Statistica Sinica **13**(2003), 461-476

# CONJUGATE PRIORS FOR GENERALIZED LINEAR MODELS

Ming-Hui Chen and Joseph G. Ibrahim

*University of Connecticut and University of North Carolina*

*Abstract:* We propose a novel class of conjugate priors for the family of generalized linear models. Properties of the priors are investigated in detail and elicitation issues are examined. We establish theorems characterizing the propriety and existence of moments of the priors under various settings, examine asymptotic properties of the priors, and investigate the relationship to normal priors. Our approach is based on the notion of specifying a prior prediction $y_0$ for the response vector of the current study, and a scalar precision parameter $a_0$ which quantifies one's prior belief in $y_0$. Then $(y_0, a_0)$, along with the covariate matrix $X$ of the current study, are used to specify the conjugate prior for the regression coefficients $\beta$ in a generalized linear model. We examine properties of the prior for $a_0$ fixed and for $a_0$ random, and study elicitation strategies for $(y_0, a_0)$ in detail. We also study generalized linear models with an unknown dispersion parameter. An example is given to demonstrate the properties of the prior and the resulting posterior.

*Key words and phrases:* Conjugate prior, generalized linear models, Gibbs sampling, historical data, logistic regression, poisson regression, predictive elicitation.

## 1. Introduction

Conjugate priors play an important role in Bayesian inference, since it is desirable to have posterior distributions with the same functional form and similar properties as the prior. Conjugate priors often have desirable features important in interpretation, data analysis, and computations. They are straightforward to construct for many models in the *i.i.d.* setting. In fact well known classes of conjugate priors are available in exponential family models, likelihood-prior combinations include the normal-normal, binomial-beta, Poisson-gamma, and gamma-gamma models. Diaconis and Ylvisaker (1979), and Morris (1982, 1983) examine general classes of conjugate priors for exponential family models. However, in regression settings, the development of conjugate priors for regression coefficients is much more complicated and the construction is not at all clear. For the class of generalized linear models (GLM's), we are not aware of any papers that develop conjugate priors for the regression coefficient vector $\beta$.

We propose a class of conjugate priors for the family of generalized linear models (GLM's). Our construction is predictive in nature and focuses on observable quantities, it is based on specifying a prior prediction $y_0$ for the response vector, and a scalar precision parameter $a_0$ which quantifies one's prior belief in $y_0$. Then $(y_0, a_0)$, along with the covariate matrix $X$ of the current study, are used to specify a conjugate prior for the regression coefficients $\beta$ in the GLM. The motivation is that the investigator often has prior information on the observables from similar previous studies or from case-specific information on the subjects in the current study. This information is often quantifiable in the form of a vector of prior predictions for the response vector of the current study. In addition, it is easier to think of observable quantities when eliciting priors, rather than specifying priors for regression parameters directly, since parameters are always unobserved. Our approach is especially appealing for variable selection problems since there are many parameters arising from different models and with different physical meaning, therefore making direct prior elicitation quite difficult. A recent article which addresses informative prior specifications for generalized linear models is Bedrick, Christensen, and Johnson (1996). Our approach focusses on a direct prior elicitation for the regression coefficients, as opposed to their Conditional Means Priors (CMP) and Data Augmentation Priors (DAP) which are based on evaluation of the prior at $p$ locations in the predictor space, where $p$ is the dimension of the regression coefficient vector.

The rest of this article is organized as follows. In Section 2, we discuss the GLM and propose a general class of conjugate priors for the GLM and investigate its properties. In Section 3, we discuss elicitation issues in detail and provide some practical guidelines for eliciting the hyperparameters of the conjugate prior. In Sections 4 and 5, we investigate some extensions of the proposed prior and, in particular, examine the case in which $a_0$ is random, as well as the case of unknown dispersion parameters in GLM's. In Section 6, we present an illustrative example.

## 2. The Prior

Suppose $y_1, \ldots, y_n$ are independent observations, where $y_i$ has a density in the exponential family

$$p(y_i|\theta_i, \tau) = \exp\left\{a_i^{-1}(\tau)(y_i\theta_i - b(\theta_i)) + c(y_i, \tau)\right\}, \quad i = 1, \ldots, n, \tag{2.1}$$

indexed by the canonical parameter $\theta_i$ and the scale parameter $\tau$. The functions $b$ and $c$ determine a particular family in the class, such as the binomial, normal, Poisson, etc. The functions $a_i(\tau)$ are commonly of the form $a_i(\tau) = \tau^{-1}w_i^{-1}$, where the $w_i$'s are known weights. For ease of exposition, we take $w_i = 1$ throughout. Now suppose the $\theta_i$'s satisfy

$$\theta_i = \theta(\eta_i), \quad i = 1, \ldots, n, \tag{2.2}$$

$$\eta = X\beta, \tag{2.3}$$

where $\eta_i$ are the components of $\eta$, $X$ is an $n \times p$ full rank matrix of covariates, $\beta = (\beta_0, \ldots, \beta_{p-1})'$ is a $p \times 1$ vector of regression coefficients, and $\theta$ is a monotone differentiable function. Models given by (2.1)−(2.3) are called generalized linear models (GLM's). The function $\theta$ is sometimes referred to as the $\theta$-link to distinguish it from the conventional link $g(\mu_i)$ which relates $\eta_i$ to the mean $\mu_i$ of $y_i|\theta_i$. We refer to $g(\mu_i)$ as the $\mu$-link. When $\theta_i = \eta_i$, the link is said to be a canonical link.

We specify a conjugate prior for the regression coefficients $\beta$ in a GLM by first adapting the results of Diaconis and Ylvisaker (1979). Toward this goal, let the canonical parameters in the GLM be independently distributed a priori, and let $\theta = (\theta_1, \ldots, \theta_n)'$ and $y = (y_1, \ldots, y_n)'$. Following the construction of Diaconis and Ylvisaker (1979), we get the joint prior

$$\pi(\theta|\tau, y_0, a_0) \propto \prod_{i=1}^{n} \exp\{a_0\tau(y_{0i}\theta_i - b(\theta_i))\} = \exp\{a_0\tau(y_0'\theta - J'b(\theta))\}, \tag{2.4}$$

where $a_0 > 0$ is a scalar prior parameter, $y_0 = (y_{01}, \ldots, y_{0n})'$ is an $n \times 1$ vector of prior parameters, $J$ is an $n \times 1$ vector of ones, and $b(\theta) = (b(\theta_1), \ldots, b(\theta_n))'$ is an $n \times 1$ vector of the $b(\theta_i)$'s. We mention here that (2.4) assumes that the $\theta_i$'s are independent a priori. This construction is consistent with the notion that, given $\theta_i$, the $y_{0i}$'s are independent. That is, the sampling distribution of the $y_{0i}$'s is identical to the response variables of the current experiment. This is a reasonable assumption to make if $y_0$ in fact represents a prior guess for $y$. Moreover, we note that the $\theta_i$'s are independent a priori before the covariates enter into the model. Once covariates are introduced, as in (2.6) below, none of the parameters in the prior are independent a priori. Thus (2.4) is not a restrictive assumption.

