Bernoulli **5**(2), 1999, 315–331

Asymptotic normality of posterior distributions in high-dimensional linear models

SUBHASHIS GHOSAL

Faculty of Mathematics and Computer Science, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands. e-mail: ghosal@cs.vu.nl

We study consistency and asymptotic normality of posterior distributions of the regression coefficient in a linear model when the dimension of the parameter grows with increasing sample size. Under certain growth restrictions on the dimension (depending on the design matrix), we show that the posterior distributions concentrate in neighbourhoods of the true parameter and can be approximated by an appropriate normal distribution.

Keywords: high dimension; linear model; normal approximation; posterior consistency; posterior distribution

1. Introduction

Consider the linear regression model

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are p-dimensional non-stochastic regressors, $\boldsymbol{\beta}$ is an unknown p-vector of the regression coefficients, $y_1 \ldots, y_n$ are observations of a real-valued dependent variable and $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed random errors having a density $f(\cdot)$. To infer about $\boldsymbol{\beta}$, a Bayesian puts a prior on it and looks at the posterior distribution. A primary question related to the Bayesian analysis is posterior consistency, i.e. whether the posterior distributions concentrate near the unknown true value of the parameter. Secondly, the posterior distributions are, in general, very complicated and good approximations are desired. It is therefore of importance to study whether the posterior distribution is consistent and asymptotically normal. The normal approximation has a theoretical importance even if the posterior is actually calculated by some other methods such as the Markov chain Monte Carlo method. For finite-dimensional smooth families, the phenomenon of the normal approximation to the posterior distributions is well known as the Bernstein–von Mises theorem or the Bayesian central limit theorem. Several researchers have contributed in this area including Le Cam (1953), Bickel and Yahav (1969) and Johnson (1970). For a recent work, see Ghosal *et al.* (1995) where a necessary and sufficient condition for posterior convergence is obtained in a general framework. When the dimension p of the parameter $\boldsymbol{\beta}$ in

1350–7265 © 1999 ISI/BS

(1.1) remains fixed as the sample size n increases, results on consistency and approximate normality of the posterior for linear models may be obtained from Theorem 1 of Ghosal *et al.* (1995). However, to the best of the present author's knowledge, the problem of approximating posterior distributions in the regression problem (even for the fixed dimension) has not been addressed explicitly in the literature.

In this paper, we study the behaviour of the posterior distribution as the sample size n tends to infinity where the dimension of the parameter space $p = p_n$ is also allowed to grow to infinity with increasing n. This problem is of significant practical importance since, in data analysis, one often uses a delicate model (i.e. with a large number of parameters) if one has enough data. In other words, one allows the dimension of the parameter to grow with increasing sample size. Moreover, nonparametric models can be approximated by parametric models with increasing dimension as discussed by Shibata (1981) and Diaconis and Freedman (1993). The frequentist version of this problem, namely consistency and asymptotic normality of M estimates, has been studied by Huber (1973), Yohai and Maronna (1979), Ringland (1983) and Portnoy (1984; 1985; 1986). In this paper we show that, under certain growth restrictions on the dimension depending on the design variables, the posterior distributions concentrate in the neighbourhoods of the true value of the parameter and admit a normal approximation. It seems that the present paper is the first attempt to study Bayesian asymptotic properties in models of increasing dimension. We observe that the condition required on the growth rate of the dimension p_n is more stringent than its frequentist counterparts. Although no claim is made about the necessity of this condition on the growth of p_n, we believe that there are at least three reasons to expect some difficulties if p_n grows very rapidly with increasing n. First, there is a long tail area which may substantially contribute to the posterior probabilities although the likelihood is small there. Secondly, our choice of the L^1 metric to measure the distance between densities is quite strong in high dimension. Finally, we approximate the posterior distribution of the entire parameter while the available frequentist results concern asymptotic normality of linear functionals of the frequentist estimate. It may be expected that the conditions required for the validity of the normal approximation of the posterior distribution of a linear functional may be weaker, particularly if the distance between densities is measured in some weaker sense such as the Kolmogorov–Smirnov distance or the weak topology.

In this paper, we assume for simplicity that there is no unknown scale parameter present in the model (1.1). Results can be extended to the case with an unknown scale parameter along a similar line (see Remark 2.2). Extensions to nonlinear models, stochastic regressors and regressors measured with errors are also expected to follow in a similar fashion.

The paper is organized as follows. In Section 2, we formally describe the assumptions and prove the main theorem on the asymptotic normality of the posterior distributions. The consistency of the posterior distributions and some other related facts are easily obtained from the arguments used in the theorem. For clarity of the proof, the main theorem is split into several auxiliary lemmas whose proofs are presented in Section 3.

Our notation will be as follows. We shall use bold letters to denote vectors and matrices. Vectors will be represented in the column form and $\mathbf{x}^{\mathrm{T}}$ will stand for the transpose of a vector $\mathbf{x}$. The capital and lower-case oh notation O, o, O_p and o_p will be used with their usual meanings. The symbol $\rightarrow_p$ will indicate convergence in probability.

2. Main results

2.1. Set-up and assumptions

Consider a triangular array setup of the linear regression model (1.1) where the y_i, $\mathbf{x}_i$, ϵ_i, p and f all can depend on n. Let $\mathbf{X}_n$ stand for the design matrix defined by $\mathbf{X}_n^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Put $\mathbf{A}_n = \mathbf{X}_n^{\mathrm{T}}\mathbf{X}_n$. We assume that $\mathbf{A}_n$ is non-singular. Let $\mathbf{A}_n^{1/2}$ denote the positive definite square root of $\mathbf{A}_n$ and $\mathbf{A}_n^{-1/2} = (\mathbf{A}_n^{1/2})^{-1}$. For vectors, let $\|\cdot\|$ stand for the usual Euclidean norm while, for a matrix $\mathbf{A}$, let $\|\mathbf{A}\|$ stand for the operator norm defined by $\|\mathbf{A}\| = \sup\{\|\mathbf{Ax}\|: \|\mathbf{x}\| \leqslant 1\}$. Set $\eta_n = \max_{1\leqslant i\leqslant n}\|\mathbf{A}_n^{-1/2}\mathbf{x}_i\|$ and $\delta_n = \|\mathbf{A}_n^{-1/2}\|$ (i.e. $\eta_n^2 = \max_{1\leqslant i\leqslant n}(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1}\mathbf{x}_i)$ and $\delta_n^2 = \|\mathbf{A}_n^{-1}\|$). Fix a (sequence of) parameter point(s) $\boldsymbol{\beta}_0$. All the probability statements are made under the true parameter $\boldsymbol{\beta}_0$. As a general convention, we shall not write the subscript n if its presence is solely due to the triangular array situation. Although not reflected in our notation, the model density $f(\cdot)$ is allowed to vary with n for which we implicitly assume the uniform version (in n) of assumption (A1) below.

