Volume 6, No. 3 (1997), pp. 332–348 Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

NORMAL APPROXIMATION TO THE POSTERIOR DISTRIBUTION FOR GENERALIZED LINEAR MODELS WITH MANY COVARIATES

SUBHASHIS GHOSAL

Statistics and Mathematics Unit, Indian Statistical Institute
203, B. T. Road, Calcutta 700 035, India

We consider generalized linear models and study the asymptotic properties of the posterior distribution where the dimension of the parameter is allowed to grow to infinity with the sample size. Under certain growth restrictions on the dimension, we show that the posterior distribution is consistent and admits a normal approximation. This result can be used to construct procedures with asymptotic Bayesian validity.

Key words: High dimension, Dempster model, exponential family, generalized linear model, normal approximation, posterior consistency, posterior distribution.

AMS 1991 Subject Classification: Primary 62F15, 62J05.

1. Introduction

In many situations that arise in practice, we encounter independent outcomes $\mathbf{x}_1, \dots, \mathbf{x}_n$ of a variable of interest where the data (possibly after a suitable transformation) comes from an m-dimensional (standard) exponential family, i.e., $\mathbf{x}_i$ has a density (with respect to a σ-finite measure on $\mathbb{R}^m$)

$$f(\mathbf{x}_i, \theta_i) = \exp[\mathbf{x}_i^T \theta_i - \psi(\theta_i)], \qquad i = 1, \dots, n, \tag{1.1}$$

where the parameters $\theta_1, \dots, \theta_n$ are determined by a smaller number of parameters $\beta_1, \dots, \beta_p$, $p < n$. Dempster [3] considers the situation where $\mathbf{x}_i$ is observed along with p (real valued) covariates $z_{i1}, \dots, z_{ip}$ and θ_i has the linear expansion

$$\theta_i = z_{i1}\beta_1 + \cdots + z_{ip}\beta_p, \qquad i = 1, \dots, n. \tag{1.2}$$

Haberman [7] terms the above a Dempster model. One constructs estimates of θ_i's through the estimates of β_j's. Thus the estimate of θ_i "borrows strength" from all the observations, a situation similar to small area estimation. Logistic regression model for binomial probability, Poisson regression model and normal regression model with homoscedastic errors are simple examples of Dempster models. Dempster models are, however, special cases of Generalized Linear Models (GLMs) introduced by Nelder and Wedderburn [12] which are defined by the relation

$$\theta_i = g(z_{i1}\beta_1 + \cdots + z_{ip}\beta_p), \quad i = 1, \ldots, n, \tag{1.3}$$

where the link function $g\colon \mathbb{R}^m \to \mathbb{R}^m$ is a one-to-one, continuously differentiable function with a non-zero Jacobian everywhere (i.e., a diffeomorphism). Dempster models are thus GLMs with the canonical link and enjoy the property of having sufficient statistics equal in dimension to $(\beta_1, \ldots, \beta_p)$, see McCullagh and Nelder [11], p. 32. Among the familiar examples satisfying (1.3) are probit regression model and overdispersed models, see McCullagh and Nelder [11] and Fahrmeir and Tutz [4] for description of these models and further examples.

When a Bayesian analysis is intended, one looks at the posterior distribution. It is therefore important to know whether the minimal requirement of posterior consistency holds or not. Further, the posterior distribution is often very complicated and it is desirable to have simpler approximations. The aim of the present paper is to establish consistency and a normal approximation of the posterior distribution in GLMs, where it is also allowed that the dimension $p = p_n$ increases to infinity with the sample size.

When p remains fixed, these results are relatively easy to derive from the general results obtained in Ghosal *et al.* [6] (Proposition 1, Theorem 1 and Example 1). Normal approximation to the posterior distribution, usually called the Bernstein–von Mises theorem, is well known in some simpler situations and there have been a number of contributors in this area. To name a few, we cite Le Cam [10], Bickel and Yahav [2], and Johnson [9]. Even if one is reluctant to use asymptotic approximations when the sample size is not very large, the normal approximation may be used for other purposes such as the approximating density in importance sampling from the posterior, see, e.g., Tanner [13] (p. 117). Finally, since a first order approximation is always free from the particular choice of the prior, it may be thought of as a sort of "non-informative posterior" which may be appealing to an objective Bayesian, or even a frequentist.

From practical considerations, it is desirable to have limit theorems when p is relatively large, i.e., $p \to \infty$ as $n \to \infty$, subject to some growth restrictions. The reason is twofold. Since a data analyst usually uses a relatively more complicated model (i.e., with a higher dimension) if the sample size is larger, on one hand, limit theorems with dimension tending to infinity justify the use of various asymptotic approximations. On the other hand, for a given model, such results give an idea about the sample size required for the safe applications of asymptotic theory. However, the situation where the dimension tends to infinity is technically much more involved and a very careful consideration of the various error terms is necessary. The frequentist version of this problem, viz., consistency and asymptotic normality of the maximum likelihood estimate (MLE) in Dempster models was solved by Haberman [7] as a special case of the treatment of his more general exponential response models. Haberman [7] obtained his results essentially under the growth condition $p^3/n \to 0$ (the condition $p^2/n \to 0$ suffices for consistency). To the best of our knowledge, similar results for GLMs are not available in the literature. In this paper, we show that the posterior distribution in GLMs admits a normal approximation, even if the dimension $p \to \infty$. It is worth mentioning here that the posterior distribution in a GLM is extremely complicated and can only be computed with the help of highly computationally intensive Markov chain Monte-Carlo methods. This is because of the lack of existence of a conjugate prior and the inapplicability of numerical integration in dimension higher than 2. On the other hand, the normal approximation to the posterior distribution of β is easy to compute and also readily yields an approximation to the posterior distribution of a θ_i, which is sometimes of more direct interest. We assume a little more stringent growth restriction $(p^4 \log p)/n \to 0$ on p than that assumed in the frequentist counterpart. Although the growth condition required here is stronger, in our view it is not totally unexpected. The main reason is the presence of a vast tail region which substantially contribute to the posterior probabilities although the likelihood may be very small in that region. Moreover, we prove our results in terms of a strong distance measure, viz., the L^1-distance and consider the entire parameter simultaneously while the frequentist counterpart concerns weak convergence of linear functionals of the MLE. It will, however, be of considerable interest if asymptotic normality of the posterior distribution of linear functionals can be established under weaker growth conditions.

Asymptotic normality of posterior distributions with a growing number of parameters has not been established before in the literature except in the recent work of the author (Ghosal [5]), where similar results have been proved for linear models. The two problems are, however, largely non-overlapping with the normal regression model being the only common example.

In our results, we restrict ourselves to univariate GLMs, i.e., $m = 1$. The general case is notationally much more complicated while the univariate case already includes many models used in practice. However, we believe that the treatment of the multivariate case $m > 1$ is essentially the same. We shall switch to the non-bold scalar notations and write $x_1, \dots, x_n$, $\theta_1, \dots, \theta_n$, $\beta_1, \dots, \beta_p$ etc. Moreover, we set $z_i = (z_{i1}, \dots, z_{ip})^T$, $\theta = (\theta_1, \dots, \theta_n)^T$ and $\beta = (\beta_1, \dots, \beta_p)^T$. We shall also make another simplification. Although we are in a situation involving a triangular array, the extra suffix denoting the stage will often be suppressed.