Disregarding for the moment any relationship of $\theta$ to the regression coefficients $\beta$, we describe the choice of the parameters of this prior for $\theta$. As shown in Diaconis and Ylvisaker (1979), $y_0 = E(\dot{b}(\theta))$, where $\dot{b}(\theta)$ is the gradient vector of $b(\theta)$ and the expectation is taken with respect to the prior distribution in (2.4). Since for GLM's $E(y|\theta) = \dot{b}(\theta)$, we have

$$E(y) = E_\theta[E(y|\theta)] = E(\dot{b}(\theta)) = y_0. \tag{2.5}$$

Thus (2.5) shows that $y_0$ is the marginal mean of $y$ and could be interpreted as a prior prediction (or guess) for $E(y)$. The parameter $a_0$ can be viewed as a prior sample size. In the present context, it would represent $\frac{n_0}{n}$, where $n_0$ is a sample size judged equivalent to the information in the prior. Using the parameters $y_0$ and $a_0$, we proceed to specify a prior for $\beta$. The prior on $\theta$ in (2.4) induces a prior on $\beta$ since $\beta$ is functionally related to $\theta$ via $X$. However, this induced prior is not tractable, and is not conjugate in general. Here, we propose a conjugate prior for $\beta$ by directly substituting $\theta$ as a function of $\beta$ into (2.4). As shown below, this results in a proper conjugate prior for $\beta$ given $\tau$. We thus write the prior as

$$\pi(\beta|a_0, y_0, \tau) \propto \exp\{a_0\tau[y_0'\theta(\eta) - J'b(\theta(\eta))]\} \equiv \exp\{a_0\tau[y_0'\theta(X\beta) - J'b(\theta(X\beta))]\}. \tag{2.6}$$

We denote the prior in (2.6) by $(\beta|a_0, y_0, \tau) \sim D(y_0, a_0)$, where $(y_0, a_0)$ are the specified hyperparameters. We see that (2.6) depends on the covariate matrix $X$, which is the same covariate matrix that appears in the likelihood function of $\beta$. Since we view the covariates as fixed a priori, our prior is *not* data dependent. In fact, the dependence of our prior on $X$ gives $y_0$ a more appealing interpretation. The dependence of our prior on the covariate matrix $X$ is also a nice feature in the sense that the idea easily extends to other types of models, such as random effects models and nonlinear models. From (2.6), we see that the $ith$ component of $y_0$ is linked to the covariate vector $x_i$ for the $ith$ subject. This link, along with (2.5), implies that $y_{0i}$ is precisely a prior prediction for the marginal mean $E(y_i)$ of $y_i$. Thus, in eliciting $y_0$, the user must focus on a prediction (or guess) for $E(y)$, which narrows the possibilities. Moreover, the specification of all $y_{0i}$ equal has an appealing interpretation: the prior modes of the regression coefficients corresponding to the covariates in the regression model are the same, but the prior modes of the intercept in the regression model vary. This is intuitive since in this case, the prior prediction on $y_{0i}$ does not depend on the $ith$ subject's case specific covariate information. The parameter $a_0$ in (2.6) can be viewed as a precision parameter that quantifies the strength of our prior belief in $y_0$. One of the main roles of $a_0$ is that it controls the heaviness of the tails of the prior for $\beta$. The smaller the $a_0$, the heavier the tails. When $a_0 = 0$, (2.6) reduces to a uniform improper prior for $\beta$; as $a_0$ gets large, (2.6) becomes more informative in $\beta$ and, as $a_0 \to \infty$, the prior reduces to a point mass at its mode. We discuss elicitation of $(a_0, y_0)$ in more detail in Section 3.

We note here that (2.6) is related to, but quite different, from the DAP priors of Bedrick, Christensen, and Johnson (1996). First, in constructing (2.6), we preserve the dimension of $y_0$ to be the same as that of $y$. Thus, $y_0$ precisely represents a prior guess for $E(y)$. In addition, we use the same covariate matrix $X$ as the current experiment to construct (2.6). Finally, we specify a weight parameter $a_0$ that acts as an effective prior sample size for the prior. Hence, (2.6) requires a specification of $(y_0, X, a_0)$. This is quite different from the framework of Bedrick et al. (1996), where they specify $p$ "prior observations" $(\tilde{y}_i, \tilde{x}_i, \tilde{w}_i, i = 1, \ldots, p)$ to construct their prior, where $\tilde{y}_i$ represent potentially observable response variables taken at some covariate vector $\tilde{x}_i$, which may or may not be related to the covariates $X$ of the current experiment. In addition, the DAP priors do not lead to conjugate priors for the class of GLM's in general. Finally, the $\tilde{w}_i$'s are the prior weights for $(\tilde{y}_i, \tilde{x}_i)$. Thus, $(\tilde{y}_i, \tilde{x}_i, \tilde{w}_i, i = 1, \ldots, p)$ have a completely different interpretation than $(y_0, X, a_0)$ and play a fundamentally different role in the prior construction. Thus, the DAP priors and the elicitation strategies for them are quite different than those of (2.6).

As an example of (2.6), we consider the normal linear regression model with canonical link and error precision $\tau = 1$, i.e., $y|X, \beta \sim N_n(X\beta, I)$. For this model $b(\theta_i) = \theta_i^2/2$, so that

$$\pi(\beta|a_0, y_0) \propto \exp\{a_0[y_0' X\beta - J' b(X\beta)]\} \propto \exp\left\{-\frac{a_0}{2}(\beta - \mu_0)'(X'X)(\beta - \mu_0)\right\}, \tag{2.7}$$

where $\mu_0 = (X'X)^{-1}X'y_0$. Thus $(\beta|a_0, y_0) \sim N_p(\mu_0, a_0^{-1}\ (X'X)^{-1})$. In this example, we see the precise roles of $y_0$ and $a_0$. In (2.7), $y_0$ corresponds to the "response vector" in a linear regression of $y_0$ on $X$, and $\mu_0$ is the least squares estimate of $\beta$ from this regression. From (2.7), we see that $a_0$ is a precision parameter that quantifies the degree of prior belief in $\mu_0$, and hence $y_0$.

Although (2.6) does not have a closed form in general for most GLM's, it lends itself to several theoretical and computational properties given below. The first result deals with the existence of the moment generating function (MGF) of (2.6).