The conditions are as follows.

(A1) The function $h(z) = \log f(z)$ is thrice differentiable in z and for some $\delta > 0$, $\sup_{|t-z|<\delta}|h'''(t)| \leqslant H(z)$ where $\int H^2(z)f(z)\,\mathrm{d}z < \infty$. The Fisher information $\gamma = \int(h'(z))^2 f(z)\,\mathrm{d}z = -\int h''(z)f(z)\,\mathrm{d}z$ satisfies $0 < \gamma < \infty$. Also, $\int(h''(z))^2 f(z)\,\mathrm{d}z < \infty$.

(A2) The true parameter satisfies

$$\max_{1\leqslant i\leqslant n}|\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0| \leqslant K, \tag{2.1}$$

where K is a constant. Note that if $\int|z|f(z)\,\mathrm{d}z < \infty$, (2.1) is equivalent to saying that $\max_{1\leqslant i\leqslant n}\mathrm{E}|y_i|$ is bounded.

We also assume that all $\boldsymbol{\beta}$ in the parameter space ($\boldsymbol{\Theta}$, say) satisfy

$$\max_{1\leqslant i\leqslant n}|\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}| \leqslant K', \tag{2.2}$$

where K' is a constant not changing with n.

From a frequentist view, (2.2) is a compactness assumption used to ensure that the posterior mass does not escape to infinity. From a Bayesian point of view, (2.2) can be regarded as a belief in (2.1) while proposing the model or the prior. The knowledge about the upper bound K helps us to reduce the parameter space to a suitable compact set. In practice, it may not be very difficult to find an upper bound for K subjectively.

As a referee pointed out, (2.1) is satisfied always (in fact, can take $\boldsymbol{\beta}_0 = \mathbf{0}$) if we redefine the model (1.1), although we then need to know the unknown true value of the parameter. For (2.2) however, a rough conservative upper bound (independent of n) of K (in the original form) suffices as a choice of K'.

(A3) The prior distribution is proper, has a density $\pi(\cdot)$ which satisfies (at $\boldsymbol{\beta}_0$) the positivity requirement

$$\pi(\boldsymbol{\beta}_0) > \eta_0^p \qquad \text{for some } \eta_0 > 0 \tag{2.3}$$

and the Lipschitz condition

$$|\log \pi(\boldsymbol{\beta}) - \log \pi(\boldsymbol{\beta}_0)| \leqslant L_n(C)\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|, \quad \text{whenever } \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leqslant Cp(\log p)^{1/2}\delta_n, \tag{2.4}$$

where the Lipschitz constant $L_n(C)$ is subject to some growth restriction (see assumption (A4) below).

Note that, if the components of $\boldsymbol{\beta}$ are *a priori* independently distributed with the jth component β_j following a density $\pi_j(\cdot)$, $j = 1, \ldots, p$, and for some M, δ, $\eta_0 > 0$ and for all $j = 1, \ldots, p$, $\pi_j(\beta_{0j}) > \eta_0$ and

$$|\log \pi_j(\beta_j) - \log \pi_j(\beta_{0j})| \leqslant M|\beta_j - \beta_{0j}|, \quad \text{whenever } |\beta_j - \beta_{0j}| \leqslant \delta, \tag{2.5}$$

then (2.3) and (2.4) are satisfied with $L_n(C) = Mp^{1/2}$ (for all sufficiently large n) provided that $p(\log p)^{1/2}\delta_n \to 0$.

(A4) The growth of the dimension p is restricted by the constraints

$$\forall C > 0,\ L_n(C)\delta_n p(\log p)^{1/2} \to 0 \quad \text{and} \quad p^{3/2}(\log p)^{1/2}\eta_n \to 0, \tag{2.6}$$

where $L_n(C)$ is as defined in (2.4). Further, the design satisfies

$$\operatorname{tr}(\mathbf{A}_n) = \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2 = O(np). \tag{2.7}$$

Note that a condition on the smallness of η_n is a uniform asymptotic negligibility condition while smallness of δ_n is a natural requirement on the normalizer. The last condition on the trace of $\mathbf{A}_n$ is also used by Portnoy (1984, Condition X4) and is a mild requirement. When the $\mathbf{x}_i$ behave like a random sample from a non-singular distribution on $\mathbb{R}^p$ (Portnoy 1984) and $L_n(C) = O(p^{1/2})$, the growth restrictions stated above are satisfied if $(p^4 \log p)/n \to 0$.

To see an example, consider the p-population problem where each population corresponds to a value of θ in the location family $f(\cdot - \theta)$. Suppose that we have $n_1, \ldots, n_p$ observations respectively from each of these populations and let $n = n_1 + \cdots + n_p$. With the help of p categorical or dummy variables, the model can be expressed in the form (1.1). In this case, it is easy to see that $\delta_n = \eta_n = (\min\{n_1, \ldots, n_p\})^{-1/2}$. Let us think of n as an indexing variable also and suppose that $n_1, \ldots, n_p$ as well as p are allowed to grow with increasing n. Assumption (A2) will be satisfied if the possible values of θ lie in a compact set. A prior satisfying (A3) is easy to specify. For example, the p-fold product of a continuous and positive density (on the compact range of the possible values of θ) satisfies the required conditions. Condition (2.7) is satisfied always as $\operatorname{tr}(\mathbf{A}_n) = n$. If all $n_1, \ldots, n_p$ are of the same order (i.e. of the order of n/p), then (2.6) holds if $(p^4 \log p)/n \to 0$.

In the proof, we shall actually assume that some power of p grows faster than n, i.e. $\log p$ and $\log n$ are of the same order. When this condition fails, the situation is very close to the classical case of fixed dimension and can be treated using reasonings similar to (although simpler than) those presented here; the only difference is that one has to use different central and tail regions (see (2.21) below). For definiteness, a splitting into $\|\mathbf{u}\| \leqslant n^{1/4}$ and $\|\mathbf{u}\| > n^{1/4}$ in the proof of Theorem 2.1 suffices in this case.