The organization of the paper is as follows. In Section 2, we state and prove our main result (Theorem 2.1) on the asymptotic normality of the posterior distribution. The proof is somewhat lengthy and for the sake of a better presentation, we split it into several auxiliary lemmas. Proofs of the lemmas are given in Section 3. Consistency and other related questions are answered from simple corollaries to the main theorem and its lemmas. In Section 4, we discuss applications of the results and check the numerical accuracy of the approximations by means of a simulation study.

2. Setup and Main Results

Let $x_1, \dots, x_n$ be independent observations with x_i having a density $f_i(\cdot)$ (with respect to a σ-finite measure ν) defined by

$$f_i(x_i) = f(x_i; \theta_i) = \exp[x_i\theta_i - \psi(\theta_i)], \qquad i = 1, \dots, n, \tag{2.1}$$

where $\theta_i = g(\mathbf{z}_i^T\beta)$ and $g(\cdot)$ is the link function. For a given prior $\pi(\cdot)$, the posterior distribution of β given the observations $x_1, \dots, x_n$ is defined by

$$(2.2)\quad \pi_n(\beta) \propto \pi(\beta)\prod_{i=1}^{n} f(x_i; g(\mathbf{z}_i^T\beta)) = \pi(\beta)\exp\left[\sum_{i=1}^{n}\left\{x_i g(\mathbf{z}_i^T\beta) - \psi(g(\mathbf{z}_i^T\beta))\right\}\right].$$

We fix a (sequence of) parameter point(s) β_0 and will agree to the convention of dropping β_0 in the probability statements. We transform the parameter β to $\mathbf{u} = \mathbf{B}_n^{1/2}(\beta - \beta_0)$, where $\mathbf{B}_n = \sum_{i=1}^{n} \psi''(g(\mathbf{z}_i^T\beta_0))(g'(\mathbf{z}_i^T\beta_0))^2\mathbf{z}_i\mathbf{z}_i^T$ and $\mathbf{B}_n^{1/2}$ is its positive definite square root. Then the likelihood ratio, as a function of $\mathbf{u}$, is given by

$$(2.3)\quad Z_n(\mathbf{u}) = \exp\left[\sum_{i=1}^{n}\left\{x_i\left(g(\mathbf{z}_i^T\beta_0 + \mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u}) - g(\mathbf{z}_i^T\beta_0)\right) - \left[\psi(g(\mathbf{z}_i^T\beta_0 + \mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})) - \psi(g(\mathbf{z}_i^T\beta))\right]\right\}\right], \quad \mathbf{u} \in \mathbf{B}_n^{1/2}(\Theta_n - \beta_0).$$

We set $Z_n(\mathbf{u}) = 0$ if $\mathbf{u} \notin \mathbf{B}_n^{1/2}(\Theta_n - \beta_0)$. The posterior distribution of $\mathbf{u}$ is then given by

$$(2.4)\qquad \pi_n^*(\mathbf{u}) = \frac{\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})Z_n(\mathbf{u})}{\int \pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{w})Z_n(\mathbf{w})d\mathbf{w}}.$$

The following is the main result of this paper.

Theorem 2.1 *Under Conditions* (A0)–(A3) *described below,*

$$(2.5)\qquad \int \left|\pi_n^*(\mathbf{u}) - \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)\right| d\mathbf{u} \to_p 0,$$

where $\Delta_n = \sum_{i=1}^{n}\left(x_i - \psi'(g(\mathbf{z}_i^T\beta_0))\right)g'(\mathbf{z}_i^T\beta_0)\mathbf{B}_n^{-1/2}\mathbf{z}_i$, $\phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ *stands for the density of* $N_p(\boldsymbol{\mu}, \Sigma)$ *and* $\mathbf{I}_p$ *is the identity matrix of order* p.

To prove Theorem 2.1, we shall assume that the following regularity conditions (A0)–(A3) hold.

(A0) The link function $g(\cdot)$ is a one-to-one thrice continuously differentiable function on $\mathbb{R}$. The matrix $\mathbf{A}_n$ defined by the relation $\mathbf{A}_n = \sum_{i=1}^{n} \mathbf{z}_i\mathbf{z}_i^T$ is positive definite.

(A1) As n varies, $\max_{1\le i\le n}|\theta_{0i}|$ remains bounded, where θ_{0i} stand for $\mathbf{z}_i^T\beta_0$, the ith component of the true value of θ_i.

This would seem to be a reasonable hypothesis particularly if the data is cleaned from extreme outliers.

With the above in mind, we can restrict the parameter space according to our convenience. The allowable values of β are assumed to satisfy

$$(2.6)\qquad \max_{1\le i\le n}|\mathbf{z}_i^T\beta| \le K \qquad \text{(say)},$$

i.e., the sequence $\max_{1\le i\le n}|\theta_i|$ is bounded. We denote the set of all β satisfying (2.6) by Θ_n. The actual specification of K is, however, not necessary for calculating the normal approximation. To a frequentist, our results apply if the prior has support in Θ_n. To a Bayesian, (2.6) means simply a belief in (A1).

Set $\eta_n = \max_{1\le i\le n}\|\mathbf{A}_n^{-1/2}\mathbf{z}_i\|$ and $\delta_n = \|\mathbf{A}_n^{-1/2}\|$ where $\mathbf{A}_n^{-1/2}$ (respectively, $\mathbf{A}_n^{1/2}$) is the positive definite square root of $\mathbf{A}_n^{-1}$ (respectively, $\mathbf{A}_n$) and $\|\cdot\|$ stands for the Euclidean norm for vectors and operator norm for matrices.

(A2) The prior density $\pi(\cdot)$ of β is proper and satisfies

$$\pi(\beta_0) > \eta_0^p \qquad \text{for some} \quad \eta_0 > 0 \tag{2.7}$$

and the condition of Lipschitz continuity

$$|\log\pi(\beta) - \log\pi(\beta_0)| \le K_n\|\beta-\beta_0\|, \qquad \|\beta-\beta_0\| \le Cp(\log p)^{1/2}\delta_n. \tag{2.8}$$

The Lipschitz constant $K_n = K_n(C)$ will be required to satisfy some growth restriction to be described in Assumption (A3) below.

The conditions (2.7) and (2.8) described above are satisfied with $K_n = Mp^{1/2}$ in the common situation where the components of β are a priori independently distributed with the jth component β_j following a density $\pi_j(\cdot)$, $j = 1,\dots,p$, and for some $M,\delta,\eta_0 > 0$ and for all $j = 1,\dots,p$, $\pi_j(\beta_{0j}) > \eta_0$ and

$$|\log\pi_j(\beta_j) - \log\pi_j(\beta_{0j})| \le M|\beta_j - \beta_{0j}|, \qquad |\beta_j - \beta_{0j}| \le \delta, \tag{2.9}$$

provided $p(\log p)^{1/2}\delta_n \to 0$.