**Theorem 2.1.** *Let $a_0 > 0$ and take $y_0 \in \mathcal{Y}$, where $\mathcal{Y}$ is the interior of the convex hull of the support for the density in* (2.1). *Assume that* $\exp\{\tau(y_{0i}\theta_i - b(\theta_i))\}$ *is bounded. Then,* (i) *under a canonical link, i.e., $\theta = \eta$, the moment generating function* (*MGF*) *of $\beta$ exists;* (ii) *under a non-canonical link, a sufficient condition for the MGF of $\beta$ to exist is that the one dimensional integral*

$$\int_{\Theta_i} \left|\frac{d}{dr_i}\theta^{-1}(r_i)\right| \exp(s_0|\theta^{-1}(r_i)|)\ \exp\{a_0\tau(y_{0i}r_i - b(r_i))\}\, dr_i < \infty \tag{2.8}$$

*for some $s_0 > 0$. Here $\Theta_i$ denotes the parameter space of the* (*univariate*) *canonical parameter $r_i$.*

A proof of Theorem 2.1 is given in the Appendix.

The next theorem states the conjugacy of (2.6).

**Theorem 2.2.** *If $(\beta|a_0, y_0, \tau) \sim D(y_0, a_0)$, then $D$ is a conjugate prior for $(\beta|a_0, y_0, \tau)$, with the posterior given by*

$$(\beta|y, y_0, a_0, \tau) \sim D\left(\frac{a_0 y_0 + y}{a_0 + 1}, a_0 + 1\right). \tag{2.9}$$

The proof follows from a straightforward multiplication of the likelihood in (2.1) and the prior in (2.6), then recognition of the resulting posterior.

The prior defined by (2.6) may also be viewed as a posterior density of $(\beta|a_0, y_0, \tau)$ with $y_0$ as the data, based on an initial uniform prior for $\beta|\tau$. It can be shown that as $n \to \infty$, (2.6) converges to a $p$ dimensional multivariate normal distribution. This is formally stated in the following theorem.

**Theorem 2.3.** *Consider the prior in* (2.6). *Then, as* $n \to \infty$,

$$\pi(\beta|\tau, a_0, y_0) \to N_p(\hat{\beta}, a_0^{-1}\tau^{-1}\hat{T}^{-1}), \tag{2.10}$$

$$T = X'\hat{\Delta}^2\hat{V}X, \tag{2.11}$$

$\hat{\beta}$ *is the mode* (*MLE*) *of* $\beta|\tau$ *using* $y_0$ *as the data,* $\hat{\Delta}$ *and* $\hat{V}$ *are* $n \times n$ *diagonal matrices with ith diagonal elements* $\delta_i \equiv \delta_i(x_i'\beta) = d\theta_i/d\eta_i$ *and* $v_i \equiv v_i(x_i'\beta) = d^2b(\theta_i)/d\theta_i^2$ *evaluated at* $\hat{\beta}$*, and* $x_i'$ *is the ith row of* $X$.

The proof of this theorem is omitted here for the sake of brevity.

We mention here that (2.6) is related to, but quite different from the power priors proposed in Ibrahim and Chen (2000). First, the latter are not conjugate in the sense of (2.9). Second, the power priors in Ibrahim and Chen (2000) assume the existence of historical data for the construction of the prior, take $y_0$ to be the response vector corresponding to the raw historical data, and take the covariate matrix to be the covariate matrix corresponding to the historical data.

## 3. Elicitation of $y_0$ and $a_0$

Taking (2.6) as the prior for the regression coefficients, we now consider elicitation schemes for $(y_0, a_0)$. According to (2.1), $y_0$ must be in the interior of the convex hull of the sampling density of $y|\theta$, with $a_0 > 0$. One possible strategy for eliciting $y_0$ is to use expert opinion or case-specific information on each subject. Another strategy is to elicit $y_0$ from forecasts or predictions obtained from a theoretical prediction model. In this case, we could obtain a point prediction of the form

$$y_0 = h(X_0), \tag{3.1}$$

where $X_0$ is a matrix of covariates based on a previous similar study and $h(.)$ is a specified function. Specifically, the investigator may have substantive prior information in the form of training data, historical data, or summary statistics for eliciting $y_0$. For example, in the context of logistic regression, $y_0$ is a vector of probabilities and we can take $y_{0i}$ to be of the form $y_{0i} = \exp(x_{i0}'\tilde{\beta})/(1 + \exp(x_{i0}'\tilde{\beta}))$, $i = 1, \ldots, n$, $x_{i0}'$ is the *ith* row of $X_0$ and $\tilde{\beta}$ is an estimate of $\beta$ from the training data, historical data, or summary statistics. If the above methods are not available, they can alternatively specify "vague" choices for $y_0$. For example, in the context of logistic regression, if we take $y_0 = (0.5, \ldots, 0.5)'$ the prior mode of $\beta$ is 0. Asymptotically, this choice of $y_0$ results in a $N_p(0, a_0^{-1}T^{-1})$ for $\beta$, where

$T$ is defined by (2.11) with $X$ replaced by $X_0$. Thus if $a_0$ is taken to be small, this choice of $y_0$ results in a noninformative prior for $\beta$. Similar choices can be employed for other GLM's.

The methods described above provide a *direct* elicitation of $y_0$. We can also specify $y_0$ *indirectly* through a prior specification for the mode of $\beta$. To fix ideas, let $\mu_0$ be a specified $p \times 1$ vector, the desired prior mode of $\beta$ for (2.6). We emphasize here that $\mu_0$ does *not* depend on $X$. Now we ask the question: What is the corresponding $y_0$ that yields this $\mu_0$ from (2.6)? The answer is given in the following theorem.

**Theorem 3.1.** *Let $\mu_0$ be any prespecified $p \times 1$ vector. Let*

$$y_0 = \dot{b}(\theta) = \dot{b}(\theta(X\mu_0)). \tag{3.2}$$

*Then, the prior given by* (2.6) *yields a prior mode of $\beta$ equal to $\mu_0$.*

The proof of Theorem 3.1 follows directly from the fact that when $y_0$ takes the form (3.2), $\beta = \mu_0$ is a solution of

$$\frac{\partial \ln \pi(\beta|a_0, y_0, \tau)}{\partial \beta} = a_0 \tau \left( y_0 \circ \frac{\partial \theta}{\partial \eta} - \dot{b}(\theta) \circ \frac{\partial \theta}{\partial \eta} \right)' X = 0, \tag{3.3}$$

where $\circ$ denotes the direct product. Theorem 3.1 also implies that, as $n \to \infty$, the choice of $y_0$ given in (3.2) yields the same prior mean as a normal prior for $\beta$.

**Remark 3.1** When $\pi(\beta|a_0, y_0, \tau)$ is log-concave, $y_0 = \dot{b}(\theta(X\mu))$ yields a unique prior mode of $\beta = \mu_0$, i.e., the solution of (3.3) is unique. We note that the log-concavity is true for many members in the GLM family, such as the GLM's with canonical links (Diaconis and Ylvisaker (1979)), and for many GLM's with noncanonical links (see Wedderburn (1976)).

