Journal of Statistical Planning and Inference 84 (2000) 121–137

journal of statistical planning and inference

www.elsevier.com/locate/jspi

# Power prior distributions for generalized linear models

Ming-Hui Chen[a], Joseph G. Ibrahim[b,*], Qi-Man Shao[c]

[a] *Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280, USA*

[b] *Department of Biostatistics, Harvard University School of Public Health and Dana Farber Cancer Institute 44 Binney St. Boston, MA 02115, USA*

[c] *Department of Mathematics, University of Oregon, Eugene, OR 97403-1222, USA*

Received 8 October 1998; received in revised form 7 June 1999; accepted 28 June 1999

**Abstract**

In this article, we propose a class of prior distributions called the *power prior distributions*. The power priors are based on the notion of the availability of historical data, and are of great potential use in this context. We demonstrate how to construct these priors and elicit their hyperparameters. We examine the theoretical properties of these priors in detail and obtain some very general conditions for propriety as well as lower bounds on the normalizing constants. We extensively discuss the normal, binomial, and Poisson regression models. Extensions of the priors are given along with numerical examples to illustrate the methodology. © 2000 Elsevier Science B.V. All rights reserved.

*MSC*: Primary 62A15; Secondary 62A99

*Keywords*: Prior elicitation; Posterior distribution; Propriety; Variable selection

## 1. Introduction

Prior elicitation is one of the most important issues in Bayesian data analysis. When no prior information is available, a non-informative prior such as a uniform prior, Jeffreys prior, or reference prior can be used (see Kass and Wasserman, 1996) for a list of such non-informative priors. However, real prior information such as historical data or data from previous similar studies is often available in applied research settings where the investigator has access to previous studies measuring the same response and

* Corresponding author.

*E-mail addresses:* mhchen@wpi.edu (M-H. Chen), ibrahim@jimmy.harvard.edu (J.G. Ibrahim), qmshao@darkwing.uoregon.edu (Q-M. Shao)



PII: S0378-3758(99)00140-8

covariates as the current study. For example, in many cancer and AIDS clinical trials, current studies often use treatments that are very similar or slight modifications of treatments used in previous studies. In carcinogenicity studies, large historical databases exist for the control animals from previous experiments. In experiments conducted over time, data from previous time periods can often be used as prior information. We shall generically refer to the data from a previous study (or studies) as historical data throughout. In all of these situations, it is natural to incorporate the historical data into the current study by quantifying it with a suitable prior distribution on the model parameters.

Our approach is based on the notion of specifying an $n_0 \times 1$ historical data response vector, $y_0$, along with an $n_0 \times k$ covariate matrix $X_0$ corresponding to $y_0$. Then, the historical data $D_0 = (n_0, y_0, X_0)$, along with a precision parameter $a_0$ that weights the historical data relative to the data from the current study, is used to specify an automated parametric informative prior for the regression coefficients. The basic general idea is that the prior is taken to be the likelihood function based on the historical data $D_0$, raised to the power $a_0$. Thus, the power parameter $a_0$, $0 \leqslant a_0 \leqslant 1$, controls the influence of the historical data on the current data. Moreover, a prior distribution is specified for $a_0$. We will refer to priors constructed in this fashion as *power priors* throughout. The power priors can be extended to any parametric or semi-parametric models including generalized linear models, non-linear models, random effects models, and models for survival data. In this paper, we focus our theoretical development on the class of generalized linear models. In carcinogenicity studies, $D_0$ is taken as the raw historical data (see Ibrahim, et al., 1998), and in clinical trials, $D_0$ typically consists of a similar previous study using the same treatment. In general, $D_0$ can be elicited via a theoretical prediction model, expert opinion, or case-specific information. However, the most natural specification of $D_0$ is the raw historical data from a similar previous study.

The initial idea of the power prior can be traced to Diaconis and Ylvisaker (1979) and Morris (1983), where they studied conjugate priors for exponential families. However, these two authors did not consider the regression problem, and in addition, only considered the situation in which the power $a_0$, is a fixed constant. In the regression setting with $a_0$ random, the formulation becomes quite complicated and the theoretical properties of the power priors remain largely unknown. In this article, we provide a comprehensive study of the theoretical properties of power priors for the class of generalized linear models.

The outline of this article is given as follows. In Section 2, we give the necessary notation and development of the power prior distributions for the class of generalized linear models. In Section 3, we explore various properties of the power prior distributions, including the role of prior parameters, present several theorems characterizing necessary and sufficient conditions for propriety of the priors, and give results concerning the lower bound of the normalizing constants. We demonstrate the priors with a real example in Section 4. We conclude the article with a discussion section.

## 2. Development of power prior distributions

Suppose that $\{(x_i, y_i), i = 1, 2, \ldots, n\}$ is a sample of independent observations from the current study, where each $y_i$ is the response variable, $x_i = (x_{i1}, x_{i2}, \ldots, x_{ik})'$ is a $k \times 1$ random vector of covariates with $x_{i1} = 1$ denoting an intercept. We use the generic label $f(u_1|u_2)$ to denote the conditional density of $u_1$ given $u_2$ throughout, where $u_1$ may be discrete or continuous. Suppose that given $x_i$, $y_i$ has a density in the exponential class with the form

$$f(y_i|x_i, \theta_i, \tau) = \exp\left\{\alpha_i^{-1}(\tau)(y_i\theta_i - \psi(\theta_i)) + \phi(y_i, \tau)\right\}, \quad i = 1, \ldots, n, \tag{2.1}$$

indexed by the canonical parameter $\theta_i$ and the scale parameter $\tau$. The functions $\psi$ and $\phi$ determine a particular family in the class, such as the binomial, normal, Poisson, etc. The function $\alpha_i(\tau)$ is commonly of the form $\alpha_i(\tau) = \tau^{-1}s_i^{-1}$, where the $s_i$'s are known weights. Further suppose the $\theta_i$'s satisfy the equations

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

and

$$\eta_i = x_i'\beta, \tag{2.3}$$

where $h$ is a monotone differentiable function, often referred to as the link function and $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a $k \times 1$ vector of regression coefficients. We consider a few special cases of (2.1).

*Normal*: $y_i|\theta_i \sim \mathrm{N}(\theta_i, \sigma^2)$. We take $\theta_i = \eta_i = x_i'\beta$, $\tau = \sigma^2$, $\alpha_i(\tau) = \tau$, $\psi(\theta_i) = \theta_i^2/2$, and $\phi(y_i, \tau) = -\frac{1}{2}(y_i^2/\sigma^2 + \log(2\pi\sigma^2))$.

*Binomial*: $y_i|n_i, \theta_i \sim b(n_i, F(\eta_i))$, $\theta_i = \log(F(\eta_i)/(1 - F(\eta_i)))$, $F$ is a cumulative distribution function with $0 < F(\eta_i) < 1$, $\tau = 1$, $\alpha_i(\tau) = \tau = 1$, $\psi(\theta_i) = -n_i \log(1 + \theta_i)$, $\phi(y_i, \tau) = 0$. This class of models includes the logit, probit, and complementary log–log models. Here $b(n_i, p_i)$ denotes the binomial distribution with success probability $p_i$ and sample size $n_i$.