(A3) The dimension p can grow to infinity subject to the following constraints:

$$K_n\delta_n p(\log p)^{1/2} \to 0 \quad \text{and} \quad p^{3/2}(\log p)^{1/2}\eta_n \to 0, \tag{2.10}$$

where K_n is as defined in (2.8). Further, the design satisfies

$$\operatorname{tr}(\mathbf{A}_n) = \sum_{i=1}^{n}\sum_{j=1}^{p} z_{ij}^2 = O(np). \tag{2.11}$$

To explain condition (A3), we remark that a condition on the smallness of η_n is a uniform asymptotic negligibility condition while smallness of δ_n is a basic requirement on the normalizer for any kind of asymptotics. The factors involving p are redundant in the fixed dimension case, but are crucial for asymptotics in increasing dimension. The last condition on the trace of $\mathbf{A}_n$ is a mild requirement. When $\mathbf{z}_i$'s behave like a random sample from a nonsingular distribution on $\mathbb{R}^p$, and $K_n = O(p^{1/2})$, the condition $(p^4\log p)/n \to 0$ is sufficient to imply (2.10).

We also assume that some power of p grows faster than n, i.e., $\log p$ and $\log n$ are of the same order. If this fails, then the situation is very close to the classical case of fixed dimension. Theorem 2.1 is still valid in this case but a little change in the proof is required. It can be treated using similar (in fact, simpler) arguments, but one has to use a different break-up of central and tail regions in (2.25) below.

For example, the arguments go through if we split into the regions $||\mathbf{u}|| \le n^{1/4}$ and $||\mathbf{u}|| > n^{1/4}$.

Recall our notation $\mathbf{B}_n = \sum_{i=1}^n \psi''\big(g(\mathbf{z}_i^T\beta_0)\big)\big(g'(\mathbf{z}_i^T\beta_0)\big)^2 \mathbf{z}_i\mathbf{z}_i^T$. By the strict convexity of $\psi(\cdot)$, the assumption on the link function $g(\cdot)$ [vide (A0)] and boundedness of $\mathbf{z}_i^T\beta_0$ [vide (A1)], it follows that there are $K_0 > k_0 > 0$ such that

$$k_0\mathbf{A}_n \le \mathbf{B}_n \le K_0\mathbf{A}_n; \tag{2.12}$$

here $\mathbf{A} \le \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is nonnegative definite. If we denote $||\mathbf{B}_n^{-1/2}||$ by δ_n^* and $\max_{1\le i\le n} ||\mathbf{B}_n^{-1/2}\mathbf{z}_i||$ by η_n^*, then it follows from (2.12) that δ_n and δ_n^* are of the same order. For the same reason η_n and η_n^* are also of the same order. Hence

$$K_n\delta_n^* p(\log p)^{1/2} \to 0, \quad p^{3/2}(\log p)^{1/2}\eta_n^* \to 0 \quad \text{and} \quad \operatorname{tr}(\mathbf{B}_n) = O(np). \tag{2.13}$$

Also observe that $E\Delta_n = \mathbf{0}$ and $E(\Delta_n\Delta_n^T) = \mathbf{I}_p$, the identity matrix of order p, where, as in Theorem 2.1, $\Delta_n = \sum_{i=1}^n \big(x_i - \psi'(g(\mathbf{z}_i^T\beta_0))\big)g'(\mathbf{z}_i^T\beta_0)\mathbf{B}_n^{-1/2}\mathbf{z}_i$. As a consequence, $||\Delta_n|| = O_p(p^{1/2})$ by Chebyshev's inequality.

The proof of Theorem 2.1 is a little lengthy. For the sake of a better presentation, we start with a series of technical lemmas; the proofs of the lemmas will be postponed till Section 3. Lemma 2.1 gives the local expansion of the likelihood ratio. Lemma 2.2 will be used to estimate the contribution of the central portion to the L^1-distance in (2.5). Contribution of the tail of the actual posterior density is estimated in Lemma 2.5 using a technique due to Ibragimov and Has'minskii [8] (Lemma I.5.2). Lemmas 2.3 and 2.4 are preparatory lemmas for Lemma 2.5. We will bound the integral of the actual posterior over an intermediate region with the aid of Lemma 2.6. Finally, Lemma 2.7 gives the estimates of the tail of the approximating normal density.

Lemma 2.1. *For any $C > 0$, we have the following:*

(a) *With probability approaching unity, uniformly in $||\mathbf{u}|| \le Cp(\log p)^{1/2}$,*

$$\big|\log Z_n(\mathbf{u}) - (\mathbf{u}^T\Delta_n - \tfrac{1}{2}||\mathbf{u}||^2)\big| \le \lambda_n||\mathbf{u}||^2 \tag{2.14}$$

and

$$\log Z_n(\mathbf{u}) \le \mathbf{u}^T\Delta_n - \tfrac{1}{2}||\mathbf{u}||^2(1 - 2\lambda_n), \tag{2.15}$$

where $\lambda_n = O(p(\log p)^{1/2}\eta_n^*)$.

(b) *With probability approaching unity, uniformly in $||\mathbf{u}|| \le C(p\log p)^{1/2}$,*

$$\big|\log Z_n(\mathbf{u}) - (\mathbf{u}^T\Delta_n - \tfrac{1}{2}||\mathbf{u}||^2)\big| \le \lambda_n^*||\mathbf{u}||^2 \tag{2.16}$$

and

$$\log Z_n(\mathbf{u}) \le \mathbf{u}^T\Delta_n - \frac{1}{2}||\mathbf{u}||^2(1 - 2\lambda_n^*), \tag{2.17}$$

where $\lambda_n^* = O((p\log p)^{1/2}\eta_n^*)$.

Let $\widetilde{Z}_n(\mathbf{u}) = \exp[\mathbf{u}^T\Delta_n - \|\mathbf{u}\|^2/2]$.

Lemma 2.2. *For any $C > 0$, there exists $B' > 0$ such that for all sufficiently large n, with any pre-assigned large probability*

$$(2.18) \qquad \Big(\int \widetilde{Z}_n(\mathbf{u})\,d\mathbf{u}\Big)^{-1} \int_{\|\mathbf{u}\|\le C(p\log p)^{1/2}} |Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})|\,d\mathbf{u} \le B' p \lambda_n^{*}.$$

Lemma 2.3. *There exist $B_0, \varepsilon_1 > 0$ such that*

$$(2.19) \qquad E|Z_n^{1/2}(\mathbf{u}_1) - Z_n^{1/2}(\mathbf{u}_2)|^2 \le B_0\|\mathbf{u}_1 - \mathbf{u}_2\|^2, \qquad \mathbf{u}_1, \mathbf{u}_2 \in \mathbf{B}_n^{1/2}(\Theta_n - \beta_0),$$

and

$$(2.20) \qquad EZ_n^{1/2}(\mathbf{u}) \le \exp[-\varepsilon_1\|\mathbf{u}\|^2], \qquad \mathbf{u} \in \mathbf{B}_n^{1/2}(\Theta_n - \beta_0).$$

Lemma 2.4. *For any $0 < \delta < 1$, we have*

$$(2.21) \qquad P\Big\{\int Z_n(\mathbf{u})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})d\mathbf{u} < \pi(\beta_0)\delta^p/4\Big\} \le 4B_0^{1/2}\delta,$$

where B_0 is the constant obtained in Lemma 2.3.