**Remark 3.2** In the context of binary regression, (3.2) reduces to $y_0 = F(X\mu_0) = (F(x_1'\mu_0), F(x_2'\mu_0), \ldots, F(x_n'\mu_0))'$, where $F$ is the cumulative distribution function used for the link in the binary regression. In particular, for binary regression models with a symmetric link, which includes the probit, logit, and $t$-link as special cases, a prior mode of $\mu_0 = 0$ yields $y_0 = (0.5, \ldots, 0.5)'$. However, for the complementary log-log link, when $\mu_0 = 0$, (3.2) simply takes the form $y_0 = (1-\exp(-1), \ldots, 1-\exp(-1))'$. For Poisson regression with a canonical link, a prior mode of $\mu_0 = 0$ yields $y_0 = (1, 1, \ldots, 1)'$. For the exponential regression model with a log-link, a prior mode of $\mu_0 = 0$ yields $y_0 = (1, 1, \ldots, 1)'$.

**Remark 3.3** For binary regression models with symmetric links, the unique $y_0$ that satisfies (3.2) with $\mu_0 = 0$ yields a symmetric prior for (2.6) about its mode, which is 0. However, when $y_0$ satisfies (3.2) with $\mu_0 \neq 0$, the resulting prior for (2.6) is no longer symmetric in general except for special structures of the covariate matrix $X$.

The hyperparameter $y_0$ only affects the location of $\beta$ in (2.6), and plays no role in the dispersion. Thus, the location of $\beta$ is primarily regulated by $y_0$. In addition $y_0$ also plays a large role in the symmetry of the prior distribution (2.6) (see Remark 3.3). On the other hand, $a_0$ primarily controls the dispersion in the prior distribution. From (2.6), we see that the prior mean of $\beta$ will indeed depend on $a_0$, but the the prior mode of $\beta$ never depends on $a_0$. In certain cases, (2.6) can be quite skewed, as demonstrated in Section 6. However, as $n \to \infty$, (2.6) does become more symmetric due to (2.10), and in this case the prior mean converges to the prior mode and the prior is symmetric about its mode. Also, as $a_0 \to \infty$, (2.6) becomes more symmetric about its prior mode. Thus, making $a_0$ large results in a more symmetric prior regardless of the value of $n$. This indicates some overlap in the roles of $(y_0, a_0)$ in (2.6).

The elicitation of $a_0$ is less straightforward than that of $y_0$. If $y_0$ is based on training data, historical data, or summary statistics based on a sample size of $n_0$, then a possible choice for $a_0$ is $a_0 = n_0/n$. In general, if training data, historical data, or summary statistics are not available for specifying $(y_0, a_0)$, we recommend the following guidelines for specifying $(y_0, a_0)$ in practice.

(1) For an initial choice of $y_0$, we use the value $\tilde{y}_0$ that yields a prior mode of $\beta$ equal to 0, found by solving (3.2) using $\mu_0 = 0$. Then we do several sensitivity analyses about $\tilde{y}_0$. We call $\tilde{y}_0$ the *guide value* for $y_0$.

(2) A value of $a_0 = 1$ is a reasonable starting value, since it gives equal weight to the likelihood and the prior. Using $a_0 = 1$ as our *guide value*, we do sensitivity analyses about this guide using other values such as $a_0 = 0, 0.1, 10, 100, 1000$.

## 4. Random $a_0$

Since a single value of $a_0$ may be difficult to specify a priori, we can express our uncertainty about $a_0$ by specifying a gamma prior for it. This leads to the joint prior

$$\pi(\beta, a_0 | y_0, \tau) \propto \exp\{a_0[\tau(y_0'\theta(\eta) - J'b(\theta(\eta))) + J'c(y_0, \tau)]\}\ a_0^{\alpha_0 - 1} \exp(-\lambda_0 a_0), \tag{4.1}$$

where $c(y_0, \tau)$ is a $n \times 1$ vector of the $c(y_{0i}, \tau)$'s, and $(\alpha_0, \lambda_0)$ are specified prior parameters. One attractive feature of (4.1) is that it creates heavier tails for the marginal prior of $\beta$ than the prior (2.6), which assumes $a_0$ is a fixed value. We now give a theorem characterizing the propriety of (4.1).

**Theorem 4.1.** *Take $y_0 \in \mathcal{Y}$, where $\mathcal{Y}$ is the interior of the convex hull of the support for the density in* (2.1). *Assume that* $\exp\{\tau(y_{0i}\theta_i - b(\theta_i)) + c(y_{0i}, \tau)\}$ *is* *bounded,* $\alpha_0 > p + k$, *and* $\lambda_0 > \max\{0, \sup_{\beta \in R^p} [\tau(y_0'\theta(\eta) - J'b(\theta(\eta))) + J'c(y_0, \tau)]\}$. *Then*

$$\int_{\Theta_i}^{\infty} \left| \frac{d}{dr_i} \theta^{-1}(r_i) \right| \exp(s_0 |\theta^{-1}(r_i)|) \ \exp\{\tau(y_{0i} r_i - b(r_i))\} dr_i < \infty \tag{4.2}$$

*for some* $s_0 > 0$, *where* $\Theta_i$ *denotes the parameter space of the* (*univariate*) *canonical parameter* $r_i$, *and*

$$\int_{R^p} \int_0^{\infty} ||\beta||^k \pi(\beta, a_0 | y_0, \tau) da_0 d\beta < \infty, \tag{4.3}$$

*where* $||\beta|| = (\beta'\beta)^{1/2}$.

The proof is given in the Appendix. We note that, in general, the MGF of $\beta$ does not exist when $a_0$ is random. This can be clearly seen from the normal linear regression model with canonical link and $\tau = 1$, since in this case, the marginal prior of $\beta$ is a $t$ distribution.

## 5. Random $\tau$

In this section, we consider GLM's with an unknown dispersion parameter. For the moment let $a_0$ be fixed and let $\pi(\tau)$ denote an initial prior for $\tau$. Then, the joint prior for $(\beta, \tau)$ has the form

$$\pi(\beta, \tau | y_0, a_0) \propto \exp\{a_0[\tau(y_0'\theta(\eta) - J'b(\theta(\eta))) + J'c(y_0, \tau)]\} \ \pi(\tau). \tag{5.1}$$

Similar to Theorem 2.2, it can be shown that $\pi(\beta, \tau | y_0, a_0)$ is a conjugate prior.