For a specified prior $\pi(\cdot)$, the posterior distribution of $\boldsymbol{\beta}$ given the observations $y_1, \ldots, y_n$ is given by

$$\pi_n(\boldsymbol{\beta}) \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} f(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}). \tag{2.8}$$

Normalize $\boldsymbol{\beta}$ as $\mathbf{u} = \mathbf{A}_n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, put

$$Z_n(\mathbf{u}) = \prod_{i=1}^{n} \frac{f(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})}{f(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)}, \quad \mathbf{u} \in \mathbf{A}_n^{1/2}(\boldsymbol{\Theta} - \boldsymbol{\beta}_0), \tag{2.9}$$

and set $Z_n(\mathbf{u}) = 0$ otherwise. The posterior distribution of $\mathbf{u}$ is then given by

$$\pi_n^*(\mathbf{u}) = \frac{\pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u}) Z_n(\mathbf{u})}{\int \pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{w}) Z_n(\mathbf{w})\,\mathrm{d}\mathbf{w}}. \tag{2.10}$$

Also set $\boldsymbol{\Delta}_n = -\sum_{i=1}^{n} h'(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\mathbf{A}_n^{-1/2}\mathbf{x}_i$. Note that $\mathrm{E}\boldsymbol{\Delta}_n = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{\Delta}_n\boldsymbol{\Delta}_n^{\mathrm{T}}) = \gamma\mathbf{A}_n^{-1/2}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}})\mathbf{A}_n^{-1/2} = \gamma\mathbf{I}$, where $\mathbf{I}$ is the identity matrix of order p. In particular, it follows from the Chebyshev inequality that $\|\boldsymbol{\Delta}_n\| = O_p(p^{1/2})$.

2.2. Statements of the results

The main result of this paper is given below.

Theorem 2.1. *Under assumptions* (A1)–(A4),

$$\int |\pi_n^*(\mathbf{u}) - \phi(\mathbf{u}; \gamma^{-1}\boldsymbol{\Delta}_n, \gamma^{-1}\mathbf{I})|\,\mathrm{d}\mathbf{u} \to_p 0, \tag{2.11}$$

where $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{I}$ is the identity matrix of order p.

To prove Theorem 2.1, we state some auxiliary lemmas. The proofs of the lemmas are deferred to Section 3.

The following lemma gives an approximation to the likelihood ratio $Z_n(\mathbf{u})$.

Lemma 2.1. *Let $C > 0$ be any given constant.*

(a) *With probability tending to one, uniformly on $\|\mathbf{u}\| \leqslant Cp(\log p)^{1/2}$,*

$$\left| \log Z_n(\mathbf{u}) - \left(\mathbf{u}^{\mathrm{T}}\boldsymbol{\Delta}_n - \frac{\gamma}{2}\|\mathbf{u}\|^2 \right) \right| \leqslant \lambda_n \|\mathbf{u}\|^2, \tag{2.12}$$

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

(b) *With probability tending to one, uniformly on $\|\mathbf{u}\| \leqslant C(p\log p)^{1/2}$,*

$$\left| \log Z_n(\mathbf{u}) - \left(\mathbf{u}^{\mathrm{T}}\boldsymbol{\Delta}_n - \frac{\gamma}{2}\|\mathbf{u}\|^2 \right) \right| \leqslant \lambda_n^* \|\mathbf{u}\|^2, \tag{2.13}$$

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

The following lemma will be used to bound the central portion of the integral in (2.11). Set $\tilde{Z}_n(\mathbf{u}) = \exp\{\mathbf{u}^{\mathrm{T}}\mathbf{\Delta}_n - (\gamma/2)\|\mathbf{u}\|^2\}$.

Lemma 2.2. *With probability tending to one, for any* $C > 0$*, there exist* $B' > 0$ *such that*

$$\left(\int \tilde{Z}_n(\mathbf{u})\,\mathrm{d}\mathbf{u}\right)^{-1} \int_{\|\mathbf{u}\| \leqslant C(p\log p)^{1/2}} |Z_n(\mathbf{u}) - \tilde{Z}_n(\mathbf{u})|\,\mathrm{d}\mathbf{u} \leqslant B' p \lambda_n^*. \tag{2.14}$$

The most difficult part of a theorem on approximation of the posterior is obtaining a good estimate of the tail integral of the posterior density. We borrow a very powerful technique from Ibragimov and Has'minskii (1981, Lemma I.5.2) for this purpose and suitably adapt in our set-up (see Lemma 2.5 below). It is worthwhile to mention here that the classical technique of Wald (1949) used by Bickel and Yahav (1969), Johnson (1970) and others is not useful in models where the dimension increases. In fact, for linear models, Wald's technique is not helpful even if the dimension is fixed.

To exploit the above-mentioned technique of Ibragimov and Has'minskii (1981), we need good bounds on $\mathrm{E}|Z_n^{1/2}(\mathbf{u}_1) - Z_n^{1/2}(\mathbf{u}_2)|^2$ and $\mathrm{E}Z_n^{1/2}(\mathbf{u})$. These bounds are obtained in the following lemma.

Lemma 2.3. *There exist* B_0, $\epsilon_1 > 0$ *such that*

$$\mathrm{E}|Z_n^{1/2}(\mathbf{u}_1) - Z_n^{1/2}(\mathbf{u}_2)|^2 \leqslant B_0\|\mathbf{u}_1 - \mathbf{u}_2\|^2, \quad \text{if } \mathbf{u}_1, \mathbf{u}_2 \in \mathbf{A}_n^{1/2}(\mathbf{\Theta} - \boldsymbol{\beta}_0), \tag{2.15}$$

and

$$\mathrm{E}Z_n^{1/2}(\mathbf{u}) \leqslant \exp(-\epsilon_1\|\mathbf{u}\|^2), \quad \text{if } \mathbf{u} \in \mathbf{A}_n^{1/2}(\mathbf{\Theta} - \boldsymbol{\beta}_0). \tag{2.16}$$

The next lemma will be used to estimate the posterior probability of the tail region in Lemma 2.5. It asserts that, on a set of large probability, the denominator in (2.10) is not too small.

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

$$P\left(\int Z_n(\mathbf{u})\pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u})\,\mathrm{d}\mathbf{u} < \frac{\pi(\boldsymbol{\beta}_0)\delta^p}{4}\right) \leqslant 4B_0^{1/2}\delta, \tag{2.17}$$

where B_0 *is the constant obtained in Lemma* 2.3.

Lemma 2.5. *For any* $m \geqslant 0$*, there exists* B_1*,* $C > 0$ *such that*

$$\mathrm{E}\left(\int_{\|\mathbf{u}\| > Cp(\log p)^{1/2}} \pi_n^*(\mathbf{u})\,\mathrm{d}\mathbf{u}\right) \leqslant B_1 p^{-m}. \tag{2.18}$$

The following lemma is needed to obtain a bound on the posterior probability of an intermediate region.

Lemma 2.6. *For any C_2, $c > 0$ there exist B_2, $C_1 > 0$ such that, with probability tending to one,*

$$\int_{C_1(p\log p)^{1/2}\leqslant\|\mathbf{u}\|\leqslant C_2\, p(\log p)^{1/2}} Z_n(\mathbf{u})\,\mathrm{d}\mathbf{u} \leqslant B_2\exp(-cp\log p). \tag{2.19}$$

Finally, the next lemma estimates the contribution of the tail of the approximating normal density.