Lemma 2.5. *For any $m \ge 0$, there are constants $B_1, C > 0$ such that*

$$(2.22) \qquad E\Big(\int_{\|\mathbf{u}\|>Cp(\log p)^{1/2}} \pi_n^*(\mathbf{u})d\mathbf{u}\Big) \le B_1 p^{-m}.$$

Lemma 2.6. *For any $C_2, c > 0$, we can find $B_2, C_1 > 0$ such that with probability approaching one,*

$$(2.23) \qquad \int_{C_1(p\log p)^{1/2}\le\|\mathbf{u}\|\le C_2p(\log p)^{1/2}} Z_n(\mathbf{u})d\mathbf{u} \le B_2\exp[-cp\log p].$$

Lemma 2.7. *For any $c > 0$, there exists $C > 0$ such that with any pre-assigned probability,*

$$(2.24) \qquad \int_{\|\mathbf{u}\|>Cp^{1/2}} \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)d\mathbf{u} \le \exp[-cp].$$

We now prove Theorem 2.1. Now onwards, B will stand for a positive generic constant which need not have the same value in each appearance.

Proof of Theorem 2.1. Let $C > 0$ and set $F = \{\mathbf{u} : \|\mathbf{u}\| \le Cp(\log p)^{1/2}\}$. Then

$$(2.25)\quad \int |\pi_n^*(\mathbf{u}) - \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)|\, d\mathbf{u} \le \int_F \left| \frac{Z_n(\mathbf{u})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})}{\int Z_n(\mathbf{w})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{w})d\mathbf{w}} - \frac{\pi(\beta_0)\widetilde{Z}_n(\mathbf{u})}{\int \pi(\beta_0)\widetilde{Z}_n(\mathbf{w})d\mathbf{w}} \right| d\mathbf{u} + \int_{F^c} \pi_n^*(\mathbf{u})\, d\mathbf{u} + \int_{F^c} \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)\, d\mathbf{u},$$

where $\widetilde{Z}_n(\mathbf{u}) = \exp[\mathbf{u}^T\Delta_n - \|\mathbf{u}\|^2/2]$.

By applications of Lemmas 2.5 and 2.7 respectively, the last two terms can be made as small as desired with probability arbitrarily close to unity by choosing C sufficiently large. Now the first term on the right hand side (RHS) of (2.25) is at most

$$\int_F \left| \frac{Z_n(\mathbf{u})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})}{\int Z_n(\mathbf{w})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{w})d\mathbf{w}} - \frac{Z_n(\mathbf{u})\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})}{\int \widetilde{Z}_n(\mathbf{w})\pi(\beta_0)d\mathbf{w}} \right| d\mathbf{u} + \left(\int \widetilde{Z}_n(\mathbf{w})\pi(\beta_0)d\mathbf{w}\right)^{-1} \int_F \left|Z_n(\mathbf{u})\pi(\beta_0 + \mathbf{A}_n^{-1/2}\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})\pi(\beta_0)\right| d\mathbf{u},$$

which is further dominated by

$$(2.26)\quad \int_{F^c} \pi_n^*(\mathbf{u})\, d\mathbf{u} + \int_{F^c} \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)\, d\mathbf{u} + 3 \sup_F \left| \frac{\pi(\beta_0 + \mathbf{B}_n^{-1/2}\mathbf{u})}{\pi(\beta_0)} - 1 \right| \frac{\int_F Z_n(\mathbf{u})d\mathbf{u}}{\int \widetilde{Z}_n(\mathbf{u})d\mathbf{u}} + 3\left(\int \widetilde{Z}_n(\mathbf{u})\, d\mathbf{u}\right)^{-1} \int_F |Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})|\, d\mathbf{u}.$$

Observe that

$$(2.27)\quad \left(\int \widetilde{Z}_n(\mathbf{u})d\mathbf{u}\right)^{-1} \int_F |Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})|\, d\mathbf{u} \le (2\pi)^{-p/2} \int_{E^c \cap F} Z_n(\mathbf{u})d\mathbf{u} + \int_{E^c} \phi_p(\mathbf{u}; \Delta_n, \mathbf{I}_p)\, d\mathbf{u} + (2\pi)^{-p/2} \int_E |Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})|\, d\mathbf{u},$$

where $E = \{\mathbf{u} : \|\mathbf{u}\| \le C_1(p\log p)^{1/2}\}$ and C_1 is to be chosen shortly.

Lemma 2.5 implies that the first term on the RHS of (2.27) is small while, with the aid of Lemma 2.7, it follows that the second term is small with probability arbitrarily close to one, provided we choose C_1 large enough. Since $p\lambda_n^* \to 0$ (see (2.13)), it follows by Lemma 2.2 that the last term on the RHS of (2.27) also goes to zero in probability.

It follows from (2.8) that for adequately large n,

$$\sup_{\mathbf{u}\in F}\left|\frac{\pi(\beta_0+\mathbf{B}_n^{-1/2}\mathbf{u})}{\pi(\beta_0)}-1\right|\to 0. \tag{2.28}$$

Since the last term on the RHS of (2.27) goes to zero, $\int_F Z_n(\mathbf{u})\,d\mathbf{u}/\int \widetilde{Z}_n(\mathbf{u})d\mathbf{u}$ remains bounded in probability. Hence the expression in (2.26) is small with probability approaching one and the proof is complete. □

The following is a consequence of Lemma 2.5.

Corollary 2.1 *If Conditions* (A0), (A1), *and* (A2) *hold,* $p(\log p)^{1/2}\eta_n \to 0$ *and* $K_n\delta_n \to 0$, *then for any given* $\delta > 0$, *with probability approaching unity, the posterior distribution of* β *concentrates in the* δ*-neighbourhood of* β_0. *Further, the assertion holds almost surely if the observations* y_i*'s of different stages share the same sample space.*

Remark 2.1. Arguments similar to those used in Theorem 2.1 imply the following moment convergence:

$$\int \|\mathbf{u}\|\,|\pi_n^*(\mathbf{u})-\phi_p(\mathbf{u};\Delta_n,\mathbf{I}_p)|d\mathbf{u}\to_p 0, \tag{2.29}$$

provided (A0), (A1), and (A2) hold and (A3) is strengthened to

$$K_n\delta_n p^{3/2}\log p\to 0 \quad \text{and} \quad p^2(\log p)\eta_n\to 0.$$

This, in particular, implies that the posterior mean $\widetilde{\beta}$ of β admits the linearization

$$\mathbf{B}_n^{1/2}(\widetilde{\beta}-\beta_0)=\Delta_n+o_p(1). \tag{2.30}$$

Thus (2.30), (A3) and Lindeberg's central limit theorem together imply that for any unit vector $\mathbf{e}$, $\mathbf{e}^T\mathbf{B}_n^{1/2}(\widetilde{\beta}-\beta_0)\to_d N(0,1)$.

Theorem 2.1, although important for theoretical reasons, cannot itself be used for the actual approximation of the posterior (unless one is simulating) since the approximation involves the unknown value β_0. We now obtain a variation of it by plugging-in a good estimate of β_0.