*Poisson*: $y_i|\theta_i \sim \mathrm{Poisson}(\exp(\theta_i))$. We take $\theta_i = \eta_i$, $\tau = 1$, $\alpha_i(\tau) = \tau = 1$, $\psi(\theta_i) = \exp(\theta_i)$, and $\phi(y_i, \tau) = -\log(y_i!)$.

We assume that $\tau$ is known and denote $\alpha_i = \alpha_i(\tau)$ and $\phi(y_i) = \phi(y_i, \tau)$ throughout the remainder of the paper. For the binomial and Poisson regression models, $\tau$ is intrinsically equal to 1.

Let $y = (y_1, \ldots, y_n)'$ denote an $n \times 1$ vector of independent responses and $X$ denotes an $n \times k$ matrix of fixed covariates with the $i$th row $x_i'$, $i = 1, 2, \ldots, n$. Let $D = (n, y, X)$ denote the data for the current study. Then, the likelihood function based on the current data $D$ is given by

$$L(\beta|D) = \prod_{i=1}^{n} \exp\left\{\alpha_i^{-1}(y_i\theta_i - \psi(\theta_i)) + \phi(y_i)\right\}. \tag{2.4}$$

Our power prior construction is based on the notion of the existence of historical data measuring the same response variable and covariates as the current study. For ease of exposition, we assume only one historical dataset. However, the extension to multiple historical datasets will be considered in Section 3.4. To this end, let $n_0$ denote sample size for the historical data, $y_0$ be an $n_0 \times 1$ response vector for the historical data, and $X_0$ is an $n_0 \times k$ matrix of covariates corresponding to $y_0$. Also, let $y_{0i}$ denote the $i$th component of $y_0$, denote $x'_{0i} = (x_{0i1}, x_{0i2}, \ldots, x_{0ik})$ to be the $i$th row of $X_0$ with $x_{0i1} = 1$ corresponding to an intercept, $\eta_{0i} = x'_{0i}\beta$ denotes the linear predictor based on the historical data, and $D_0 = (n_0, y_0, X_0)$ denotes the historical data. Then, the likelihood function of $\beta$ based on the historical data $D_0$ is given by

$$L(\beta|D_0) = \prod_{i=1}^{n_0} \exp\left\{\alpha_i^{-1}(y_{0i}\theta_{0i} - \psi(\theta_{0i})) + \phi(y_{0i})\right\}, \tag{2.5}$$

where

$$\theta_{0i} = h(\eta_{0i}).$$

The power prior for the class of GLMs takes the form

$$\pi(\beta|D_0, a_0) \propto \left[L(\beta|D_0)\right]^{a_0}, \tag{2.6}$$

where $0 \leqslant a_0 \leqslant 1$ is a scalar prior precision parameter that weights the historical data relative to the likelihood of the current study. Small values of $a_0$ give little prior weight to the historical control data relative to the likelihood of the current study whereas values of $a_0$ close to 1, for example, give roughly equal weight to the prior and the likelihood of the current study. The case $a_0 = 0$ results in an improper uniform prior on $\beta$, resulting in no incorporation of historical data. The parameter $a_0$ allows the investigator to control the influence of the historical data on the current study. Such control is important in cases where there is heterogeneity between the historical data and the current study, or when the sample sizes of the two studies are quite different.

The prior specification is completed by specifying a prior distribution for $a_0$. We take a beta prior for $a_0$, and thus we propose a joint prior distribution for $(\beta, a_0)$ of the form

$$\pi(\beta, a_0|D_0) \propto \left[L(\beta|D_0)\right]^{a_0} a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1}, \tag{2.7}$$

where $(\delta_0, \lambda_0)$ are specified prior hyperparameters. We see that (2.7) will not have a closed form in general, but has several attractive properties as discussed in the subsequent sections. The joint posterior distribution of $\beta$ and $a_0$ is given by

$$p(\beta, a_0|D, D_0) \propto L(\beta|D) \left[L(\beta|D_0)\right]^{a_0} a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1}. \tag{2.8}$$

In the next several sections, we will explore various properties of power prior $\pi(\beta, a_0|D_0)$ given by (2.7).

## 3. Properties of power priors

### 3.1. Roles of prior parameters

When historical data are available, it can be seen from (2.7) that the power prior $\pi(\beta, a_0|D_0)$ is proportional to the product of the likelihood function of $\beta$ based on the historical data $D_0$ taken to the power $a_0$ multiplied by the (beta) prior of $a_0$.

The prior parameter $a_0$ is a precision parameter. One of the main roles of $a_0$ in (2.7) is that $a_0$ controls the heaviness of the tails of the prior for $\beta$. As $a_0$ becomes smaller, the tails of (2.6) become heavier. Taking $a_0$ to be random instead of fixed allows the tails of the marginal distribution of $\beta$ to be heavier than the tails with $a_0$ fixed, and this may be more desirable. Another advantage of allowing $a_0$ to be random is that it allows flexibility in expressing our uncertainty about $a_0$ via a prior mean and variance. Then the weight given to the historical data relative to the current study will be dictated by $D_0$ as well as the prior parameters $\delta_0$ and $\lambda_0$.

The main reasons for taking a beta prior for $a_0$ are (i) the beta prior is simple and natural; (ii) when $\lambda_0 \to \infty$, $\pi(\beta, a_0|D_0)$ becomes an improper uniform prior for $\beta$, resulting in no incorporation of the historical data; (iii) when $\delta_0 \to \infty$, the historical data and the current data become equally weighted. For elicitation purposes, it is easier to work with the prior mean and standard deviation of $a_0$, that is, $\mu_{a_0} = \delta_0/(\delta_0 + \lambda_0)$ and $\sigma_{a_0} = (\mu_{a_0}(1-\mu_{a_0}))^{1/2}(\delta_0+\lambda_0+1)^{-1/2}$. It is typically easier to specify $(\mu_{a_0}, \sigma_{a_0})$ and then solve for $(\delta_0, \lambda_0)$ from the implied equations. The investigator may choose $\mu_{a_0}$ to be small if he/she wishes to have low prior weight on the historical data. If a large prior weight is desired, then $\mu_{a_0} \geqslant 0.5$ may be suitable. It is reasonable to choose $\sigma_{a_0}$ to be of the form $\sigma_{a_0} = (p_0 \mu_{a_0})^{1/2}$, where $0 < p_0 < 1$ is a chosen scalar. In practice, we recommend that several choices of $(\mu_{a_0}, \sigma_{a_0})$ be used, including ones that give small and large weight to the historical data, and several sensitivity analyses be conducted. We do not recommend doing an analysis based on one set of prior parameters. The choices recommended here can be used as starting points from which sensitivity analyses can be based.