Lemma 2.7. *For any $c > 0$, there exists $C > 0$ such that, with probability tending to one,*

$$\int_{\|\mathbf{u}\|>Cp^{1/2}} \phi(\mathbf{u};\,\gamma^{-1}\boldsymbol{\Delta}_n,\,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u} \leqslant \exp(-cp). \tag{2.20}$$

2.3. Proof of the main theorem

We are now in a position to prove Theorem 2.1. In what follows and in Section 3, B will stand for a generic positive constant.

Proof of Theorem 2.1. Observe that, for any $C > 0$,

$$\begin{aligned}\int|\pi_n^*(\mathbf{u}) - \phi(\mathbf{u};\,\gamma^{-1}\boldsymbol{\Delta}_n,\,\gamma^{-1}\mathbf{I})|\,\mathrm{d}\mathbf{u}\\
\leqslant \int_{\|\mathbf{u}\|\leqslant Cp(\log p)^{1/2}}&\left|\frac{Z_n(\mathbf{u})\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})}{\int Z_n(\mathbf{w})\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{w})\,\mathrm{d}\mathbf{w}} - \frac{\pi(\boldsymbol{\beta}_0)\tilde{Z}_n(\mathbf{u})}{\int\pi(\boldsymbol{\beta}_0)\tilde{Z}_n(\mathbf{w})\,\mathrm{d}\mathbf{w}}\right|\mathrm{d}\mathbf{u}\\
&+\int_{\|\mathbf{u}\|>Cp(\log p)^{1/2}}\pi_n^*(\mathbf{u})\,\mathrm{d}\mathbf{u} + \int_{\|\mathbf{u}\|>Cp(\log p)^{1/2}}\phi(\mathbf{u};\,\gamma^{-1}\boldsymbol{\Delta}_n,\,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u},\end{aligned} \tag{2.21}$$

where, as before, $\tilde{Z}_n(\mathbf{u}) = \exp\{\mathbf{u}^{\mathrm{T}}\boldsymbol{\Delta}_n - (\gamma/2)\|\mathbf{u}\|^2\}$.

Using Lemmas 2.5 and 2.7 respectively, the last two terms can be made as small as we please with probability arbitrarily close to one by choosing C large enough. For this chosen C, let F denote the set $\{\mathbf{u}\colon \|\mathbf{u}\| \leqslant Cp(\log p)^{1/2}\}$. Using Lemma A.1 of Appendix 1, we obtain the following upper bound for the first term on the right-hand side of (2.21):

$$\begin{aligned}\int_{F^c}\pi_n^*(\mathbf{u})\,\mathrm{d}\mathbf{u} + \int_{F^c}\phi(\mathbf{u};\,\gamma^{-1}\boldsymbol{\Delta}_n,\,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u} + 3\left(\int\tilde{Z}_n(\mathbf{u})\pi(\boldsymbol{\beta}_0)\,\mathrm{d}\mathbf{u}\right)^{-1}\\
\times\int_F|Z_n(\mathbf{u})\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u}) - \tilde{Z}_n(\mathbf{u})\pi(\boldsymbol{\beta}_0)|\,\mathrm{d}\mathbf{u}.\end{aligned} \tag{2.22}$$

Also observe that

$$\left(\int \tilde{Z}_n(\mathbf{u})\pi(\boldsymbol{\beta}_0)\,\mathrm{d}\mathbf{u}\right)^{-1}\int_F |Z_n(\mathbf{u})\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})-\tilde{Z}_n(\mathbf{u})\pi(\boldsymbol{\beta}_0)|\,\mathrm{d}\mathbf{u}$$
$$\leqslant \left|\frac{\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})}{\pi(\boldsymbol{\beta}_0)}-1\right|\frac{\int_F Z_n(\mathbf{u})\,\mathrm{d}\mathbf{u}}{\int \tilde{Z}_n(\mathbf{u})\,\mathrm{d}\mathbf{u}}+\left(\int \tilde{Z}_n(\mathbf{u})\,\mathrm{d}\mathbf{u}\right)^{-1}\int_F |Z_n(\mathbf{u})-\tilde{Z}_n(\mathbf{u})|\,\mathrm{d}\mathbf{u}.$$

Now for large n, uniformly on F,

$$\left|\frac{\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})}{\pi(\boldsymbol{\beta}_0)}-1\right| \leqslant 2|\log\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})-\log\pi(\boldsymbol{\beta}_0)| \leqslant 2L_n(C)\delta_n p(\log p)^{1/2}\to 0. \tag{2.23}$$

Also,

$$\left(\int \tilde{Z}_n(\mathbf{u})\,\mathrm{d}\mathbf{u}\right)^{-1}\int_F |Z_n(\mathbf{u})-\tilde{Z}_n(\mathbf{u})|\,\mathrm{d}\mathbf{u} \leqslant (2\pi\gamma^{-1})^{-p/2}\int_{E^c\cap F} Z_n(\mathbf{u})\,\mathrm{d}\mathbf{u} + \int_{E^c}\phi(\mathbf{u};\gamma^{-1}\boldsymbol{\Delta}_n,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u} + (2\pi\gamma^{-1})^{-p/2}\int_E |Z_n(\mathbf{u})-\tilde{Z}_n(\mathbf{u})|\,\mathrm{d}\mathbf{u}, \tag{2.24}$$

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

The first term on the right-hand side of (2.24) is small by Lemma 2.6 while the second term is small by Lemma 2.7 with probability arbitrarily close to one, if we choose C_1 large enough. Since $p\lambda_n^*\to 0$ by assumption (see (2.6)), it follows by Lemma 2.2 that the last term on the right-hand side of (2.24) goes to zero in probability. In particular, the last assertion implies that $(\int_F Z_n(\mathbf{u})\,\mathrm{d}\mathbf{u})/(\int \tilde{Z}_n(\mathbf{u})\,\mathrm{d}\mathbf{u})$ remains bounded in probability. Hence (2.22) is small with probability approaching unity, proving the theorem. □

2.4. Some remarks

As a corollary to Lemma 2.5, we get the following.