Theorem 2.2. *Let* $\widehat{\beta}$ *be an estimate of* β *satisfying*

$$\mathbf{B}_n^{1/2}(\widehat{\beta}-\beta_0)=\Delta_n+o_p(1). \tag{2.31}$$

Let $\widehat{\pi}_n(\mathbf{v})$ *stand for the posterior density of* $\mathbf{v}=\widehat{\mathbf{B}}_n^{1/2}(\beta-\widehat{\beta})$, *where*

$$\widehat{\mathbf{B}}_n=\sum_{i=1}^n \psi''(\mathbf{z}_i^T\widehat{\beta})\left(g'(\mathbf{z}_i^T\widehat{\beta})\right)^2\mathbf{z}_i\mathbf{z}_i^T.$$

Then under (A0)–(A3),

$$\int |\widehat{\pi}_n(\mathbf{v}) - \phi_p(\mathbf{v}; \mathbf{0}, \mathbf{I}_p)| d\mathbf{v} \to_p 0. \tag{2.33}$$

If, moreover, $\widehat{\beta}$ *is the MLE, then* (2.31) *holds, provided* $\widehat{\beta}$ *is consistent in the sense that*

$$\max_{1 \le i \le n} |\mathbf{z}_i^T(\widehat{\beta} - \beta_0)| \to_p 0. \tag{2.33}$$

Proof. Since L^1-distance is an invariant of a change of variable, by Theorem 2.1, we obtain

$$\int |\widehat{\pi}_n(\mathbf{v}) - \phi_p(\mathbf{v}; \boldsymbol{\mu}, \Sigma)| d\mathbf{v} \to_p 0, \tag{2.34}$$

where $\boldsymbol{\mu} = \widehat{\mathbf{B}}_n^{1/2}(\mathbf{B}_n^{-1/2}\Delta_n - (\widehat{\beta} - \beta_0))$ and $\Sigma = \widehat{\mathbf{B}}_n^{1/2}\mathbf{B}_n^{-1}\widehat{\mathbf{B}}_n^{1/2}$. Thus we have to show that

$$\int |\phi_p(\mathbf{v}; \boldsymbol{\mu}, \Sigma) - \phi_p(\mathbf{v}; \mathbf{0}, \mathbf{I}_p)| d\mathbf{v} \to_p 0. \tag{2.35}$$

It suffices to show that the Kullback–Leibler distance

$$\int \log\left(\frac{\phi_p(\mathbf{v}; \mathbf{0}, \mathbf{I}_p)}{\phi_p(\mathbf{v}; \boldsymbol{\mu}, \Sigma)}\right) \phi_p(\mathbf{v}; \mathbf{0}, \mathbf{I}_p) d\mathbf{v}$$

is $o_p(1)$. The twice Kullback–Leibler distance between these two normal densities is

$$\operatorname{tr}(\Sigma^{-1} - \mathbf{I}_p) + \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \log\det\Sigma. \tag{2.36}$$

First observe that

$$p \max_{1 \le i \le n} |\mathbf{z}_i^T(\widehat{\beta} - \beta_0)| \to_p 0. \tag{2.37}$$

Indeed, $\|\mathbf{B}_n^{1/2}(\widehat{\beta} - \beta_0)\| = O_p(p^{1/2})$ by (2.31) and the fact that $\|\Delta_n\| = O_p(p^{1/2})$. Hence

$$|\mathbf{z}_i^T(\widehat{\beta} - \beta_0)| = |\mathbf{z}_i^T \mathbf{B}_n^{-1/2} \mathbf{B}_n^{1/2}(\widehat{\beta} - \beta_0)| \le \eta_n^* p^{1/2}$$

which implies (2.37) in view of Condition (A3).

As in (2.12), it follows that $\widehat{\mathbf{B}}_n \ge k_0 \mathbf{A}_n$, and so $\widehat{\mathbf{B}}_n^{-1} \le k_0^{-1}\mathbf{A}_n^{-1}$. Below, let $t_n = \max_{1 \le i \le n} |\mathbf{z}_i^T(\widehat{\beta} - \beta_0)|$. Using the smoothness of $\psi(\cdot)$ and $g(\cdot)$, it follows that for some constant $a > 0$,

$$\begin{aligned} |\operatorname{tr}(\Sigma^{-1} - \mathbf{I}_p)| &= |\operatorname{tr}((\mathbf{B}_n - \widehat{\mathbf{B}}_n)(\widehat{\mathbf{B}}_n)^{-1})| \\ &\le \max_{1 \le i \le n} |\psi''(\mathbf{z}_i^T\widehat{\beta})(g'(\mathbf{z}_i^T\widehat{\beta}))^2 - \psi''(\mathbf{z}_i^T\beta_0)(g'(\mathbf{z}_i^T\beta_0))^2| \sum_{i=1}^n \mathbf{z}_i^T \widehat{\mathbf{B}}_n^{-1}\mathbf{z}_i \\ &\le a t_n \sum_{i=1}^n \mathbf{z}_i^T \mathbf{A}_n^{-1}\mathbf{z}_i = apt_n, \end{aligned}$$

which is $o_p(1)$ by (2.37). The second term in (2.36) equals $\|B_n^{1/2}(\widehat{\beta}-\beta_0)-\Delta_n\|^2 = o_p(1)$ by (2.31). Also $-O(t_n)B_n \le \widehat{B}_n - B_n \le O(t_n)B_n$, so $\det \Sigma = (1+O(t_n))^p = o_p(1)$ by (2.37). This proves the first assertion.

It remains to prove that under (2.33), the MLE satisfies (2.31). To that end, we observe that the MLE $\widehat{\beta}$ of β satisfies $\sum_{i=1}^n (x_i - \psi'(z_i^T\widehat{\beta}))g(z_i^T\widehat{\beta})B_n^{-1/2}z_i = 0$. Set $\varphi(\theta) = \psi(g(\theta))$. Then by Taylor's expansion, it follows that

$$(2.38)\quad \Delta_n = \sum_{i=1}^n \left(x_i g''(z_i^T\beta_0) - \varphi''(z_i^T\beta_0)\right)B_n^{-1/2}z_iz_i^T(\widehat{\beta}-\beta_0)$$
$$+\frac{1}{2}\sum_{i=1}^n \left(x_i g'''(z_i^T\beta^*) - \varphi'''(z_i^T\beta^*)\right)B_n^{-1/2}z_i\left(z_i^T(\widehat{\beta}-\beta_0)\right)^2;$$

here β^* is an intermediate point. The rest of the proof can be completed using arguments similar to those used in the proof of Lemma 2.1 (see Section 3). □

Remark 2.2. Condition (2.33) is a consistency requirement for the MLE $\widehat{\beta}$ which means that if the sample size is large, all the θ_i's can be estimated with a given precision. The condition holds, if $\|B_n^{1/2}(\widehat{\beta}-\beta_0)\|^2 = O_p(p^3\log p)$ [vide Condition (A3)]. For the special case of Dempster models, MLE satisfies (2.33) in view of Haberman's [7] results. The condition (2.31) means that the estimator $\widehat{\beta}$ has the efficient influence function.

3. Proof of the Lemmas

Throughout this section, we set $\varphi(t) = \psi(g(t))$.