To illustrate the roles of the prior parameters in the power priors, we consider the following logistic regression model. We simulated a data set consisting $n_0 = 200$ independent Bernoulli observations with success probability

$$p_{0i} = \frac{\exp\{-0.5 + 0.5x_{0i}\}}{1 + \exp\{-0.5 + 0.5x_{0i}\}}, \qquad i = 1, \ldots, n_0 ,$$

where the $x_{0i}$ are i.i.d. normal random variables with mean 0 and standard deviation 0.5. Using the Gibbs sampler, for each given set of $(\delta_0, \lambda_0)$, we generated 50,000 iterates from the joint prior distribution $\pi(\beta, a_0|D_0)$ given by (2.7). The detailed implementation scheme of the Gibbs sampler can be found in Chen et al. (1999). Using the Gibbs output, we estimated the marginal prior densities for $\beta_1$ (intercept) and $\beta_2$ (slope). Fig. 1 shows the marginal prior densities of $\beta_1$ and $\beta_2$ for three choices of $(\mu_{a_0}, \sigma_{a_0})$. From this figure, it is easy to see that as $\mu_{a_0}$ gets smaller, both marginal prior density curves get flatter; but the prior modes of $\beta_1$ and $\beta_2$ for all three choices of $(\mu_{a_0}, \sigma_{a_0})$ are almost the same. Although it is not shown in Fig. 1, we also obtained the marginal prior densities for $\beta_1$ and $\beta_2$ for $(\delta_0, \lambda_0) = (3, 3)$, which are nearly uniform over the real line. For Fig. 1, the corresponding $(\delta_0, \lambda_0)$ values are $(50, 3)$, $(20, 20)$, and $(5, 5)$, respectively.

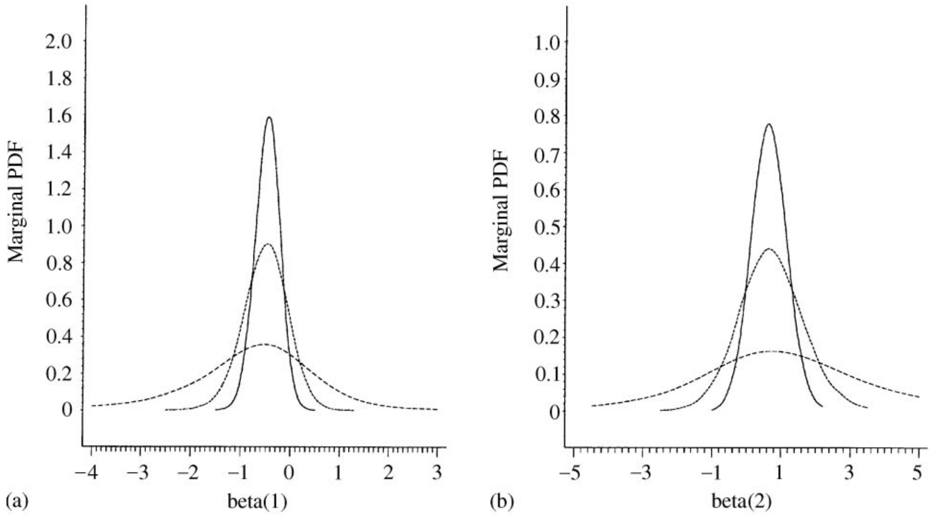


Fig. 1. Plots of marginal posterior densities for $\beta_1$ and $\beta_2$ where the solid curve is for $(\mu_{a_0},\sigma_{a_0})$=(0.94, 0.031), the dotted curve is for $(\mu_{a_0},\sigma_{a_0}) = (0.5, 0.078)$, and the dashed curve is for $(\mu_{a_0},\sigma_{a_0}) = (0.5, 0.151)$.

### *3.2. Propriety of power prior distributions*

Any informative Bayesian analysis necessarily requires a proper prior distribution. It is thus critical to examine the conditions under which the joint prior for $\beta$ and $a_0$ is proper. This issue is crucial in Bayesian variable selection as it is well known that Bayesian variable selection requires a proper prior distribution. It is also an important issue in Bayesian hypothesis testing problems, and in particular, in the calculation of Bayes factors and related quantities (see, for example, Berger, 1985, pp. 145–157). Here, we establish some very general results concerning the propriety of the joint prior distribution of $(\beta, a_0)$ for the generalized linear regression models. These results can be extended to other models as well. Towards this goal, we first introduce a useful lemma.

**Lemma 3.1.** *Let $\delta_0 > 0$, $\lambda_0 > 0$. Then there exist $K_1 = K_1(\delta_0,\lambda_0) > 0$ and $K_2 = K_2(\delta_0,\lambda_0) > 0$ such that $\forall\ 0 \leqslant t \leqslant 1$,*

$$K_1(1+\ln(1/t))^{-\delta_0} \leqslant \int_0^1 t^{a_0} a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\,\mathrm{d}a_0 \leqslant K_2(1+\ln(1/t))^{-\delta_0}. \tag{3.1}$$

The proof of the lemma is given in the appendix.

Using Lemma 3.1, we obtain the following result, which ensures the propriety of the joint prior distribution $\pi(\beta, a_0|D_0)$.

**Theorem 3.1.** *Assume that*

$$\exp\left\{\alpha_i^{-1}(y_{0i}\theta_{0i} - \psi(\theta_{0i}))\right\} \leqslant M \tag{3.2}$$

*for* $i=1,2,\ldots,n_0$, *where* $M$ *is some finite constant. Suppose there exist* $y_{0i_1}$, $y_{0i_2}$, $\ldots$, $y_{0i_k}$ $(1 \leqslant i_1 \leqslant i_2 \leqslant \cdots \leqslant i_k)$ *such that*

$$\int_{-\infty}^{\infty} \mathrm{e}^{t_0|\eta|} \exp\left\{\alpha_{i_j}^{-1}(y_{0i_j}h(\eta) - \psi(h(\eta))\right\} \mathrm{d}\eta < \infty, \tag{3.3}$$

*or*

$$\int_{-\infty}^{\infty} \mathrm{e}^{t_0\eta^2} \exp\left\{\alpha_{i_j}^{-1}(y_{0i_j}h(\eta) - \psi(h(\eta))\right\} \mathrm{d}\eta < \infty \tag{3.4}$$

*for some* $t_0>0$ *and* $j=1,2,\ldots,k$ *and the corresponding design matrix* $\left(x_{0i_1}, x_{0i_2}, \ldots, x_{0i_k}\right)'$ *has full rank* $k$. *Then, the joint prior distribution* $\pi(\beta, a_0|D_0)$ *is proper,* i.e.,

$$\int\int_0^1 \left[L(\beta|D_0)\right]^{a_0} \; a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\,\mathrm{d}a_0\,\mathrm{d}\beta < \infty \tag{3.5}$$

*if one of the following conditions is satisfied:*

(i) $\delta_0 > k$, $\lambda_0 > 0$, *and* (3.3) *holds,*

(ii) $\delta_0 > k/2$, $\lambda_0 > 0$, *and* (3.4) *holds.*

The proof of the theorem is given in the appendix.