Now assume that $\exp\{a_0 [\tau (y_{0i}\theta(\eta_i) - b(\theta(\eta_i))) + c(y_{0i}, \tau)]\}$ is bounded by $M_i(\tau | a_0) = \sup_{\eta_i} \{\exp(a_0[\tau(y_{0i}\theta(\eta_i) - b(\theta(\eta_i))) + c(y_{0i}, \tau)])\}$. Following the notation used in the proof of Theorem 2.1, we partition a row permutation of $X$ into $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, where $X_1$ is a $p \times p$ full rank matrix and $X_2$ is an $(n-p) \times p$ matrix. For ease of notation, we assume that the first $p$ rows of $X$ form the submatrix $X_1$. Then, we are led to the following theorem.

**Theorem 5.1.** *Take* $y_0$ *to be in the convex hull of the support of the density in* (2.1). *Assume that for any initial prior* $\pi(\tau)$, *proper or improper,*

$$\int_0^{\infty} \prod_{j=p+1}^{n} M_j(\tau | a_0) \Big\{ \prod_{i=1}^{p} \int_{\Theta_i} \left| \frac{d}{dr_i} \theta^{-1}(r_i) \right| \exp\left(a_0[\tau(y_{0i} r_i - b(r_i)) + c(y_{0i}, \tau)]\right) dr_i \Big\} d\tau < \infty, \tag{5.2}$$

*where* $\Theta_i$ *is defined in* (2.8). *Then the prior* (5.1) *is proper.*

The proof follows directly from (5.2) and the proof of Theorem 2.1. The details are omitted. Similar sufficient conditions are also considered in Sun, Tsutakawa and He (2001). We also note that for the normal linear model, condition (5.2) can be relaxed.

When both $a_0$ and $\tau$ are random, a joint prior of $(\beta, \tau, a_0)$ becomes cumbersome. For illustrative purposes, we consider only the normal linear regression model and propose the following joint prior:

$$\pi(\beta, \tau, a_0|y_0) \propto \exp[-a_0\tau(y_0-X\beta)'(y_0-X\beta)/2]\tau^{\zeta_0-1}\exp(-\delta_0\tau)a_0^{\alpha_0-1}\exp(-\lambda_0 a_0), \tag{5.3}$$

where $\zeta_0$, $\delta_0$, $\alpha_0$, and $\lambda_0$ are prespecified hyperparameters. The following theorem characterizes the propriety of the joint prior (5.3).

**Theorem 5.2.** *Assume that* $\zeta_0 > p/2$, $\delta_0 > 0$, $\alpha_0 > \zeta_0$, $\lambda_0 > 0$, *and* $\lambda_0 + (n/2)\ln\delta_0 > 0$. *Then, for any* $y_0 \in R^n$, *the joint prior* (5.3) *is proper.*

The proof is given in the Appendix.

## 6. Illustrative Example

Suppose $y_i|\theta_i$ are independent Bernoulli observations with probability of success $p_i = e^{x_i'\beta}/(1+e^{x_i'\beta})$, where $x_i'$ is a $1 \times p$ vector, $i = 1, \ldots, n$. The conjugate prior in (2.6) takes the form

$$\pi(\beta|a_0, y_0) \propto \exp\Big\{\sum_{i=1}^{n} a_0\left(y_{0i}\; x_i'\beta - \log(1+e^{x_i'\beta})\right)\Big\}, \tag{6.1}$$

where $y_{0i}$ is the *ith* component of $y_0$. We consider data from Finney (1947), obtained to study the effect of the rate and volume of air inspired on a transient vaso-constriction in the skin of the digits. The response variable measured is binary with 1 and 0 indicating occurrence or nonoccurrence of vaso-constriction, respectively. The dataset can also be found in Pregibon (1981). There are $n = 39$ observations in the dataset. The two covariates are $x_1 = \log(\text{volume})$ and $x_2 = \log(\text{rate})$ with $\beta_1$ and $\beta_2$ denoting the respective regression coefficients. For these data, we consider a logistic regression model along with the prior in (6.1). An intercept $\beta_0$ is also included in the model, and thus $\beta = (\beta_0, \beta_1, \beta_2)$.

The maximum likelihood estimates and the standard deviations are $-2.875$ and 1.319 for $\beta_0$, 5.179 and 1.862 for $\beta_1$, and 4.562 and 1.835 for $\beta_2$, respectively. For ease of exposition, the notation $y_0 = 0.1$ means that $y_0 = (0.1, \ldots, 0.1)'$, and so forth. Also, SD denotes standard deviation. The prior modes of $\beta$ are $(-2.197, 0, 0)'$ for $y_0 = 0.1$, $(0, 0, 0)'$ for $y_0 = 0.5$, and $(2.197, 0, 0)'$ for $y_0 = 0.9$. Thus, the prior mode of $\beta$ changes dramatically as $y_0$ is changed. When $y_0 = 0.1$, the prior mode is the same in magnitude but opposite in sign to the case $y = 0.9$.

Table 1 show various summaries of the prior distribution (6.1) and posterior estimates of $\beta$ under several choices of $(y_0, a_0)$. For a given $a_0$, we see that the prior means and standard deviations of $\beta$ are quite different as $y_0$ is varied. For example, for $(y_0, a_0) = (0.1, 1), (0.5, 1), (0.9, 1)$, the prior mean (standard deviation) of $\beta_1$ are 0.067 (1.262), 0.0049 (0.683), -0.019 (1.208), respectively. Here, we see that the prior estimates change dramatically as $y_0$ is varied. A similar phenomenon occurs with the other regression coefficients. Moreover, the prior using $(y_0, a_0) = (0.1, 1), (0.9, 1)$ is highly skewed about its mode as can be seen from the 95% highest prior density intervals. For example, for $(y_0, a_0) = (0.1, 1), (0.9, 1)$, the 95% highest prior density intervals for $\beta_0$ are (-4.328, -1.251) and (1.254, 4.274), respectively. A similar phenomenon occurs with the other regression coefficients. For a given $y_0$, as $a_0$ is increased, the prior becomes more symmetric about its mode, the prior means shrink to the prior modes, and the prior standard deviations decrease. For example, for $(y_0, a_0) = (0.1, 1)$, $(0.1, 10)$, and $(0.1, 100)$, the prior means (standard deviations) of $\beta_1$ are 0.067 (1.262), 0.003 (0.346), and -0.00002 (0.108). A similar phenomenon occurs with the other regression coefficients and other values of $y_0$. Moreover, the prior becomes more symmetric as $a_0$ increases, as can be seen from the 95% highest prior density intervals. For $(y_0, a_0) = (0.1, 1)$, $(0.1, 10)$, and $(0.1, 100)$, the 95% highest prior density intervals for $\beta_0$ are (-4.328, -1.251), (-2.649, -1.846), and (-2.322, -2.076). We mention that using $y_0 = 0.5$ results in symmetry of the prior about its mode, and this can be seen from Table 1. From the 95% highest prior density intervals, we can see that for $y_0 = 0.5$ and for all values of $a_0$, the prior is symmetric about its mode, which is 0. Moreover, for $y_0 = 0.5$ the prior means are very close to the prior mode for all values of $a_0$. This is in contrast to the prior mean behavior for $y_0 = 0.1$ and 0.9. Thus, we see from Table 1 that $y_0 = 0.5$ exhibits several nice properties of the prior (6.1), and thus is a suitable guide value for conducting sensitivity analyses. We also note that $a_0 = 0$ yields posterior estimates of $\beta$ that are very close to the maximum likelihood estimates.