Corollary 2.1. *Assume that conditions* (A1)*,* (A2) *and* (A3) *hold and suppose that* $p(\log p)^{1/2}\eta_n\to 0$ *and, for all* $C>0$*,* $L_n(C)\delta_n\to 0$*. Then for any given* $\delta>0$*, the posterior probability of the* δ *neighbourhood of* $\boldsymbol{\beta}_0$ *tends to one in probability.*

Remark 2.1. Using arguments similar to those used in Theorem 2.1, one can prove that

$$\int \|\mathbf{u}\||\pi_n^*(\mathbf{u})-\phi(\mathbf{u};\gamma^{-1}\boldsymbol{\Delta}_n,\gamma^{-1}\mathbf{I})|\,\mathrm{d}\mathbf{u}\to_p 0, \tag{2.25}$$

provided that (A1)–(A3) hold and (A4) is strengthened to the following: for all $C>0$,

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

Let $\tilde{\boldsymbol{\beta}}_n$ denote the posterior mean of $\boldsymbol{\beta}$. Then (2.25) implies that

$$\mathbf{A}_n^{1/2}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \gamma^{-1}\boldsymbol{\Delta}_n + \mathbf{o}_p(1). \tag{2.26}$$

If $\mathbf{e}$ is a unit vector in $\mathbb{R}^p$, it follows from (A4) and the Lindeberg central limit theorem that $\mathbf{e}^{\mathrm{T}}\boldsymbol{\Delta}_n$ is asymptotically $N(0, 1)$. Consequently, $\gamma^{1/2}\mathbf{e}^{\mathrm{T}}\mathbf{A}_n^{1/2}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is asymptotically $N(0, 1)$.

Remark 2.2. Theorem 2.1 should be regarded as a theoretical result and is itself not very useful for the actual approximation of the posterior as the approximation involves $\boldsymbol{\Delta}_n$, which in turn involves the unknown value $\boldsymbol{\beta}_0$ of $\boldsymbol{\beta}$. Thus it will be more useful to obtain a version of Theorem 2.1 which replaces the unknown value of the parameter by its estimate. To this end, observe that the maximum-likelihood estimate (MLE) $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$ satisfies

$$\sum_{i=1}^{n} h'(y_i - \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n)\mathbf{A}_n^{-1/2}\mathbf{x}_i = 0. \tag{2.27}$$

Thus if $\max_{1\leqslant i\leqslant n}|\mathbf{x}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)| = o_p(1)$ (which is implied by $\|\mathbf{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\|^2 = O_p(p^3\log p)$), arguments similar to those used in the proof of Lemma 2.1 (see Section 3) imply that

$$\begin{aligned}
\boldsymbol{\Delta}_n &= \sum_{i=1}^{n}\{h'(y_i - \mathbf{x}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}_n) - h'(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\}\mathbf{A}_n^{-1/2}\mathbf{x}_i \\
&= -\sum_{i=1}^{n} h''(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\mathbf{A}_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \\
&\quad + \tfrac{1}{2}\sum_{i=1}^{n} h'''(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n^*)\mathbf{A}_n^{-1/2}\mathbf{x}_i\{\mathbf{x}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\}^2 \quad (\boldsymbol{\beta}_n^* \text{ is an intermediate point}) \\
&= \gamma\mathbf{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + \mathbf{o}_p(1).
\end{aligned} \tag{2.28}$$

Now letting $\mathbf{v} = \mathbf{A}_n^{1/2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_n)$, the posterior distribution $\hat{\pi}_n(\mathbf{v})$ (say) of $\mathbf{v}$ is approximated by $\phi(\mathbf{v}; \gamma^{-1}\boldsymbol{\Delta}_n - \mathbf{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \gamma^{-1}\mathbf{I})$ which is further approximated by $\phi(\mathbf{v}; \mathbf{0}, \gamma^{-1}\mathbf{I})$. Thus asymptotically the posterior distribution of $\gamma^{1/2}\mathbf{A}_n^{1/2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_n)$ is p-dimensional standard normal; here, as before, closeness is measured by the L^1 distance and convergence is interpreted in the sense of convergence in probability. Note that, above, the MLE can be replaced by any estimate $\hat{\boldsymbol{\beta}}_n$ so that $\mathbf{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ has a linearization $\gamma^{-1}\boldsymbol{\Delta}_n$. Note that by Remark 2.1 the posterior mean has the above linearization, but it cannot be used for plugging in for the obvious reason that it cannot be evaluated without knowing the posterior. Unfortunately, neither is it known that the MLE $\hat{\boldsymbol{\beta}}_n$ satisfies $\|\mathbf{A}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)\|^2 = O_p(p^3\log p)$ nor is a frequentist estimator readily available with linearization $\gamma^{-1}\boldsymbol{\Delta}_n$. However, we expect these to be true under reasonable conditions.

Remark 2.3. Usually, in regression models, the error distribution involves an unknown scale parameter σ (say). By arguments similar to those used in the proof of Theorem 2.1, it can be checked that the joint posterior distribution of $\mathbf{A}_n^{1/2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$ and $n^{1/2}(\sigma-\sigma_0)$ can be approximated by the product of the normal distribution in Theorem 2.1 (with γ replaced by $\sigma_0^{-2}\gamma$) and $N(\tau_n, \sigma_0^2\gamma_1^{-1})$ as long as the scale parameter σ remains confined in a compact subset of $(0, \infty)$; here σ_0 stands for the true value of σ, $\tau_n = -n^{-1/2}\gamma_1^{-1}\sum_{i=1}^n\{\epsilon_i h'(\epsilon_i/\sigma_0)+\sigma_0\}$ and $\gamma_1 = \int\{1+zh'(z)\}^2 f(z)\,\mathrm{d}z$. Although the basic idea of the proof is the same as before, the expressions become complicated. For this reason, the proof is presented assuming that the scale parameter is known.

3. Proof of the lemmas

Proof of Lemma 2.1. We prove only part (a); proof of the other part is exactly analogous. Let $C>0$ be a given constant and $\|\mathbf{u}\| \leqslant Cp(\log p)^{1/2}$. By (A1),

$$\begin{aligned}\log Z_n(\mathbf{u}) &= \sum_{i=1}^n\{h(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u}) - h(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\}\\ &= \mathbf{u}^{\mathrm{T}}\boldsymbol{\Delta}_n - \tfrac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\left(\sum_{i=1}^n h''(\epsilon_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)\mathbf{A}_n^{-1/2}\mathbf{u} + R_n(\mathbf{u}),\end{aligned} \tag{3.1}$$

where

$$R_n(\mathbf{u}) = \tfrac{1}{6}\sum_{i=1}^n h'''(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*)(\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{x}_i)^3 \tag{3.2}$$

and $\boldsymbol{\beta}_*$ is an intermediate point. Thus, for n large,

$$|h'''(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_*)| \leqslant H(\epsilon_i), \quad i = 1, \ldots, n. \tag{3.3}$$

Hence with probability tending to one and uniformly on $\|\mathbf{u}\| \leqslant Cp(\log p)^{1/2}$,

$$\begin{aligned}|R_n(\mathbf{u})| &\leqslant \frac{C}{6}p(\log p)^{1/2}\eta_n\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\left(\sum_{i=1}^n H(\epsilon_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right)\mathbf{A}_n^{-1/2}\mathbf{u}\\ &= \frac{C}{6}p(\log p)^{1/2}\eta_n EH(\epsilon_1)\|\mathbf{u}\|^2\\ &\quad + \frac{C}{6}p(\log p)^{1/2}\eta_n\sum_{i=1}^n\{H(\epsilon_i) - EH(\epsilon_1)\}(\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{x}_i)^2.\end{aligned} \tag{3.4}$$