**Remark 3.1.** When $y_{0i}$ has a binomial distribution $b(n_{0i}, F(\eta_i))$, then (3.3) holds for logit and log–log link models and (3.4) holds for probit models as long as $0 < y_{0i} < n_{0i}$. It can also be observed that (3.3) holds for the Poisson model as long as $y_{0i} > 0$. Furthermore (3.4) always holds for normal models.

**Remark 3.2.** Conditions (i) and (ii) of Theorem 3.1 are sufficient but *not* necessary conditions for the propriety of (3.5). However, the conditions on $\delta_0$ are somehow necessary; see Section 3.3 for detailed discussions.

**Remark 3.3.** As discussed in Remark 3.1, the conditions stated in Theorem 3.1 hold for many generalized linear models. However, when the $y_{0i}$ are binary responses, i.e., $y_{0i} = 0, 1$, we have

$$L(\beta|D_0) = \prod_{i=1}^{n_0} \left[F(x_{0i}'\beta)\right]^{y_{0i}} \left[1 - F(x_{0i}'\beta)\right]^{1-y_{0i}}. \tag{3.6}$$

In this case, neither (3.3) nor (3.4) will be satisfied. Also as mentioned in Remark 3.2, both (3.3) and (3.4) are only sufficient. Indeed, under some additional regularity conditions on the $x_{0i}$, (3.5) still holds. The following lemma is due to Shao and Chen (1997).

**Lemma 3.2.** *Let* $I_0 = \{1 \leqslant i \leqslant n_0 : y_{0i} = 0\}$ *and* $J_0 = \{1 \leqslant j \leqslant n_0 : y_{0j} = 1\}$.

(a) *Assume the following conditions are satisfied:*

(C1) $I_0$ *and* $J_0$ *are non-empty sets;*

(C2) $\forall \beta^* = (\beta_2, \ldots, \beta_k)' \neq 0$,

$$\min_{j \in J_0} \left(\sum_{l=2}^{k} x_{0jl}\beta_l\right) < \max_{i \in I_0} \left(\sum_{l=2}^{k} x_{0il}\beta_l\right) \quad \textit{if } k \geqslant 2,$$

(C3) $\int_{-\infty}^{\infty} |u|^k \, \mathrm{d}F(u) < \infty$.

*Then*

$$\int_{-\infty}^{\infty} L(\beta|D_0)\, \mathrm{d}\beta < \infty. \tag{3.7}$$

(b) *If* $0 < F(0-) \leqslant F(0) < 1$, *then* (C1) *and* (C2) *are necessary conditions for* (3.7).

(c) *Assume that* (C1) *and* (C2) *are satisfied. If* $\int_{-\infty}^{\infty} \mathrm{e}^{t_0|u|}\, \mathrm{d}F(u) < \infty$ *for some* $t_0 > 0$, *then*

$$\int_{-\infty}^{\infty} \mathrm{e}^{t_0^* ||\beta||} L(\beta|D_0)\, \mathrm{d}\beta < \infty \tag{3.8}$$

*for some* $t_0^* > 0$.

(d) *Assume that* (C1) *and* (C2) *are satisfied. If* $\int_{-\infty}^{\infty} \mathrm{e}^{t_0 u^2}\, \mathrm{d}F(u) < \infty$ *for some* $t_0 > 0$, *then*

$$\int_{-\infty}^{\infty} \mathrm{e}^{t_0^* ||\beta||^2} L(\beta|D_0)\, \mathrm{d}\beta < \infty \tag{3.9}$$

*for some* $t_0^* > 0$.

**Remark 3.4.** As Shao and Chen (1997) showed, $(C2)$ implies that the design matrix $\left(x_{01}, x_{02}, \ldots, x_{0n_0}\right)'$ has full rank.

**Remark 3.5.** It can be shown that the condition, $\int_{-\infty}^{\infty} \mathrm{e}^{t_0|u|}\, \mathrm{d}F(u) < \infty$, holds for the logit and log–log link models while the condition, $\int_{-\infty}^{\infty} \mathrm{e}^{t_0 u^2}\, \mathrm{d}F(u) < \infty$, holds for the probit models.

Using Lemmas 3.1 and 3.2, we have the following sufficient conditions for the propriety of the power prior distribution when the $y_{0i}$ are binary responses.

**Theorem 3.2.** *If one of the following conditions is satisfied:*

(i) $\delta_0 > k$, $\lambda_0 > 0$, *and the conditions of (c) in Lemma 3.2*,

(ii) $\delta_0 > k/2$, $\lambda_0 > 0$, *and the conditions of (d) in Lemma 3.2*,

*then for the binary responses* $y_{0i}$*'s, we have*

$$\int \int_0^1 \left[L(\beta|D_0)\right]^{a_0} a_0^{\delta_0 - 1}(1 - a_0)^{\lambda_0 - 1}\, \mathrm{d}a_0 \, \mathrm{d}\beta < \infty. \tag{3.10}$$

The proof Theorem 3.2 directly follows from Lemmas 3.1 and 3.2 and the proof of Theorem 3.1.

### *3.3. Lower bounds for the normalizing constants*

In this section, we derive lower bounds for the normalizing constant of the conditional power prior density for $\beta$ given in (2.6), and then use these lower bounds to investigate whether the sufficient condition on $\delta_0$ is also necessary.

The normalizing constant for (2.6) is given by $\int \left[L(\beta|D_0)\right]^{a_0} \mathrm{d}\beta$. The following theorem gives the lower bound of the normalizing constant for binary response models.

**Theorem 3.3.** *Assume that there exist $c_0 > 0$, $c_1 > 0$ such that for any $u \geqslant 0$, one of the following conditions is satisfied:*
(I) $F(-u) \geqslant c_0 \mathrm{e}^{-c_1 u}$ *and* $1 - F(u) \geqslant c_0 \mathrm{e}^{-c_1 u}$;
(II) $F(-u) \geqslant c_0 \mathrm{e}^{-c_1 u^2}$ *and* $1 - F(u) \geqslant c_0 \mathrm{e}^{-c_1 u^2}$.
*Then for binary responses $y_{0i}$'s, there exists $A(X_0) > 0$ such that for every $0 < a_0 < 1$*

$$\int \left[L(\beta|D_0)\right]^{a_0} \mathrm{d}\beta \geqslant \begin{cases} A(X_0)a_0^{-k} & \textit{if (I) is satisfied,} \\ A(X_0)a_0^{-k/2} & \textit{if (II) is satisfied.} \end{cases} \tag{3.11}$$

The proof of this theorem is given in the appendix.