In general, in Table 1, we see that the posterior standard deviations are smaller than the corresponding prior standard deviations and the 95% highest posterior density (HPD) intervals are narrower than the corresponding 95% highest prior density intervals for all combinations of $(y_0, a_0)$. For a given $a_0$, we see that the posterior modes, means and standard deviations of $\beta$ are quite different as $y_0$ is varied. For a given $y_0$, and as $a_0$ is increased, the posterior mean of $\beta$ converges to the posterior mode of $\beta$, and the convergence is fastest when $y_0 = 0.5$. In addition, as $a_0$ increases, the prior dominates the likelihood as can be seen from the posterior estimates of $\beta$. Furthermore, for a given $y_0$, and as $a_0$ increases, the posterior standard deviations decrease and the 95% HPD intervals become narrower and more symmetric about the posterior mode, with the highest degree of symmetry occurring for $y_0 = 0.5$.

Table 1. Summary statistics from the prior and posterior distributions for finney data.

| | | | Prior | | | | Posterior | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $y_0$ | $a_0$ | Parameter | Mean | SD | 95% HPD | | Mean | SD | 95% HPD | |
| 0.1 | 1 | $\beta_0$ | -2.700 | 0.888 | (-4.328, | -1.251) | -1.997 | 0.551 | (-3.105, | -0.974) |
| | | $\beta_1$ | 0.067 | 1.262 | (-2.471, | 2.468) | 1.850 | 0.649 | ( 0.600, | 3.142) |
| | | $\beta_2$ | 0.357 | 0.986 | (-1.263, | 2.237) | 1.688 | 0.694 | ( 0.434, | 3.098) |
| | 10 | $\beta_0$ | -2.245 | 0.205 | (-2.649, | -1.846) | -2.071 | 0.194 | (-2.456, | -1.698) |
| | | $\beta_1$ | 0.003 | 0.346 | (-0.673, | 0.687) | 0.459 | 0.285 | (-0.105, | 1.013) |
| | | $\beta_2$ | 0.039 | 0.232 | (-0.398, | 0.503) | 0.344 | 0.233 | (-0.095, | 0.807) |
| | 100 | $\beta_0$ | -2.2019 | 0.063 | (-2.322, | -2.076) | -2.178 | 0.062 | (-2.300, | -2.058) |
| | | $\beta_1$ | -0.0002 | 0.108 | (-0.210, | 0.210) | 0.061 | 0.106 | (-0.147, | 0.266) |
| | | $\beta_2$ | 0.0039 | 0.070 | (-0.134, | 0.141) | 0.040 | 0.070 | (-0.096, | 0.178) |
| 0.5 | 1 | $\beta_0$ | 0.0014 | 0.406 | (-0.804, | 0.796) | -0.502 | 0.325 | (-1.147, | 0.123) |
| | | $\beta_1$ | 0.0049 | 0.683 | (-1.317, | 1.368) | 1.342 | 0.537 | ( 0.313, | 2.415) |
| | | $\beta_2$ | -0.0025 | 0.482 | (-0.979, | 0.940) | 0.897 | 0.413 | ( 0.119, | 1.729) |
| | 10 | $\beta_0$ | 0.0002 | 0.121 | (-0.235, | 0.236) | -0.071 | 0.114 | (-0.301, | 0.149) |
| | | $\beta_1$ | 0.0003 | 0.207 | (-0.415, | 0.398) | 0.213 | 0.198 | (-0.165, | 0.611) |
| | | $\beta_2$ | 0.0003 | 0.135 | (-0.269, | 0.264) | 0.127 | 0.130 | (-0.132, | 0.379) |
| | 100 | $\beta_0$ | -0.0004 | 0.038 | (-0.074, | 0.074) | -0.007 | 0.038 | (-0.079, | 0.067) |
| | | $\beta_1$ | 0.0006 | 0.065 | (-0.129, | 0.124) | 0.023 | 0.065 | (-0.104, | 0.148) |
| | | $\beta_2$ | 0.0006 | 0.042 | (-0.083, | 0.083) | 0.013 | 0.042 | (-0.070, | 0.096) |
| 0.9 | 1 | $\beta_0$ | 2.668 | 0.787 | ( 1.254, | 4.274) | 0.496 | 0.299 | (-0.086, | 1.086) |
| | | $\beta_1$ | -0.019 | 1.208 | (-2.416, | 2.370) | 1.703 | 0.637 | ( 0.494, | 2.980) |
| | | $\beta_2$ | -0.324 | 0.898 | (-2.167, | 1.286) | 0.885 | 0.365 | ( 0.194, | 1.623) |
| | 10 | $\beta_0$ | 2.243 | 0.204 | ( 1.852, | 2.649) | 1.753 | 0.156 | ( 1.444, | 2.057) |
| | | $\beta_1$ | -0.004 | 0.346 | (-0.694, | 0.665) | 0.465 | 0.299 | (-0.125, | 1.048) |
| | | $\beta_2$ | -0.038 | 0.230 | (-0.496, | 0.404) | 0.227 | 0.172 | (-0.118, | 0.554) |
| | 100 | $\beta_0$ | 2.2015 | 0.062 | ( 2.081, | 2.325) | 2.139 | 0.061 | ( 2.020, | 2.257) |
| | | $\beta_1$ | 0.0003 | 0.108 | (-0.211, | 0.215) | 0.061 | 0.107 | (-0.146, | 0.272) |
| | | $\beta_2$ | -0.0038 | 0.070 | (-0.145, | 0.133) | 0.032 | 0.068 | (-0.103, | 0.163) |

Table 2 summarizes posterior estimates of $\beta$ using a random $a_0$ with $y_0 = 0.5$. We consider three sets of hyperparameters for $a_0$. These are (i) $(\alpha_0, \lambda_0) = (0.1, 0.1)$, (ii) $(\alpha_0, \lambda_0) = (10, 10)$, and (iii) $(\alpha_0, \lambda_0) = (100, 100)$. Here, (i) implies a noninformative prior for $a_0$, (ii) implies a moderately informative prior for $a_0$, and (iii) implies an informative prior for $a_0$. From Table 2, we see that with $(\alpha_0, \lambda_0) = (0.1, 0.1)$, the posterior estimates of $\beta$ are close to the estimates corresponding to $a_0 = 0$. As the prior for $a_0$ becomes more informative, the posterior mean of $a_0$ increases, and as a result, the posterior estimates of $\beta$ change a lot. For example, when $(\alpha_0, \lambda_0) = (100, 100)$, we see that the posterior estimates of $\beta$ are close to those of Table 1 corresponding to $(y_0, a_0) = (0.5, 1)$.