We claim that, given any $\eta>0$, there exists a set E with probability greater than $1-\eta$ and a constant $B_1>0$ such that, on E,

$$\left|\sum_{i=1}^n\{H(\epsilon_i) - EH(\epsilon_1)\}(\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{x}_i)^2\right| \leqslant B_1 p^{1/2}\eta_n\|\mathbf{u}\|^2 \text{ for all } \mathbf{u} \in \mathbb{R}^p. \tag{3.5}$$

To prove (3.5), it suffices to restrict attention to the unit vectors. Let $\mathbf{e}_1, \ldots, \mathbf{e}_p$ stand for the standard basis in $\mathbb{R}^p$ and let u_j denote the jth component of a vector $\mathbf{u}$. Note that, for any unit vector $\mathbf{u}$,

$$
\begin{aligned}
&\left| \sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\}(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2 \right|^2 \\
&\quad \leqslant \left| \sum_{j=1}^{p}\sum_{k=1}^{p} \left(\sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\} u_j u_k (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k) \right) \right|^2 \\
&\quad \leqslant \left(\sum_{j=1}^{p}\sum_{k=1}^{p} u_j^2 u_k^2 \right) \sum_{j=1}^{p}\sum_{k=1}^{p} \left(\sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\} (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k) \right)^2 \\
&\quad = \sum_{j=1}^{p}\sum_{k=1}^{p} \left(\sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\} (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k) \right)^2 .
\end{aligned}
\tag{3.6}
$$

Thus

$$
\begin{aligned}
&P\left\{ \sup\left(\left| \sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\}(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2 \right| : \mathbf{u} \text{ is a unit vector} \right) \geqslant B_1 p^{1/2}\eta_n \right\} \\
&\quad \leqslant P\left\{ \sum_{j=1}^{p}\sum_{k=1}^{p} \left(\sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\} (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k) \right)^2 \geqslant B_1^2 p \eta_n^2 \right\} \\
&\quad \leqslant B_1^{-2} p^{-1} \eta_n^{-2} \sum_{j=1}^{p}\sum_{k=1}^{p} E\left(\sum_{i=1}^{n} \{H(\epsilon_i) - EH(\epsilon_1)\} (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k) \right)^2 \\
&\quad = B_1^{-2} p^{-1} \eta_n^{-2} \sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{i=1}^{n} \operatorname{var}\{H(\epsilon_i)\} (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_j)^2(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{e}_k)^2 \\
&\quad = B_1^{-2} p^{-1} \eta_n^{-2} \operatorname{var}\{H(\epsilon_1)\} \sum_{i=1}^{n} \|\mathbf{A}_n^{-1/2}\mathbf{x}_i\|^4 \\
&\quad \leqslant B_1^{-2} p^{-1} \operatorname{var}\{H(\epsilon_1)\} \operatorname{tr}\left(\sum_{i=1}^{n} \mathbf{A}_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2} \right) \\
&\quad = B_1^{-2} \operatorname{var}\{H(\epsilon_1)\}.
\end{aligned}
\tag{3.7}
$$

Therefore (3.5) follows if we choose B_1 large enough.

Using a similar argument we can show that, given any $\eta > 0$, there exists a set E' with probability greater than $1 - \eta$ and a constant $B_2 > 0$ such that, on E',

$$\left| \sum_{i=1}^{n} \{h''(\epsilon_i) + \gamma\}(\mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{x}_i)^2 \right| \leqslant B_2 p^{1/2} \eta_n \|\mathbf{u}\|^2 \text{ for all } \mathbf{u} \in \mathbb{R}^p. \tag{3.8}$$

Combining (3.1), (3.4), (3.5) and (3.8), we obtain (2.12). □

Proof of Lemma 2.2. Let $E = \{\mathbf{u}\colon \|\mathbf{u}\| \leqslant C(p \log p)^{1/2}\}$. Thus, for large n, with probability close to unity, we have uniformly on $\mathbf{u} \in E$,

$$|Z_n(\mathbf{u}) - \tilde{Z}_n(\mathbf{u})| \leqslant B\lambda_n^* \|\mathbf{u}\|^2 \tilde{Z}_n(\mathbf{u}) \exp[\lambda_n^* \|\mathbf{u}\|^2], \tag{3.9}$$

where λ_n^* is as in Lemma 2.1. Thus

$$\begin{aligned}
\int_E |Z_n(\mathbf{u}) - \tilde{Z}_n(\mathbf{u})| \, \mathrm{d}\mathbf{u} &\leqslant B\lambda_n^* \int_E \|\mathbf{u}\|^2 \exp\left\{ \mathbf{u}^{\mathrm{T}} \boldsymbol{\Delta}_n - \frac{\gamma}{2}\left(1 - \frac{2\lambda_n^*}{\gamma}\right) \|\mathbf{u}\|^2 \right\} \mathrm{d}\mathbf{u} \\
&\leqslant B\lambda_n^* \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-(1+p/2)} \exp\left\{ \frac{\gamma}{2} \|\boldsymbol{\Delta}_n\|^2 \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-1} \right\} \\
&\quad \times \int \left\| \mathbf{u} + \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-1/2} \boldsymbol{\Delta}_n \right\|^2 \exp\left(-\frac{\gamma}{2}\|\mathbf{u}\|^2\right) \mathrm{d}\mathbf{u} \\
&\leqslant B\lambda_n^* \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-(1+p/2)} \exp\left\{ \frac{\gamma}{2} \|\boldsymbol{\Delta}_n\|^2 \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-1} \right\} \\
&\quad \times \left\{ p + \left(1 - \frac{2\lambda_n^*}{\gamma}\right)^{-1} \|\boldsymbol{\Delta}_n\|^2 \right\} (2\pi\gamma^{-1})^{p/2}.
\end{aligned} \tag{3.10}$$

The result now follows since $\int \tilde{Z}_n(\mathbf{u}) \, \mathrm{d}\mathbf{u} = \exp\{(\gamma/2)\|\boldsymbol{\Delta}_n\|^2\} \, (2\pi\gamma^{-1})^{p/2}$, $\|\boldsymbol{\Delta}_n\|^2 = O_p(p)$ and $p\lambda_n^* \to 0$. □