**Remark 3.6.** From Theorem 3.3, it is easy to see that a necessary condition on $\delta_0$ for the propriety of the power prior distribution is either $\delta_0 > k$ or $\delta_0 > k/2$. This result is useful in determining the choices of hyperparameters $(\delta_0, \lambda_0)$ in the power prior distribution. However, condition (I) in Theorem 3.3 is satisfied for the logit model, while condition (II) in Theorem 3.3 is satisfied for the probit model. Although neither (I) nor (II) in Theorem 3.3 is satisfied for the complementary log–log model, a similar result can be obtained for this model, which is given in the next theorem.

**Theorem 3.4.** *Let*

$$L(\beta|D_0) = \prod_{i=1}^{n_0} \left[F(x_{0i}'\beta)\right]^{y_{0i}} \left[1 - F(x_{0i}'\beta)\right]^{1-y_{0i}},$$

*where $F(s) = \exp[-\exp(-s)]$ is the log–log link function. Then, for the log–log link model including an intercept, i.e., $x_{0i1} = 1$ for $i = 1, 2, \ldots, n_0$, we have*

$$\int \left[L(\beta|D_0)\right]^{a_0} \mathrm{d}\beta \geqslant A(X_0)a_0^{-k}. \tag{3.12}$$

A similar result is also true for the Poisson regression model.

**Theorem 3.5.** *For the Poisson regression model including an intercept, we have*

$$\int_0^{\infty} \left[L(\beta|D_0)\right]^{a_0} \mathrm{d}\beta \geqslant A(X_0, Y_0)a_0^{-k}, \tag{3.13}$$

*where*

$$L(\beta|D_0) = \prod_{i=1}^{n_0} \exp(y_{0i}x_{0i}'\beta - \log(y_{0i}!) - \mathrm{e}^{x_{0i}'\beta}).$$

The proofs of Theorems 3.4 and 3.5 are given in the appendix.

**Remark 3.7.** From Theorems 3.4 and 3.5, for the log–log link model and the Poisson regression model, a necessary condition on $\delta_0$ for the propriety of the power prior distribution is $\delta_0 > k$.

### 3.4. Multiple historical data sets

Multiple historical data sets are often available in clinical trials settings, carcinogenicity studies, and studies in which data are collected over time, such as meteorological data. The priors developed in Section 2 can be easily extended to more than one historical study. If there are $N$ historical studies, we define $D_{0j} = (n_{0j}, X_{0j}, y_{0j})$ to be the historical data based on the $j$th study, $j=1,\ldots,N$, and $D_0=(D_{01},\ldots,D_{0N})$. In this case, it may be desirable to define a weight parameter $a_{0j}$ for each historical study, and take the $a_{0j}$'s to be i.i.d. beta random variables with hyperparameters $(\delta_0, \lambda_0)$, $j=1,\ldots,N$. Letting $a_0 = (a_{01},\ldots,a_{0N})$, the prior in (2.7) can be generalized as

$$\pi(\beta, a_0|D_0) \propto \prod_{j=1}^{N} \left[L(\beta|D_{0j})\right]^{a_{0j}} \, a_{0j}^{\delta_0-1} \, (1-a_{0j})^{\lambda_0-1}. \tag{3.14}$$

Defining a different weight parameter $a_{0j}$ for each historical study provides a great deal of flexibility in the prior elicitation. Such a framework is quite useful in situations where the sample sizes for the various historical studies are quite different, or in situations in which there is much heterogeneity between historical studies. Such heterogeneity, for example, may arise because the various historical studies are based on slightly different populations. Thus, by allowing for a different $a_{0j}$ for each study, we are able to accommodate and model potential heterogeneity between studies. The prior in (3.14) is attractive, since one only needs to specify the two hyperparameters $(\delta_0, \lambda_0)$. Thus, the elicitation is no more difficult than the case of one historical study.

The conditions for the propriety of (3.14) are similar to those of (2.7). An extension of Theorem 3.1 is given as follows.

**Theorem 3.6.** *Assume that conditions* (3.2)–(3.4) *hold for each of the historical data* $D_{0j}$. *Then, the joint prior distribution* $\pi(\beta, a_0|D_0)$ *given in* (3.14) *is proper,* i.e.,

$$\int \left\{ \prod_{j=1}^{N} \int_0^1 \left[L(\beta|D_{0j})\right]^{a_{0j}} \, a_{0j}^{\delta_0-1}(1-a_{0j})^{\lambda_0-1} \, \mathrm{d}a_{0j} \right\} \mathrm{d}\beta < \infty \tag{3.15}$$

*if one of the following conditions is satisfied:*

(i) $\delta_0 > k/N$, $\lambda_0 > 0$, *and* (3.3) *holds;*

(ii) $\delta_0 > k/(2N)$, $\lambda_0 > 0$, *and* (3.4) *holds.*

The proof of the theorem is given in the appendix. Similar extensions of Theorems 3.2–3.5 can also be obtained, and thus a necessary condition on $\delta_0$ for the propriety of the power prior distribution $\pi(\beta, a_0|D_0)$ given by (3.14) is either $\delta_0 > k/N$ or $\delta_0 > k/(2N)$. Note that these sufficient and necessary conditions on $\delta_0$ for multiple historical data sets are weaker than those for the single historical data set, which is expected since more prior information is incorporated into the analysis when multiple historical studies are available.

## 4. An illustrative example: AIDS study

For illustrative purposes, we consider an analysis of the AIDS study ACTG036 using the data from ACTG019 as historical data.

The ACTG019 study was a double blind placebo-controlled clinical trial comparing zidovudine (AZT) to placebo in persons with CD4 counts less than 500. The results of this study were published in Volberding (1990). The sample size for this study, excluding cases with missing data, was $n_0 = 823$. The response variable $(y_0)$ for these data is binary with a 1 indicating death, development of AIDS, or AIDS related complex (ARC), and a 0 indicates otherwise. Several covariates were also measured. The ACTG036 study was also a placebo-controlled clinical trial comparing AZT to placebo in patients with hereditary coagulation disorders. The results of this study have been published by Merigan (1991). The sample size in this study, excluding cases with missing data, was $n = 183$. The response variable $(y)$ for these data is binary with a 1 indicating death, development of AIDS, or AIDS-related complex (ARC), and a 0 indicates otherwise. Several covariates were measured for these data. A summary of both data sets can be found in Chen et al. (1999). Therefore, we let $D_0$ denote the data from the ACTG019 study and $D$ denote the data from the ACTG036 study.