Table 2. Summary statistics from the posterior distribution with random $a_0$ for finney data.

| $(\alpha_0, \lambda_0)$ | $E(a_0\|D)$ $(SD(a_0\|D))$ | Parameter | Mean | SD | 95% HPD Interval |
|---|---|---|---|---|---|
| (0.1, 0.1) | 0.002 (0.005) | $\beta_0$ | -3.723 | 1.594 | (-6.851, -0.829) |
| | | $\beta_1$ | 6.594 | 2.607 | ( 1.978, 11.727) |
| | | $\beta_2$ | 5.837 | 2.381 | ( 1.599, 10.564) |
| (10, 10) | 0.188 (0.067) | $\beta_0$ | -1.611 | 0.814 | (-3.245, -0.151) |
| | | $\beta_1$ | 3.371 | 1.291 | ( 1.070, 5.989) |
| | | $\beta_2$ | 2.694 | 1.179 | ( 0.665, 5.049) |
| (100, 100) | 0.757 (0.078) | $\beta_0$ | -0.623 | 0.379 | (-1.382, 0.094) |
| | | $\beta_1$ | 1.611 | 0.633 | ( 0.404, 2.891) |
| | | $\beta_2$ | 1.107 | 0.507 | ( 0.189, 2.134) |

Table 3 shows posterior estimates of $\beta$ based on the asymptotic prior (2.10) using $(y_0, a_0) = (0.5, 1), (0.5, 100)$. We see from this table that the posterior estimates of $\beta$ are fairly close to the posterior estimates of Table 1, which use (6.1). For example, for $(y_0, a_0) = (0.5, 1)$, the posterior mean (standard deviation) of $\beta_1$ from Table 3 is 1.228 (0.487), compared to 1.342 (0.537) from Table 1. Thus, we see that even with a fairly small sample size of $n = 39$, the asymptotic prior in (2.10) provides a somewhat fair approximation to (6.1). As $a_0$ is increases, the posterior estimates of $\beta$ from Tables 1 and 3 are much closer together since, in this case, the prior dominates the likelihood and the priors (6.1) and (2.10) become highly peaked at the mode. Finally, Figure 1 shows three dimensional plots of the marginal prior for $(\beta_1, \beta_2)$ using $(y_0, a_0) = (0.5, 1), (0.5, 10)$, respectively. We see from these plots that the prior is symmetric, and becomes more concentrated about the mode as $a_0$ is increased.

Table 3. Posterior summaries based on asymptotic prior for finney data.

| $a_0$ | Parameter | Mean | SD | 95% HPD Interval |
|---|---|---|---|---|
| 1 | $\beta_0$ | -0.416 | 0.281 | (-0.978, 0.121) |
| | $\beta_1$ | 1.228 | 0.487 | ( 0.292, 2.199) |
| | $\beta_2$ | 0.749 | 0.330 | ( 0.104, 1.401) |
| 100 | $\beta_0$ | -0.008 | 0.037 | (-0.078, 0.069) |
| | $\beta_1$ | 0.023 | 0.064 | (-0.103, 0.149) |
| | $\beta_2$ | 0.013 | 0.042 | (-0.068, 0.095) |

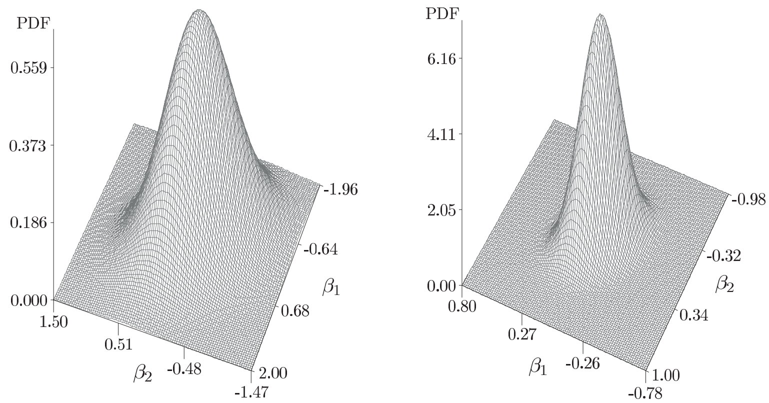


Figure 1. Joint prior distributions for $(\beta_1, \beta_2)$ with $a_0 = 1$ (left) and $a_0 = 10$ (right) for Finney data.

## Appendix : Proofs of Theorems

**Proof of Theorem 2.1.** Without loss of generality, take $\tau = 1$. We make use of a technique in Ibrahim and Laud (1991). It suffices to show, for $t$ in a neighborhood of 0, the finiteness of

$$\int_{R^p} \exp(t'\beta)\ \exp\{a_0[y_0'\theta(X\beta) - J'b(\theta(X\beta))]\}\ d\beta, \tag{A.1}$$

where $R^k$ denotes $p$-dimensional Euclidean space. Partition a row permutation of $X$ into $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, where $X_1$ is a $p \times p$ full rank matrix and $X_2$ is an $(n-p) \times p$ matrix. Correspondingly, partition $y_0$, $\theta(.)$, and $b(.)$. Now (A.1) takes the form

$$\int_{R^p} \exp(t'\beta) \exp\{a_0[y_{10}'\theta(X_1\beta) - J_1'b_1(\theta(X_1\beta))]\} \\ \times \exp\{a_0[y_{20}'\theta(X_2\beta) - J_2'b_2(\theta(X_2\beta))]\}\ d\beta. \tag{A.1}$$

Since the prior density of $\beta$ is assumed bounded, there exists a constant $K_1$ such that $\exp\{a_0[y_{20}'\theta(X_2\beta) - J_2'b_2(\theta(X_2\beta))]\} \leq K_1$. Thus (A.1) is less than or equal to

$$K_1 \int_{R^p} \exp(t'\beta) \exp\{a_0[y_{10}'\theta(X_1\beta) - J_1'b_1(\theta(X_1\beta))]\}. \tag{A.2}$$

Now make the transformations $u = X_1\beta$ and $r = \theta(u) = (\theta(u_1), \ldots, \theta(u_p))' = (r_1, \ldots, r_p)'$. After dropping unnecessary constants, (A.2) reduces to

$$\int_{\Theta} |J_2(r)| \exp(s'\theta^{-1}(r)) \exp\{a_0[y_{10}'r - J_1'b_1(r)]\}dr, \tag{A.3}$$

where $\Theta = \Theta_1 \times \cdots \times \Theta_p \subset R^p$, and $\Theta_i$ is the parameter space of the one dimensional canonical parameter $r_i$. Further, $s' = t'X_1^{-1}$, and the Jacobian of the second transformation is given by $J_2(r) = \prod_{i=1}^{p} \frac{d}{dr_i} \{\theta^{-1}(r_i)\}$. If the link is canonical, then $\theta^{-1}(r) = r$, and (A.3) reduces to

$$\int_{\Theta} \exp\left\{a_0\left[(y_{10}' + a_0^{-1}s')r \ - \ J_1'b_1(r)\right]\right\} \, dr. \tag{A.4}$$