Proof of Lemma 2.3. The arguments that we use are essentially borrowed from Ibragimov and Has'minskii (1981, Chapters I–III). By (A1) and a well-known result of Hájek (see, for example, Ibragimov and Has'minskii (1981, pp. 121–123)), the parametric family $f(\cdot - \theta)$ satisfies the condition of differentiability in quadratic mean at any θ_0, so that

$$\int \left| f^{1/2}(z - \theta) - f^{1/2}(z - \theta_0) + (\theta - \theta_0) \frac{\mathrm{d}}{\mathrm{d}z} f^{1/2}(z - \theta_0) \right|^2 \mathrm{d}z = o(|\theta - \theta_0|^2) \tag{3.11}$$

as $\theta \to \theta_0$. Thus, if $\mathbf{x}_i^{\mathrm{T}} \mathbf{A}_n^{-1/2} \mathbf{u}$ is sufficiently small,

$$\int \left| f^{1/2}(y - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u}) - f^{1/2}(y - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0) + \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u}\left(\frac{\mathrm{d}}{\mathrm{d}y} f^{1/2}(y - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\right)\right|^2 \mathrm{d}y$$
$$= o\{(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2\}. \quad (3.12)$$

Let $H_i(\mathbf{u})$ stand for the Hellinger distance between the densities $f(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})$ and $f(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)$, i.e.

$$H_i^2(\mathbf{u}) = \int |f^{1/2}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u}) - f^{1/2}(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0)|^2 \,\mathrm{d}y_i. \quad (3.13)$$

Equation (3.12) yields that there exist constants ϵ_0 and B_0 such that

$$\frac{\epsilon_0(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2}{1 + (\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2} \leqslant H_i^2(\mathbf{u}) \leqslant B_0(\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u})^2. \quad (3.14)$$

Since, by (2.2), the $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}$ values are uniformly bounded, the denominator on the left-hand side of (3.14) is bounded above. Consequently, for some constant $\epsilon_1 > 0$,

$$\begin{aligned} \mathrm{E}Z_n^{1/2}(\mathbf{u}) &= \prod_{i=1}^n \{1 - \tfrac{1}{2}H_i^2(\mathbf{u})\} \\ &\leqslant \exp\left(-\epsilon_1 \sum_{i=1}^n \mathbf{u}^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbf{A}_n^{-1/2}\mathbf{u}\right) \\ &= \exp(-\epsilon_1\|\mathbf{u}\|^2). \end{aligned} \quad (3.15)$$

Also,

$$\begin{aligned} \mathrm{E}|Z_n^{1/2}(\mathbf{u}_1) - Z_n^{1/2}(\mathbf{u}_2)|^2 &= 2\left(1 - \prod_{i=1}^n \{1 - \tfrac{1}{2}H_i^2(\mathbf{u}_1 - \mathbf{u}_2)\}\right) \\ &\leqslant \sum_{i=1}^n H_i^2(\mathbf{u}_1 - \mathbf{u}_2) \\ &\leqslant B_0\|\mathbf{u}_1 - \mathbf{u}_2\|^2. \end{aligned} \quad (3.16)$$

□

Proof of Lemma 2.4. Note that, by (2.4), for $\|\mathbf{u}\| \leqslant 1$ and sufficiently large n,

$$\pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u}) \geqslant \pi(\boldsymbol{\beta}_0)\exp(-L_n\delta_n) > \frac{\pi(\boldsymbol{\beta}_0)}{2}. \quad (3.17)$$

Now the proof of Lemma I.5.1 of Ibragimov and Has'minskii (1981) goes through verbatim. □

Proof of Lemma 2.5. The proof essentially follows the arguments of Ibragimov and Has'minskii (1981, Lemma I.5.2), but there is an important difference that the dimension p

can no longer be absorbed into the constants and must be expressed explicitly. A formal proof is presented below.

Let $H > 0$ and set $I = \int_\Gamma \pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u}) Z_n(\mathbf{u})\, d\mathbf{u}$, where $\Gamma = \{\mathbf{u}\colon H \leqslant \|\mathbf{u}\| < H+1\}$. Divide $[-(H+1), (H+1)]^p$ into L^p identical cubes, where L is to be chosen later and pick $\mathbf{u}_i \in \boldsymbol{\Delta}_i$, where $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots$ are the cubes which intersect Γ and $\mathbf{A}_n^{1/2}(\boldsymbol{\Theta} - \boldsymbol{\beta}_0)$. Set $S = \sum_i \int_{\boldsymbol{\Delta}_i} Z_n(\mathbf{u}_i) \pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{1/2}\mathbf{u})\, d\mathbf{u}$. Then using Lemma 2.3 and the fact that $\pi(\cdot)$ is proper, for any $0 < b < \frac{1}{2}$,

$$\begin{aligned} P\{S > \tfrac{1}{2}\exp(-b\epsilon_1 H^2)\} &\leqslant 2L^p \exp\left(-\frac{\epsilon_1 H^2}{2}\right)\left(\int \pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u})\, d\mathbf{u}\right)^{1/2} \\ &\leqslant 2L^p \exp\left(-\frac{\epsilon_1 H^2}{2}\right)(\det \mathbf{A}_n)^{1/4}, \end{aligned} \tag{3.18}$$

$$\begin{aligned} \mathrm{E}|S - I| &\leqslant 2B_0^{1/2} p^{1/2} L^{-1} \int \pi(\boldsymbol{\beta}_0 + \mathbf{A}_n^{-1/2}\mathbf{u})\, d\mathbf{u} \\ &\leqslant Bp^{1/2} L^{-1} (\det \mathbf{A}_n)^{1/2}. \end{aligned} \tag{3.19}$$

Choose the integer L so as to satisfy $1 \leqslant L^p \exp(-\epsilon_1 H^2/4) \leqslant 2^p$. Then

$$\begin{aligned} P\{I > \exp(-b\epsilon_1 H^2)\} &\leqslant 2^{p+1} \exp\left(-\frac{\epsilon_1 H^2}{4}\right)(\det \mathbf{A}_n)^{1/4} \\ &\quad + Bp^{1/2}(\det \mathbf{A}_n)^{1/2} \exp\left\{-\left(\frac{1}{4p} - b\right)\epsilon_1 H^2\right\}. \end{aligned} \tag{3.20}$$

Choosing $b = 1/8p$, we obtain

$$P\left\{ I > \exp\left(-\frac{\epsilon_1 H^2}{8p}\right)\right\} \leqslant Bp^{1/2}(\det \mathbf{A}_n)^{1/2} \exp\left(-\frac{\epsilon_1 H^2}{8p}\right). \tag{3.21}$$