Chen et al. (1999) use the priors given by (2.7) and the logistic regression model to carry out variable subset selection, which yields the model containing an intercept, CD4 count (cell count per mm$^3$ of serum), age, and treatment as the best model. For this model, we use the power prior (2.7) to obtain posterior estimates of the regression coefficients for various choices of $(\mu_{a_0}, \sigma_{a_0})$. The results based on the standardized covariates and the logit model are given in Table 1. The values of $(\mu_{a_0}, \sigma_{a_0})$ and the corresponding values of $(\delta_0, \lambda_0)$ are also reported in the table. We used 50,000 Gibbs iterations for all posterior computations and the Monte Carlo method of Chen and Shao (1999) to calculate 95% highest probability density (HPD) intervals for the parameters of interest. From Table 1, we see that as the weight for ACTG019 study increases, the posterior mean of $a_0$ (denoted $E(a_0|D, D_0)$) increases, the posterior standard deviations (SD) for all parameters decrease, and the 95% HPD intervals get narrower. Most noticeably, when $(\delta_0, \lambda_0) = (100, 1)$, none of the HPD intervals for the regression coefficients contain 0. Table 1 also indicates that the HPD intervals are not too sensitive for moderate changes in $(\mu_{a_0}, \sigma_{a_0})$. This is a comforting feature, since it implies that the HPD intervals are fairly robust with respect to the hyperparameters of $a_0$. This same robustness feature is also exhibited in posterior model probability calculations (see Chen et al., 1999). Although the results in Table 1 are very robust to a wide choice of $(\delta_0, \lambda_0)$, more extreme values of $(\delta_0, \lambda_0)$ can lead to results that are different from those of Table 1. We also note that if $a_0$ has an improper prior, then the joint posterior of $(\beta, a_0)$ is also improper.

Table 1
Posterior estimates for AIDS data

| $(\delta_0, \lambda_0)$ | $(\mu_{a_0}, \sigma_{a_0})$ | $E(a_0\|D, D_0)$ | Posterior variable | Posterior mean | 95% HPD | |
|---|---|---|---|---|---|---|
| | | | | | SD | Interval |
| (5,5) | (0.50, 0.151) | 0.02 | Intercept | −4.389 | 0.725 | (−5.836, −3.055) |
| | | | CD4 count | −1.437 | 0.394 | (−2.238, −0.711) |
| | | | Age | 0.135 | 0.221 | (−0.314, 0.556) |
| | | | Treatment | −0.120 | 0.354 | (−0.817, 0.570) |
| (20,20) | (0.50, 0.078) | 0.09 | Intercept | −3.803 | 0.511 | (−4.834, −2.868) |
| | | | CD4 count | −1.129 | 0.300 | (−1.723, −0.559) |
| | | | Age | 0.176 | 0.195 | (−0.214, 0.552) |
| | | | Treatment | −0.223 | 0.300 | (−0.821, 0.364) |
| (30,30) | (0.50, 0.064) | 0.13 | Intercept | −3.621 | 0.436 | (−4.489, −2.809) |
| | | | CD4 count | −1.028 | 0.265 | (−1.551, −0.515) |
| | | | Age | 0.194 | 0.185 | (−0.170, 0.557) |
| | | | Treatment | −0.259 | 0.278 | (−0.805, 0.288) |
| (50,1) | (0.98, 0.019) | 0.26 | Intercept | −3.337 | 0.323 | (−3.978, −2.715) |
| | | | CD4 count | −0.865 | 0.211 | (−1.276, −0.448) |
| | | | Age | 0.233 | 0.160 | (−0.081, 0.548) |
| | | | Treatment | −0.314 | 0.230 | (−0.766, 0.138) |
| (100,1) | (0.99, 0.010) | 0.53 | Intercept | −3.144 | 0.231 | (−3.601, −2.705) |
| | | | CD4 count | −0.746 | 0.161 | (−1.058, −0.429) |
| | | | Age | 0.271 | 0.135 | (0.001, 0.529) |
| | | | Treatment | −0.356 | 0.181 | (−0.717, −0.011) |

## 5. Conclusions

We have proposed a very useful class of informative prior distributions for generalized linear models, called the power prior distributions. These priors provide a practical and coherent way to do Bayesian data analysis when historical data are available. There are many applications of the proposed priors, including model selection, hypothesis testing, carcinogenicity studies, and clinical trials. The priors are especially attractive since few prior parameters have to be elicited. Such a property is quite desirable for example, in variable subset selection. In addition, the proposed priors are proper under some very general conditions, which are straightforward to check. Several extensions of the priors to other models, such as generalized linear mixed models, time series models, and models for survival data are also available. The priors are also computationally feasible and, since the priors resemble likelihoods, they are especially well-behaved in large samples.

## Acknowledgements

This work was partially supported by NSF grants No. DMS-9702172 and No. DMS-9802451, and NIH grants #CA 70101 and #CA 74015. We wish to thank the referees for helpful comments that have improved the manuscript.

## Appendix

**Proof of Lemma 3.1.** When $1/2 \leqslant t \leqslant 1$, (3.1) is obviously true. Therefore, it suffices to consider the case of $0 < t < 1/2$. It is easy to see that for any $0 < t < 1/2$

$$
\begin{aligned}
&\int_0^1 t^{a_0} a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}\,\mathrm{d}a_0 \\
&\quad = \int_0^{1/2} t^{a_0} a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}\,\mathrm{d}a_0 + \int_{1/2}^1 t^{a_0} a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}\,\mathrm{d}a_0 \\
&\quad \leqslant K_2 \int_0^{1/2} \exp(-a_0 \ln(1/t)) a_0^{\delta_0 - 1}\,\mathrm{d}a_0 + Kt^{1/2} \\
&\quad = K_2(\ln(1/t))^{-\delta_0} \int_0^{\ln(1/t)/2} \exp(-a_0) a_0^{\delta_0 - 1}\,\mathrm{d}a_0 + Kt^{1/2} \\
&\quad \leqslant K_2(\ln(1/t))^{-\delta_0} \\
&\quad \leqslant K_2(1+\ln(1/t))^{-\delta_0}.
\end{aligned}
$$

This proves the right-hand side inequality of (3.1). As to the left-hand side inequality, one can see that

$$
\begin{aligned}
\int_0^1 t^{a_0} a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}\,\mathrm{d}a_0 &\geqslant \int_0^{1/2} t^{a_0} a_0^{\delta_0 - 1}(1-a_0)^{\lambda_0 - 1}\,\mathrm{d}a_0 \\
&\geqslant K_1 \int_0^{1/2} \exp(-a_0 \ln(1/t)) a_0^{\delta_0 - 1}\,\mathrm{d}a_0 \\
&= K_1(\ln(1/t))^{-\delta_0} \int_0^{\ln(1/t)/2} \exp(-a_0) a_0^{\delta_0 - 1}\,\mathrm{d}a_0 \\
&\geqslant K_1(\ln(1/t))^{-\delta_0} \int_0^{(\ln 2)/2} \exp(-a_0) a_0^{\delta_0 - 1}\,\mathrm{d}a_0 \\
&\geqslant K_1(\ln(1/t))^{-\delta_0} \\
&\geqslant K_1(1+\ln(1/t))^{-\delta_0},
\end{aligned}
$$

as desired. □

**Proof of Theorem 3.1.** We first show that if (3.3) holds:

$$\int \exp\left(t_0^* ||\beta||\right) L(\beta|D_0)\,\mathrm{d}\beta < \infty \tag{A.1}$$

and if (3.4) holds:

$$\int \exp\left(t_0^* ||\beta||^2\right) L(\beta|D_0)\,\mathrm{d}\beta < \infty, \tag{A.2}$$

where $t_0^* > 0$ and $||\beta|| = \sqrt{\beta_1^2 + \cdots + \beta_k^2}$.

