$\beta 2_1 \sim Multivariate - Normal(\mu_1, \Sigma_{0,1})$$

$$\beta 2_2 \sim Multivariate - Normal(\mu_2, \Sigma_{0,2})$$

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$$

for $i = 1, \ldots, m_1$,

$$\tilde{\beta 1}^i \sim Multivariate - Normal(\beta 2_1, \Sigma_{Pop})$$

and a likelihood function

$$f(y|\tilde{\beta 1}, \beta 2) = \prod_{i=1}^{m_1} f_i(y_i|\tilde{X1}_i \tilde{\beta 1}_i + \tilde{X2}_i \beta 2_2).$$

We now show that by using a one-to-one and onto transformation of the parameters in this formulation, we can still use the above described procedure in order to generate an random draw from the posterior density associated with this model. Define the following:

$$\mu := \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma_0 := \begin{bmatrix} \Sigma_{0,1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{0,2} \end{bmatrix}$$

$$\beta 2 := \begin{bmatrix} \beta 2_1 \\ \beta 2_2 \end{bmatrix}$$

$$\beta 1_i := \tilde{\beta 1}_i - \beta 2_1, i = 1, \ldots, m_1$$

$$X1_i := \tilde{X1}_i, i = 1, \ldots, m_1$$

$$X2_i := \begin{bmatrix} \tilde{X1}_i & \tilde{X2}_i \end{bmatrix}, i = 1, \ldots, m_1.$$

We note that it follows that

$$\begin{aligned} \tilde{X1}_i \tilde{\beta 1}_i + \tilde{X2}_i \beta 2_2 &= \tilde{X1}_i(\beta 1_i + \beta 2_1) + \tilde{X2}_i \beta 2_2 \\ &= \tilde{X1}_i \beta 1_i + \tilde{X1}_i \beta 2_1 + \tilde{X2}_i \beta 2_2 \\ &= X1_i \beta 1_i + X2_i \beta 2. \end{aligned}$$

Hence the above model is equivalent to

$$\beta 2 \sim Multivariate - Normal(\mu, \mathbf{\Sigma_0})$$

$$\Sigma_{Pop} \sim Inverse - Wishart(m_0 Var_{0,Pop}, m_0)$$

for $i = 1, \ldots, m_1$,

$$\beta 1^i \sim Multivariate - Normal(\mathbf{0}, \Sigma_{Pop}), i = 1, \ldots, m_1$$

and a likelihood function given by

$$f(y|\beta 1, \beta 2) = \prod_{i=1}^{m_1} f_i(y_i | X1_i \beta 1_i + X2_i \beta 2).$$

This means that we can use the following procedure in order to generate a sample for $\beta 2, \Sigma_{Pop}$, and $\tilde{\beta 1}$:

(i) Generate a sample for $\beta 2, \Sigma_{Pop}$, and $\beta 1$ as above.
(ii) For each $i = 1, \ldots, m_1$ set $\tilde{\beta 1}_i = \beta 1_i + \beta 2_1$.

4 Examples

5 Discussion

6 References

Anderson, T.W. (1958). *An introduction to Multivariate Statistical Analysis*, Wiley, New York.

Bayes, T. (1763), "An essay towards solving a problem in the doctrine of chances," *Philosophical Transactions of the Royal Society*, 330-418. Reprinted, with biographical note by G.A. Barnard, *Biometrica*, 45, 293-315 (1958).

Breslow, N. E., and Clayton, D. G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9-25.

Dellaportas, P., and Smith, A.F.M. (1993), "Bayesian Inference for generalized linear and proportional hazard models via gibbs sampling," *Applied Statistics*, 42, Issue 3, 443-459.

Eckhardt, R. (1987), "Stan Ulam, John von Neumann, and the Monte Carlo method," *Los Alamos Science*, Special Issue 15, 131-137.

Fishman, G.S. (1999), *Monte Carlo: Concepts, Algorithms, and Applications*, Springer-Verlag, New York, NY.

Gamerman, D. (1997), "Sampling from the posterior distribution in generalized linear mixed models," *Statistics and Computing*, 7, 57-68.

Gelfand, A.E., and Smith, A.F.M. (1990), "Sample-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2000), *Bayesian Data Analysis*, CRC Press LLC, Boca Raton, Florida.

Geman, S., and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6, 721-740.

Gilks, W.R. (1992), "Derivative-free adaptive rejection sampling for Gibbs sampling," *Bayesian Statistics*, 4 (eds. Bernardo, J., Berger, J., Dawid, A.P., and Smith, A.F.M.), Oxford University Press.

Gilks, W.R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics*, 41, pp. 337-348.

Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995), "Adaptive rejection Metropolis sampling," *Applied Statistics,* 44, pp. 455-472.

Greene, W.H. (1997), *Econometric Analysis*, Prentice Hall, Upper Saddle River, New Jersey.

Hastings, W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrica*, 57, 92-109.

Hobert, J.P., and Geyer, C.J. (1998), "Geometric Ergodicity of Gibbs and Block Gibbs Samplers for a Hierarchical Random Effects Model," *Journal of Multivariate Analysis*, 67, 414-430.

Jones, G.L., and Hobert, J.P. (2003), "Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random Effects Model," Accepted for Publication in *The Annals of Statistics.*

Laplace, P-S. (1785), "Memoire sur les formules qui sont fonctions de tres grands nombres," In *Memoires de l'Academie Royale des Sciences.*

—- (1810), "Memoire sur les formules qui sont fonctions de tres grands nombres et sur leurs applications aux probabilities," In *Memoires de l'Academie des Sciences de Paris.*

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, 21, 1087-1092.

Metropolis, N., and Ulam, S. (1949), "The Monte Carlo Method," *Journal of the American Statistical Association*, 44(247), 335-341.

The Numerical Algorithms Group Ltd(2002), *The Nag C Library Manual, Mark 7*, Oxford, UK.

Nygren, K. (2003), "Exact Sampling from Posterior Densities Using Likelihood Subgradient Densities with Accept-Reject Procedures", submitted to *Journal or the American Statistical Association.*

Raiffa, H., and Schlaifer, R. (1961), *Applied Statistical Decision Theory*, Division of Research, Graduate School of Business Administration, Harvard University, Boston.

von Neumann, J. (1951), "Various techniques used in connection with random digits," *Monte Carlo Method*, Applied Mathematics Series 12, National Bureau of Standards, Washington, D.C.