Proof of Lemma 2.1. The proofs of (a) and (b) are almost identical and so we prove part (a) only. Let $\|u\| \le Cp(\log p)^{1/2}$ where $C > 0$ is a given constant. Then by Taylor's expansion, it follows that

$$(3.1)\qquad \log Z_n(u) = u^T\Delta_n - \frac{1}{2}\|u\|^2 + R_{1n}(u) + R_{2n}(u),$$

where

$$(3.2)\qquad R_{1n}(u) = \frac{1}{2}\sum_{i=1}^n [x_i - \psi'(g(z_i^T\beta_0))]g''(z_i^T\beta_0)(z_i^TB_n^{-1/2}u)^2,$$

$$(3.3)\qquad R_{2n}(u) = \frac{1}{6}\sum_{i=1}^n [x_i g'''(z_i^T\beta^*) - \varphi'''(z_i^T\beta^*)](z_i^TB_n^{-1/2}u)^3,$$

and β^* is an intermediate point. Thus, using boundedness of the parameter space (see (2.6)),

$$(3.4)\quad |R_{2n}(u)|$$
$$\le Bp(\log p)^{1/2}\eta_n^*\Big[\sup_{|t|\le K}\left(|g'''(t)|+|\varphi'''(t)|\right)\Big]\Big[\inf_{|t|\le K}\left(\psi''(g(t))\,(g'(t))^2\right)\Big]^{-1}$$
$$\times\sum_{i=1}^n (|x_i|+1)\psi''\left(g(z_i^T\beta_0)\right)\left(g'(z_i^T\beta_0)\right)^2(z_i^TB_n^{-1/2}u)^2$$

$$= Bp(\log p)^{1/2}\eta_n^* \sum_{i=1}^{n} Ew_i\psi''\big(g(\mathbf{z}_i^T\beta_0)\big)\big(g'(\mathbf{z}_i^T\beta_0)\big)^2(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2$$

$$+ Bp(\log p)^{1/2}\eta_n^* \sum_{i=1}^{n}(w_i - Ew_i)\psi''\big(g(\mathbf{z}_i^T\beta_0)\big)\big(g'(\mathbf{z}_i^T\beta_0)\big)^2(\mathbf{u}^T\mathbf{B}_n^{-1/2}\mathbf{z}_i)^2;$$

here w_i is a short hand for $(|x_i|+1)$.

Since the parameter space is bounded, the sequence $\max_{1\le i\le n} Ew_i$ is also bounded implying the first term on the RHS of (3.4) is at most $Bp(\log p)^{1/2}\eta_n^*\|\mathbf{u}\|^2$. We claim that given any $\eta > 0$, there exists a constant $B > 0$ such that the probability of the following event is greater than $1-\eta$: For all $\mathbf{u} \in \mathbb{R}^p$,

$$(3.5)\qquad \left|\sum_{i=1}^{n}(w_i - Ew_i)\psi''\big(g(\mathbf{z}_i^T\beta_0)\big)\big(g'(\mathbf{z}_i^T\beta_0)\big)^2(\mathbf{u}^T\mathbf{B}_n^{-1/2}\mathbf{z}_i)^2\right| \le Bp^{1/2}\eta_n^*\|\mathbf{u}\|^2.$$

To prove that (3.4) holds with large probability, it is enough to look at the unit vectors. Let $\mathbf{e}_1,\dots,\mathbf{e}_p$ stand for the standard basis in $\mathbb{R}^p$. From Cauchy–Schwarz inequality it follows that for any unit vector $\mathbf{u}$,

$$(3.6)\qquad \left|\sum_{i=1}^{n}(w_i - Ew_i)\psi''\big(g(\mathbf{z}_i^T\beta_0)\big)\big(g'(\mathbf{z}_i^T\beta_0)\big)^2(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2\right|^2$$
$$\le \sum_{j=1}^{p}\sum_{k=1}^{p}\left(\sum_{i=1}^{n}(w_i - Ew_i)\psi''(g(\mathbf{z}_i^T\beta_0))(g'(\mathbf{z}_i^T\beta_0)\right)^2$$
$$\times (\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{e}_j)(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{e}_k))^2.$$

The proof of the above claim now follows from Chebyshev's inequality and some algebraic manipulations. This proves that with high probability, simultaneously for $\|\mathbf{u}\| \le Cp(\log p)^{1/2}\eta_n^*$, we have $R_{2n}(\mathbf{u}) \le Bp(\log p)^{1/2}\eta_n^*(1+p^{1/2}\eta_n^*)\|\mathbf{u}\|^2$.

From similar arguments, we can show that with high probability, simultaneously for $\|\mathbf{u}\| \le C(p\log p)^{1/2}\eta_n^*$, we have $R_{1n}(\mathbf{u}) \le Bp^{1/2}\eta_n^*\|\mathbf{u}\|^2$. This completes the proof of the first part of (a). The second part is a trivial consequence of the first one. □

Proof of Lemma 2.2. Fix $C > 0$ and consider the set $E = \{\mathbf{u} : \|\mathbf{u}\| \le C(p\log p)^{1/2}\}$. Thus for large n, with probability close to unity, we have simultaneously for all $\mathbf{u} \in E$,

$$|Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})| \le B\lambda_n^*\|\mathbf{u}\|^2\widetilde{Z}_n(\mathbf{u})\exp[\lambda_n^*\|\mathbf{u}\|^2],$$

where λ_n^* is as defined in Lemma 2.1. Thus

$$(3.7)\qquad \int_E |Z_n(\mathbf{u}) - \widetilde{Z}_n(\mathbf{u})|\,d\mathbf{u}$$
$$\le B\lambda_n^*\int_E \|\mathbf{u}\|^2\exp\left[\mathbf{u}^T\Delta_n - (1-2\lambda_n^*)\|\mathbf{u}\|^2/2\right]d\mathbf{u}$$

$$\leq B\lambda_n^*(1-2\lambda_n^*)^{-(1+p/2)}\exp\left[\|\Delta_n\|^2(1-2\lambda_n^*)^{-1}/2\right]$$
$$\times\int\|\mathbf{u}+(1-2\lambda_n^*)^{-1/2}\Delta_n\|^2\exp\left[-\|\mathbf{u}\|^2/2\right]d\mathbf{u}$$
$$\leq 2B\lambda_n^*(1-2\lambda_n^*)^{-(1+p/2)}\exp\left[\|\Delta_n\|^2(1-2\lambda_n^*)^{-1}/2\right]$$
$$\times\left(p+(1-2\lambda_n^*)^{-1}\|\Delta_n\|^2\right)(2\pi)^{p/2}.$$

The result now follows from the facts that $\int\widetilde{Z}_n(\mathbf{u})d\mathbf{u}=\exp[\|\Delta_n\|^2/2](2\pi)^{p/2}$, $\|\Delta_n\|^2=O_p(p)$ and $p\lambda_n^*\to 0$. □

Proof of Lemma 2.3. Since any exponential family satisfies the condition of differentiability in quadratic mean at any θ_0, so as $\theta\to\theta_0$,

$$\text{(3.8)}\quad \int\left|f^{1/2}(x;\theta)-f^{1/2}(x;\theta_0)-(\theta-\theta_0)\frac{\partial}{\partial\theta}f^{1/2}(x;\theta_0)\right|^2\nu(dx)=o\left(|\theta-\theta_0|^2\right).$$

Thus if $\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u}$ is sufficiently small,

$$\text{(3.9)}\quad \int\Big|f^{1/2}(x;g(\mathbf{z}_i^T\beta_0+\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u}))-f^{1/2}(x;g(\mathbf{z}_i^T\beta_0))$$
$$-\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u}\left(\frac{\partial}{\partial\theta}f^{1/2}(x;g(\mathbf{z}_i^T\beta_0))g'(\mathbf{z}_i^T\beta_0)\right)\Big|^2\nu(dx)$$
$$=o\left(\left(g(\mathbf{z}_i^T\beta_0+\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})-g(\mathbf{z}_i^T\beta_0)\right)^2\right)=o\left((\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2\right).$$