Observe that

$$\det \mathbf{A}_n \leqslant \left(\frac{\operatorname{tr} \mathbf{A}_n}{p}\right)^p = \{O(n)\}^p \tag{3.22}$$

by (2.7). Thus, if C is large and $H > Cp(\log p)^{1/2}$, the bound in (3.21) can be reduced to the form $B\exp(-\epsilon H^2/p)$ for some $\epsilon > 0$. By Lemma 2.4 and making ϵ smaller if necessary, we obtain for $0 < \delta < 1$

$$\mathrm{E}\left(\int_{H\leqslant\|\mathbf{u}\|<H+1}\pi_n(\mathbf{u})\,\mathrm{d}\mathbf{u}\right)\leqslant P\left(\int Z_n(\mathbf{u})\pi(\boldsymbol{\beta}_0+\mathbf{A}_n^{-1/2}\mathbf{u})\,\mathrm{d}\mathbf{u}<\frac{\pi(\boldsymbol{\beta}_0)\delta^p}{4}\right)$$
$$+P\left\{I>\exp\left(-\frac{\epsilon H^2}{p}\right)\right\}+4\{\pi(\boldsymbol{\beta}_0)\}^{-1}\delta^{-p}\exp\left(-\frac{\epsilon H^2}{p}\right)$$
$$\leqslant B\delta+B\eta_0^{-p}\delta^{-p}\exp\left(-\frac{\epsilon H^2}{p}\right). \tag{3.23}$$

The choice $\delta=\eta_0^{-1}\exp(-H^2/2p^2)$ will reduce the right-hand side of (3.23) to $B\exp(-\epsilon H^2/2p^2)$. Replacing H by $Cp(\log p)^{1/2}+r$, $r=0, 1, \ldots,$ and adding the corresponding bounds, we obtain (2.18) by choosing C sufficiently large. □

Proof of Lemma 2.6. Fix C_2, $c>0$ and note that, by (2.12), with probability approaching unity,

$$\int_{B(p\log p)^{1/2}\leqslant\|\mathbf{u}\|\leqslant C_2 p(\log p)^{1/2}} Z_n(\mathbf{u})\,\mathrm{d}\mathbf{u}$$
$$\leqslant\int_{B(p\log p)^{1/2}\leqslant\|\mathbf{u}\|\leqslant C_2 p(\log p)^{1/2}}\exp\left\{\mathbf{u}^{\mathrm{T}}\boldsymbol{\Delta}_n-\frac{\gamma}{2}\|\mathbf{u}\|^2\left(1-\frac{2\lambda_n}{\gamma}\right)\right\}$$
$$\leqslant\int_{B(p\log p)^{1/2}\leqslant\|\mathbf{u}\|\leqslant C_2 p(\log p)^{1/2}}\exp\left\{-\frac{\gamma}{2}\|\mathbf{u}\|^2\left(1-\frac{2\lambda_n}{\gamma}\right)\right\}$$
$$\leqslant\{C_2 p(\log p)^{1/2}\}^p\exp\left(-\frac{\gamma}{4}Bp\log p\right)$$
$$\leqslant\exp(-cp\log p), \tag{3.24}$$

by choosing $C_1=B$ sufficiently large; in the above, we have used the fact that $\|\boldsymbol{\Delta}_n\|=O_p(p^{1/2})$. □

Proof of Lemma 2.7. Fix $c>0$ and recall that $\|\boldsymbol{\Delta}_n\|=O_p(p^{1/2})$. Thus, with probability tending to one,

$$\int_{\|\mathbf{u}\|>Bp^{1/2}}\phi(\mathbf{u};\,\gamma^{-1}\boldsymbol{\Delta}_n,\,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u}\leqslant\int_{\|\mathbf{u}\|>Bp^{1/2}}\phi(\mathbf{u};\,\mathbf{0},\,\gamma^{-1}\mathbf{I})\,\mathrm{d}\mathbf{u}$$
$$\leqslant\Pr(\chi_p^2>Bp). \tag{3.25}$$

Thus the result follows from the well-known large deviation inequality for the chi-squared distribution (Bahadur 1971) by choosing $C=B$ sufficiently large. □

Appendix 1

Lemma A.1. *Let f and g be two non-negative integrable functions not identically zero on a measurable space S and let $F \subset S$. Then*

$$\int_F \left| \frac{f}{\int f} - \frac{g}{\int g} \right| \leqslant \frac{\int_{F^c} f}{\int f} + \frac{\int_{F^c} g}{\int g} + 3\left(\int g\right)^{-1} \int_F |f-g|.$$

Proof. The expression on the left-hand side is dominated by the sum of $|(\int f)^{-1} - (\int g)^{-1}| \int_F f$ and $(\int g)^{-1} \int_F |f-g|$. Note that the first expression is also expressible as

$$\left(\int f\right)^{-1} \left(\int g\right)^{-1} \left(\int_F f\right) \left| \int_F f - \int_F g \right|$$

and the last factor is at most $\int_{F^c} f + \int_{F^c} g + \int_F |f-g|$. Noting that $\int_F f \leqslant \int_F g + \int_F |f-g|$ also, the rest follows easily. □

Acknowledgements

The author wishes to express his gratitude to Professor J. K. Ghosh who drew the author's attention to the problem of approximating posterior distributions in models with a large number of parameters. The suggestions of the referees have been very useful for a better presentation of the paper. The research of the author is supported by the National Board of Higher Mathematics, Department of Atomic Energy, Government of India.

References

Bahadur, (1971) *Some Limit Theorems in Statistics*. Pennsylvania: SIAM.

Bickel, P. and Yahav, J. (1969) Some contributions to the asymptotic theory of Bayes solutions. *Z. Wahrscheinlichkeitstheorie Verw. Geb*, **11**, 257–275.

Diaconis, P. and Freedman, D. (1993) Nonparametric binary regression: a Bayesian approach. *Ann. Statist.*, **21**, 2108–2137.

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

Huber, P. (1973) Robust regression: asymptotics, conjectures, and Monte Carlo. *Ann. Statist.*, **1**, 799–821.

Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag.

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

Le Cam, L. (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.*, **1**, 277–330.

Portnoy, S. (1984) Asymptotic behavior of *M*-estimators of p regression parameters when p^2/n is large. I: Consistency. *Ann. Statist.*, **12**, 1298–1309.

Portnoy, S. (1985) Asymptotic behavior of *M*-estimators of p regression parameters when p^2/n is large. II: Normal approximation. *Ann. Statist.*, **13**, 1403–1417.
Portnoy, S. (1986) Asymptotic behavior of the empiric distribution of *M*-estimated residuals from a regression model with many parameters. *Ann. Statist.*, **14**, 1152–1170.
Ringland, J. (1983) Robust multiple comparisons. *J. Amer. Statist. Assoc.*, **78**, 145–151.
Shibata, R. (1981) An optimal selection of regression variables. *Biometrika*, **68**, 45–54.
Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **20**, 595–601.
Yohai, V.J. and Maronna, R.A. (1979) Asymptotic behavior of *M*-estimators for the linear model. *Ann. Statist.*, **7**, 258–268.

Received September 1995 and revised January 1997