Since the exponential family density in (2.1) is obtained as a product of $n$ exponential densities on subsets of $R^1$, the integrand in (A.4) is an exponential family density with the observable in $R^p$ and canonical parameter $u \in \Theta \subset R^p$. Denoting by $\mathcal{Z}$ the interior of the convex hull of the support set of the latter exponential family density, we see that $y_0 \in \mathcal{Y}$ implies $y_{10} \in \mathcal{Z}$. (Both $\mathcal{Y}$ and $\mathcal{Z}$ are, in fact, open rectangles). Now, since $\mathcal{Z}$ is open, there exists an open neighborhood of 0 such that for every $s$ in this neighborhood, $y_{10} + a_0^{-1}s \in \mathcal{Z}$. Application of Theorem 1 of Diaconis and Ylvisaker (1979, p.272) to (A.4) proves part (i).

For (ii), (A.3) can be written as a product of the $p$ one dimensional integrals

$$\int_{\Theta_i} \left|\frac{d}{dr_i}\theta^{-1}(r_i)\right| \exp\{a_0[y_{10i}r_i + a_0^{-1}s_i\theta^{-1}(r_i) - b_{1i}(r_i)]\} \, dr_i, \quad i = 1, \ldots, p, \tag{A.5}$$

where $y_{10i}$ is the $ith$ component of $y_{10}$, and $b_{1i}(r_i)$ is the $ith$ component of $b_1(r)$. Thus (A.2) is finite if each integral in (A.5) is. This proves part (ii).

**Proof of Theorem 4.1.** Let $L(\beta|y_0,\tau) = \exp\{\tau[y_0'\theta(\eta) - J'b(\theta(\eta))] + J'c(y_0,\tau)\}$. Since $\lambda_0 > \sup_{\beta \in R^p}[\tau(y_0'\theta(\eta) - J'b(\theta(\eta))) + J'c(y_0,\tau)] = \ln L(\beta|y_0,\tau)$, it is easy to see that

$$\int_0^{\infty} [L(\beta|y_0,\tau)]^{a_0} a_0^{\alpha_0 - 1} \exp(-\lambda_0 a_0) da_0 = K_0[\lambda_0 - \ln L(\beta|y_0,\tau)]^{-\alpha_0}, \tag{A.6}$$

where $K_0$ is a constant independent of $\beta$. Using (A.6), for some $t_0^* > 0$, we have

$$\begin{aligned}
&\int_{R^p}\int_0^{\infty} ||\beta||^k \pi(\beta, a_0|y_0,\tau) da_0 d\beta \\
&= K_0 \int_{R^p} ||\beta||^k [\lambda_0 - \ln L(\beta|y_0,\tau)]^{-\alpha_0} 1_{\{L(\beta|y_0,\tau) > \exp(-t_0^*||\beta||)\}} d\beta \\
&\quad + K_0 \int_{R^p} ||\beta||^k [\lambda_0 - \ln L(\beta|y_0,\tau)]^{-\alpha_0} 1_{\{L(\beta|y_0,\tau) \le \exp(-t_0^*||\beta||)\}} d\beta \\
&\le K_1 \int_{R^p} ||\beta||^k L(\beta|y_0,\tau) \exp(t_0^*||\beta||) d\beta + K_0 \int_{R^p} ||\beta||^k (\lambda_0 + t_0^*||\beta||)^{-\alpha_0} d\beta < \infty,
\end{aligned}$$

where $K_1 > 0$ is a constant. Theorem 2.1 ensures that the first integral is finite, while the second integral is finite since $\alpha_0 > p + k$. This proves (4.3).

**Proof of Theorem 5.2.** Integrating out $\tau$ yields

$$\int_0^\infty \pi(\beta,\tau,a_0|y_0)d\tau$$
$$\leq K_1[\delta_0 + a_0(y_0 - X\beta)'(y_0 - X\beta)/2]^{-(a_0 n/2+\zeta_0)} a_0^{\alpha_0 - 1}\exp(-\lambda_0 a_0), \quad \text{(A.7)}$$

where $K_1 > 0$ is a constant. Since $X$ is of full rank, there exists a positive constant $K_2$ so that the right hand side of (A.7) is less than $K_2[\delta_0 + a_0||\beta||^2/2]^{-(a_0 n/2+\zeta_0)} a_0^{\alpha_0-1}\exp(-\lambda_0 a_0)$. For some $s_0 > 0$ and $K_3 > 0$, we have

$$\int_0^\infty \int_{R^p} \left[\delta_0 + a_0||\beta||^2/2\right]^{-(a_0 n/2+\zeta_0)} a_0^{\alpha_0-1}\exp(-\lambda_0 a_0)d\beta da_0$$
$$\leq K_3 \int_{||\beta||\leq s_0}\int_0^\infty \delta_0^{-a_0 n/2} a_0^{\alpha_0-1}\exp(-\lambda_0 a_0)da_0 d\beta$$
$$+K_3\int_{||\beta||>s_0} ||\beta||^{-2\zeta_0} d\beta \int_0^\infty a_0^{-(a_0 n/2+\zeta_0)} a_0^{\alpha_0-1}\exp(-\lambda_0 a_0) < \infty$$

if the assumptions given in Theorem 5.2 hold. This completes the proof.

## References


Bedrick, E. J., Christensen, R. and Johnson, W. (1996). A new perspective on priors for generalized linear models. *J. Amer. Statist. Assoc.* **91**, 1450-1461.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7**, 269-281.

Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34**, 320-334.

Ibrahim, J. G. and Chen, M.-H (2000). Power prior distributions for regression models. *Statist. Sci.* **15**, 46-60.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *J. Amer. Statist. Assoc.* **86**, 981-986.

Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10**, 65-80.

Morris, C. N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.* **11**, 515-529.

Pregibon, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9**, 705-724.

Sun, D., Tsutakawa, R. K. and He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statist. Sinica* **11**, 77-95.



Department of Statistics, University of Connecticut, 215 Glenbrock Road, U-4120, Storrs, CT, 06269-4120, U.S.A.

E-mail: mhchen@stat.uconn.edu

Department of Biostatistics, University of North Carolina, MaGavran-Greenberg Hall, CB#7420, Chapel Hill, NC 27559.

E-mail: ibrahim@bios.unc.edu


(Received July 2001; accepted October 2002)