Let $H_i(\mathbf{u};\beta^*)$ stand for the Hellinger distance between the densities $f(x_i;\mathbf{z}_i^T\beta^*+\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})$ and $f(x_i;\mathbf{z}_i^T\beta^*)$, i.e.,

$$\text{(3.10)}\quad H_i^2(\mathbf{u};\beta^*)=\int\left|f^{1/2}(x_i;\mathbf{z}_i^T\beta^*+\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})-f^{1/2}(x_i;\mathbf{z}_i^T\beta^*)\right|^2\nu(dx_i).$$

Relation (3.9) yields that there exist constants ε_0 and B_0 such that

$$\text{(3.11)}\quad \frac{\varepsilon_0(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2}{1+(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2}\leq H_i^2(\mathbf{u};\beta^*)\leq B_0(\mathbf{z}_i^T\mathbf{B}_n^{-1/2}\mathbf{u})^2;$$

here we have used convexity of $\psi(\cdot)$ and boundedness of the parameter space of θ_i's to conclude that

$$\inf_{|t|>\varepsilon}\int\left|f^{1/2}(x;\theta+t)-f^{1/2}(x;\theta)\right|^2dx>0,\quad\varepsilon>0.$$

Since by (2.6), $\mathbf{z}_i^T\beta$'s are uniformly bounded, the denominator on the left hand side of (3.11) is bounded above. The rest of the argument is standard, see Ibragimov and Has'minskii [8] (p. 53–54). □

Proof of Lemma 2.4. The proof is almost identical with that of Lemma I.5.1 of Ibragimov and Has'minskii [8] and hence is omitted. □

Proof of Lemma 2.5. Except for the fact that the dimension p cannot be absorbed into the constants, the proof goes along the lines of Lemma I.5.2 of Ibragimov and Has'minskii [8]. A formal proof may be found in Ghosal [6] (Lemma 2.5). □

Proof of Lemma 2.6. Fix $C_2, c > 0$. Note that by (2.15) and the fact that $\|\Delta_n\| = O_p(p^{1/2})$, we have with high probability,

$$
\begin{aligned}
(3.12)\quad & \int_{C_1(p\log p)^{1/2}\le\|\mathbf{u}\|\le C_2 p(\log p)^{1/2}} Z_n(\mathbf{u})\,d\mathbf{u} \\
&\le \int_{C_1(p\log p)^{1/2}\le\|\mathbf{u}\|\le C_2 p(\log p)^{1/2}} \exp\left[\mathbf{u}^T\Delta_n - \|\mathbf{u}\|^2(1-2\lambda_n)/2\right] \\
&\le \int_{C_1(p\log p)^{1/2}/2\le\|\mathbf{u}\|\le C_2 p(\log p)^{1/2}} \exp\left[-\|\mathbf{u}\|^2(1-2\lambda_n)/2\right] \\
&\le \left(2C_2 p(\log p)^{1/2}\right)^p \exp\left[-\frac{1}{16}C_1^2 p\log p\right].
\end{aligned}
$$

The result now follows if we choose C_1 sufficiently large. □

Proof of Lemma 2.7. As $\|\Delta_n\| = O_p(p^{1/2})$, the result is a consequence of the large deviation estimates associated with the chi-square distribution, see Bahadur [1]. □

4. Applications and Illustration by a Simulation Study

The asymptotic normality of the posterior distribution of $\boldsymbol{\beta}$ may be used to test approximately whether a particular component of the parameter is zero. Note that the jth component β_j of $\boldsymbol{\beta}$ is zero if and only if there is no effect of the jth covariate. Need for such testing arises in the context of model selection, testing whether there is a variation between different groups, testing the relative effectiveness of one drug over its competitor (when many other factors are also present) and so on. By Theorem 2.2, the posterior distribution of $\boldsymbol{\beta}$ is approximately $N_p(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{B}}_n^{-1})$, where $\widehat{\mathbf{B}}_n = \sum_{i=1}^n \psi''(\mathbf{z}_i^T\widehat{\boldsymbol{\beta}})\mathbf{z}_i\mathbf{z}_i^T$ and $\widehat{\boldsymbol{\beta}}$ is a good estimate like the MLE. Marginalizing to the jth component, we obtain that the posterior distribution of β_j is approximately $N(\widehat{\beta}_j, \sigma^{jj})$, where $((\sigma^{jj})) = \widehat{\mathbf{B}}_n^{-1}$. Hence the approximate region of Highest Posterior Density having posterior probability content about $(1-\alpha)$ is the interval $(\widehat{\beta}_j - \tau_{\alpha/2}\sqrt{\sigma^{jj}}, \widehat{\beta}_j + \tau_{\alpha/2}\sqrt{\sigma^{jj}})$, where τ_α is the $(1-\alpha)$th quantile of the standard normal variable. Thus the hypothesis $\beta_j = 0$ may be accepted if this interval contains zero and rejected otherwise. The procedure is essentially a frequentist test. Our results give its asymptotic Bayesian justification even when the number of regressors grows to infinity.

Sometimes the inference on the more direct parameters θ_i's are of more importance. The normal approximation to the posterior distribution of $\boldsymbol{\beta}$ readily produces an approximation to the posterior distribution of any particular θ_i, which can be used to make inference on θ_i. In fact, the posterior distribution of $g^{-1}(\theta_i)$ is approximately $N(\mathbf{z}_i^T\widehat{\boldsymbol{\beta}}, \mathbf{z}_i^T\widehat{\mathbf{B}}_n^{-1}\mathbf{z}_i)$. By an application of the delta-method, the posterior distribution of θ_i can also be approximated by $N(g(\mathbf{z}_i^T\widehat{\boldsymbol{\beta}}), (g'(\mathbf{z}_i^T\widehat{\mathbf{B}}_n^{-1}\mathbf{z}_i))^2)$.

Below, we investigate the numerical accuracy of the normal approximation to the posterior distributions by means of a simulation study. Although much real data of considerable importance are available in the literature (see McCullagh and Nelder [11] and Fahrmeir and Tutz [4]), we prefer to work with simulated data to avoid the effect of any possible model misspecification. We consider the Poisson regression model: $x_i \overset{\text{ind}}{\sim} \text{Poisson}(\varphi_i)$, $\theta_i = \log\varphi_i = \sum_{j=1}^{p} z_{ij}\beta_j$, $i = 1, \ldots, n$; here the z_{ij}'s are the covariates and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is unknown. Note that when parametrized in terms θ, the exponential family is standard. We generate the z_{ij}'s from $N(0,1)$ but fix their values once these are obtained. Choosing $\boldsymbol{\beta}_0$ (the true value of $\boldsymbol{\beta}$) as the vector having the first 5 components equal to 1 and the rest equal to 0, we then generate the x_i's from the above model. We take $n = 100$ and compare the exact and approximate posterior for $p = 5$, 10 and 20.