Since $\left(x_{0i_1}, x_{0i_2}, \ldots, x_{0i_{m_0}}\right)'$ has full rank, without loss of generality, we assume that $X_{0k} = \left(x_{0i_1}, x_{0i_2}, \ldots, x_{0i_k}\right)'$ is a $k \times k$ full rank matrix. Then, from (2.5) and (3.2), we have

$$L(\beta|D_0) \leqslant M^* \prod_{j=1}^{k} \exp\left\{\alpha_{i_j}^{-1}(y_{0i_j} h(x_{0i_j}'\beta) - \psi(h(x_{0i_j}'\beta))\right\}, \tag{A.3}$$

where

$$M^* = M^{n_0-k} \times \left(\prod_{i=1}^{n_0} \exp\left\{\phi(y_{0i})\right\}\right).$$

Now we make the transformation $u=(u_1, u_2, \ldots, u_k)' = X_{0k}\beta$. This is a one-to-one linear transformation in $k$ dimensions. Thus, we have

$$||\beta|| \leqslant c_1 \sum_{j=1}^{k} |u_j| ||\beta||^2 \leqslant c_2 \sum_{j=1}^{k} u_j^2, \tag{A.4}$$

where $c_1$ and $c_2$ are two positive constants. It is easy to see that (A.3) and (A.4) lead to

$$\begin{aligned}
&\int \exp\left(t_0^* ||\beta||\right) L(\beta|D_0)\, \mathrm{d}\beta \\
&\quad \leqslant M^* |X_{0k}|^{-1} \prod_{j=1}^{k} \left\{\int_{-\infty}^{\infty} \exp\left(t_0^* c_1 |u_j|\right) \exp\left(\alpha_{i_j}^{-1}(y_{0i_j} h(u_j) - \psi(h(u_j))\right)\, \mathrm{d}u_j\right\}.
\end{aligned} \tag{A.5}$$

Therefore (A.1) directly follows from (A.5) and (3.3). Similarly, we can show (A.2).

To prove (3.5), we use Lemma 3.1, (A.1), and (A.2). Without loss generality, we assume that $L(\beta|D_0) \leqslant 1$. (This is because $L(\beta|D_0)$ is bounded due to (3.2).) When (i) is satisfied, by (A.1) and Lemma 3.1

$$\begin{aligned}
&\int \int_0^1 \left[L(\beta|D_0)\right]^{a_0} a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}\, \mathrm{d}a_0\, \mathrm{d}\beta \\
&\leqslant K_2 \int (1 - \ln L(\beta|D_0))^{-\delta_0}\, \mathrm{d}\beta \\
&\leqslant K_2 \int (1 - \ln L(\beta|D_0))^{-\delta_0} 1_{\{L(\beta|D_0) > \mathrm{e}^{-t_0^*(||\beta||+1)}\}}\, \mathrm{d}\beta \\
&\quad + K_2 \int (1 - \ln L(\beta|D_0))^{-\delta_0} 1_{\{L(\beta|D_0) \leqslant \mathrm{e}^{-t_0^*(||\beta||+1)}\}}\, \mathrm{d}\beta &\text{(A.6)} \\
&\leqslant K_2 \int L(\beta|D_0) \mathrm{e}^{t_0^*(||\beta||+1)}\, \mathrm{d}\beta + K_2 \int (1 + ||\beta||)^{-\delta_0}\, \mathrm{d}\beta \\
&< \infty. &\text{(A.7)}
\end{aligned}$$

Similarly, (A.2) and Lemma 3.1 lead to (3.5) if (ii) is satisfied. □

**Proof of Theorem 3.3.** Note that

$$L(\beta|D_0) \geqslant \prod_{i=1}^{n_0} \left[F(-|x_{0i}'\beta|)\right]^{y_{0i}} \left[1 - F(|x_{0i}'\beta|)\right]^{1-y_{0i}}$$

$$\geqslant \begin{cases} \prod_{i=1}^{n_0} [c_0 \exp(-c_1|x_{0i}'\beta|)] & \text{if (I) is satisfied} \\ \prod_{i=1}^{n} [c_0 \exp(-c_1|x_{0i}'\beta|^2)] & \text{if (II) is satisfied} \end{cases}$$

$$\geqslant \begin{cases} c_0^{n_0} \exp\left(-A(X_0)\sum_{j=1}^{k} |\beta_j|\right) & \text{if (I) is satisfied} \\ c_0^{n_0} \exp\left(-A(X_0)\sum_{j=1}^{k} |\beta_j|^2\right) & \text{if (II) is satisfied} \end{cases}$$

We have

$$\int \left[L(\beta|D_0)\right]^{a_0} \mathrm{d}\beta \geqslant \begin{cases} c_0^{n_0} \int \exp\left(-A(X_0)a_0\sum_{j=1}^{k} |\beta_j|\right) \mathrm{d}\beta & \text{if (I) is satisfied} \\ c_0^{n_0} \int \exp\left(-A(X_0)a_0\sum_{j=1}^{k} |\beta_i|^2\right) \mathrm{d}\beta & \text{if (II) is satisfied} \end{cases}$$

$$\geqslant \begin{cases} a_0^{-k} c_0^{n_0} \int \exp\left(-A(X_0)\sum_{j=1}^{k} |\beta_j|\right) \mathrm{d}\beta & \text{if (I) is satisfied} \\ a_0^{-k/2} c_0^{n_0} \int \exp\left(-A(X_0)\sum_{j=1}^{k} |\beta_j|^2\right) \mathrm{d}\beta & \text{if (II) is satisfied} \end{cases}$$

$$\geqslant \begin{cases} A(X_0)a_0^{-k} & \text{if (I) is satisfied} \\ A(X_0)a_0^{-k/2} & \text{if (II) is satisfied} \end{cases} \quad \square$$

**Proof of Theorem 3.4.** Let $x_{0,\min} = \min\{x_{0ij}, j = 1,2,\dots,k, i = 1,2,\dots,n_0\}$. Consider the transformation $x_{0ij}^* = x_{0ij} - x_{0,\min}$ and let $x_{0i}^* = (x_{0i1}^*, x_{0i1}^*, \dots, x_{0ik}^*)'$. Also let $\beta^* = (\beta_1^*, \beta_2, \dots, \beta_k)$ with $\beta_1^* = \beta_1 + x_{0,\min}\sum_{j=2}^{k} \beta_j$.