To find out the normal approximation, we first compute the MLE of $\boldsymbol{\beta}$ by Newton–Raphson method starting with the initial estimate which is the least square solution of the system $\log(1 + x_i) = \sum_{j=1}^{p} z_{ij}\beta_j$, $i = 1, \ldots, n$. The motivation for this estimate is as follows: We expect a reasonable estimate of $\boldsymbol{\beta}$ if we replace φ_i by an estimate solely based on x_i, in the exact relation $\log\varphi_i = \sum_{j=1}^{p} z_{ij}\beta_j$. The estimate $(1+x_i)$ is the Bayes estimate of φ_i based on the improper uniform prior on φ_i and data x_i. The obvious estimate x_i cannot be used here since it may vanish.

To compute the exact posterior, we have to rely on Markov chain Monte-Carlo methods because numerical integration is inapplicable in high dimension. Since finding the constants of normalization is the main difficulty, we find that the Metropolis algorithm (see Tanner [13], p. 176) is particularly suitable for our purpose. The transition probability function is taken as the (centered) multivariate normal density with dispersion matrix equal to the inverse of the information matrix at the MLE. Starting with a randomly selected value, we run the chain and collect 100 values of $\boldsymbol{\beta}$ with a lag of 100, after discarding the first 5000 outcomes. The whole procedure is repeated 10 times independently to get 1000 samples from the posterior distribution of $\boldsymbol{\beta}$. Finally, the posterior density is obtained by the kernel method of density estimation.

For the purpose of illustration, we display various posterior characteristics of β_1 and θ_1 (which produced the observation $x_1 = 0$) in Table 1 and 2, respectively. For $p = 5$ and $p = 10$, we plot the exact and approximate posterior densities of β_1 in Figures 1 and 2 respectively. We thus observe that, as expected, the normal approximation to the posterior density of β_1 is pretty accurate for $p = 5$ while the quality of the approximation deteriorates as p increases. A similar phenomenon is observed for the posterior density of θ_1 also, which is not shown here to save space. Nevertheless, even with $p = 20$, the approximation is still moderately accurate.

Acknowledgement. The author wishes to thank the referee for suggestions that improved the presentation. This work was supported by a post doctoral grant from the National Board of Higher Mathematics, Department of Atomic Energy, Bombay, India.

TABLE 1. Posterior characteristics of β_1

dim.	post. mean	post. median	MLE	post. SD	SD of normal approximation
$p = 5$	0.9822	0.9848	0.9817	0.0396	0.0468
$p = 10$	0.9546	0.9555	0.9493	0.0589	0.581
$p = 20$	0.9894	0.9868	1.0188	0.0728	0.0712

TABLE 2. Posterior characteristics of θ_1

dim.	post. mean	post. median	MLE	post. SD	SD of normal approximation
$p = 5$	−1.4143	−1.4091	−1.4142	0.0859	0.0979
$p = 10$	−1.2551	−1.2483	−1.2294	0.1529	0.1595
$p = 20$	−1.1604	−1.1675	−1.2057	0.1836	0.1879

FIGURE 1. Exact and approximate posterior density of β_1 for $p = 5$

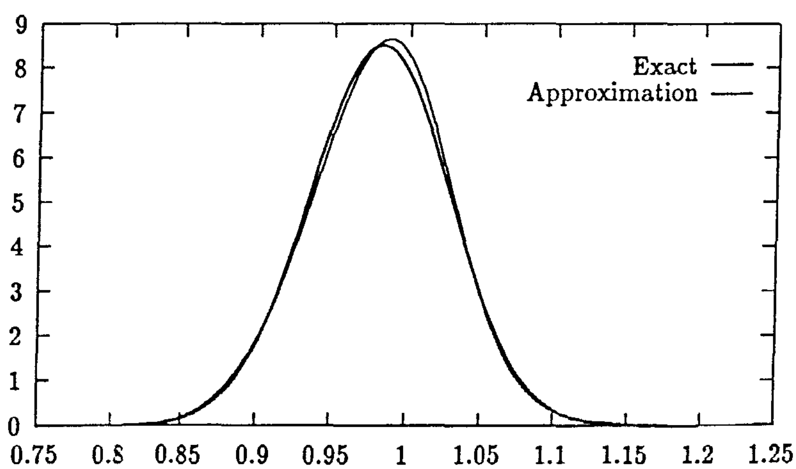
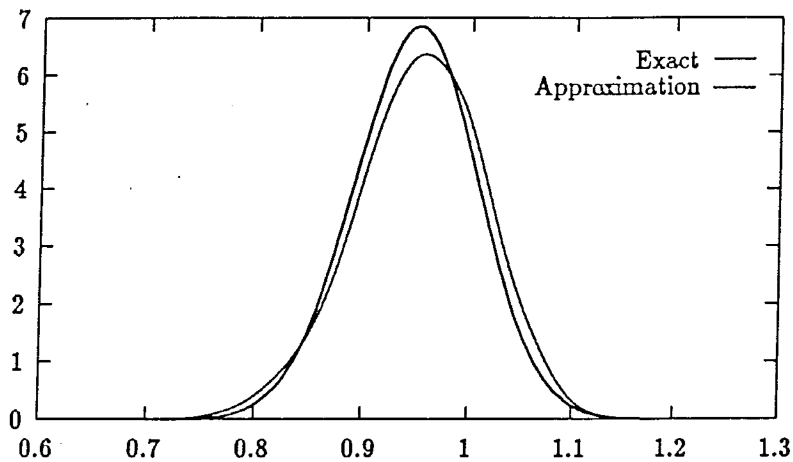


FIGURE 2. Exact and approximate posterior density of β_1 for $p = 10$



References

[1] R. R. Bahadur, *Some Limit Theorems in Statistics*, SIAM, Pennsylvania, 1971.

[2] P. J. Bickel and J. Yahav, *Some contributions to the asymptotic theory of Bayes solutions*, Z. Wahrsch. Verw. Gebiete, 11 (1969), 257–275.

[3] A. P. Dempster, *An overview of multivariate data analysis*, J. Multivariate Anal., 1 (1971), 316–346.

[4] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York, 1994.

[5] S. Ghosal, *Asymptotic normality of posterior distributions in high dimensional linear models*, Technical Report, Indian Statistical Institute, 1995.

[6] S. Ghosal, J. K. Ghosh, and T. Samanta, *On convergence of posterior distributions*, Ann. Statist., 23 (1995), 2145–2152.

[7] S. J. Haberman, *Maximum likelihood estimates in exponential response models*, Ann. Statist., 5 (1977), 815–841.

[8] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer, New York, 1981.

[9] R. A. Johnson, *Asymptotic expansions associated with posterior distribution*, Ann. Math. Statist., 42 (1970), 1241–1253.

[10] L. Le Cam, *On some asymptotic properties of maximum likelihood estimates and related Bayes estimates*, University of California Publications in Statistics, 1 (1953), 277–330.

[11] P. McCullagh, P. A. Nelder, and J. A. Nelder, *Generalized Linear Models*, 2nd ed., Chapman and Hall, London, 1989.

[12] J. A. Nelder and R. W. N. Wedderburn, *Generalized linear models*, J. Roy. Statist. Soc., Ser. A, 35 (1972), 370–384.

[13] M. A. Tanner, *Tools of Statistical Inference: Methods for the Exploration of the Posterior Distribution and Likelihood Function*, 3-rd ed., Springer, New York, 1996.

[*Received November* 1995]