Since the model includes an intercept, we have $x_{0i}^* \geqslant 0$ and

$$\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{k} [L(\beta|D_0)]^{a_0} \,\mathrm{d}\beta$$

$$= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{k} \left\{ \prod_{i=1}^{n_0} \left[F((x_{0i}^*)'\beta^*)\right]^{y_{0i}} \left[1 - F((x_{0i}^*)'\beta^*)\right]^{1-y_{0i}} \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant \underbrace{\int_{0}^{\infty} \cdots \int_{0}^{\infty}}_{k} \left\{ \prod_{i=1}^{n_0} \left[F((x_{0i}^*)'\beta^*)\right]^{y_{0i}} \left[1 - F((x_{0i}^*)'\beta^*)\right]^{1-y_{0i}} \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant \underbrace{\int_{0}^{\infty} \cdots \int_{0}^{\infty}}_{k} \left\{ \prod_{i=1}^{n_0} [F(0)]^{y_{0i}} \left[1 - \exp(-\exp(-(x_{0i}^*)'\beta^*))\right] \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant [F(0)]^{n_0} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{k} \left\{ \prod_{i=1}^{n_0} \left[\exp(-(x_{0i}^*)'\beta^*)/2\right] \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant [F(0)/2]^{n_0} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{k} \exp(-a_0 \sum_{i=1}^{n_0} (x_{0i}^*)'\beta^*)\, \mathrm{d}\beta^*$$

$$\geqslant A(X_0) a_0^{-k},$$

as desired. □

**Proof of Theorem 3.5.** Using the same notation as in the proof of Theorem 3.4, we have

$$\underbrace{\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty}_{k} [L(\beta|D_0)]^{a_0}\, \mathrm{d}\beta$$

$$= \underbrace{\int_{-\infty}^\infty \cdots \int_{-\infty}^\infty}_{k} \left\{ \prod_{i=1}^{n_0} \exp\left( y_{0i}(x_{0i}^*)'\beta^* - \log(y_{0i}!) - e^{(x_{0i}^*)'\beta^*} \right) \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant \underbrace{\int_{-\infty}^0 \cdots \int_{-\infty}^0}_{k} \left\{ \prod_{i=1}^{n_0} \exp( y_{0i}(x_{0i}^*)'\beta^* - \log(y_{0i}!) - e^{(x_{0i}^*)'\beta^*} ) \right\}^{a_0} \mathrm{d}\beta^*$$

$$\geqslant \underbrace{\int_{-\infty}^0 \cdots \int_{-\infty}^0}_{k} \left\{ \prod_{i=1}^{n_0} \exp( y_{0i}(x_{0i}^*)'\beta^* - \log(y_{0i}!) - 1) \right\}^{a_0} \mathrm{d}\beta^*$$

$$= \exp(-n_0 - a_0 \sum_{i=1}^{n} \log(y_{0i}!)) \underbrace{\int_{-\infty}^0 \cdots \int_{-\infty}^0}_{k} \exp\left( -a_0 \sum_{i=1}^{n_0} y_{0i}(x_{0i}^*)'\beta^* \right) \mathrm{d}\beta^*$$

$$\geqslant A(X_0, y_0) a_0^{-k}.$$

This proves (3.13). □

**Proof of Theorem 3.6.** Similar to the proof of Theorem 3.1, it can be shown that results (A.1) and (A.2) are true for each $D_{0j}$. Without loss of generality, we assume that $L(\beta|D_{0j}) \leqslant 1$ for $j = 1, 2, \ldots, N$. Then, following the proof of Theorem 3.1 and using Lemma 3.1 and (A.1), we have

$$\int \left\{ \prod_{j=1}^{N} \int_0^1 \left[L(\beta|D_{0j})\right]^{a_{0j}} a_{0j}^{\delta_0 - 1} (1 - a_{0j})^{\lambda_0 - 1}\, \mathrm{d}a_{0j} \right\} \mathrm{d}\beta$$

$$\leqslant \int K_2^N \prod_{j=1}^{N} (1 - \ln L(\beta|D_{0j}))^{-\delta_0}\, \mathrm{d}\beta$$

$$
\begin{aligned}
&= K_2^N \int \prod_{j=1}^{N} (1 - \ln L(\beta|D_{0j}))^{-\delta_0} 1_{\{\max_{1\leqslant j\leqslant N} L(\beta|D_{0j}) > \mathrm{e}^{-t_0^*(||\beta||+1)}\}} \, \mathrm{d}\beta \\
&\quad + K_2^N \int \prod_{j=1}^{N} (1 - \ln L(\beta|D_{0j}))^{-\delta_0} 1_{\{\max_{1\leqslant j\leqslant N} L(\beta|D_{0j}) \leqslant \mathrm{e}^{-t_0^*(||\beta||+1)}\}} \, \mathrm{d}\beta \\
&\leqslant K_2^N \int 1_{\{\max_{1\leqslant j\leqslant N} L(\beta|D_{0j}) > \mathrm{e}^{-t_0^*(||\beta||+1)}\}} \, \mathrm{d}\beta + K_2^N \int \prod_{j=1}^{N} [t_0^*(||\beta|| + 1)]^{-\delta_0} \, \mathrm{d}\beta \\
&\leqslant K_2^N \sum_{1\leqslant j\leqslant N} \int L(\beta|D_{0j}) \mathrm{e}^{t_0^*(||\beta||+1)} \, \mathrm{d}\beta + K_2^N \int (||\beta|| + 1)^{-N\delta_0} \, \mathrm{d}\beta \\
&< \infty
\end{aligned}
$$

for some $t_0^* > 0$ when (i) is satisfied. Similarly, (3.15) is true when (ii) is satisfied. □

## References

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis. 2nd Edition, Springer, New York.

Chen, M.-H., Ibrahim, J.G., Yiannoutsos, C., 1999. Prior elicitation, variable selection and Bayesian computation for logistic regression models. J. Roy. Statist. Soc. Ser. B 61, 223–242.

Chen, M.-H., Shao, Q.-M., 1999. Monte Carlo estimation of bayesian credible and HPD intervals. J. Comput. Graphical Statist. 8, 69–92.

Diaconis, P., Ylvisaker, D., 1979. Conjugate priors for exponential families. Ann. Statist. 7, 269–281.

Ibrahim, J.G., Ryan, L.M., Chen, M.-H., 1998. Use of historical controls to adjust for covariates in trend tests for binary data. J. Am. Statist. Assoc. 93, 1282–1293.

Kass, R.E., Wasserman, L., 1996. The selection of prior distributions by formal rules. J. Am. Statist. Assoc. 91, 1343–1370.

Merigan, T.C., 1991. Placebo-controlled trial to evaluate zidovudine in treatment of human immunodeficiency virus infection in asymptomatic patients with hemophilia. Blood. 900–906.

Morris, C.N., 1983. Natural exponential families with quadratic variance functions: statistical theory. Ann. of Statist. 11, 515–529.

Shao, Q.-M., Chen, M.-H., 1997. Propriety of posterior distribution for dichotomous quantal response models with general link functions. Technical Report, Department of Mathematical Sciences, Worcester Polytechnic Institute.

Volberding, P.A., 1990. Zidovudine in asymptomatic human immunodeficiency virus infection. New England J. Med. 322.