1. Introduction and summary

In the context of statistical inference, the asymptotic properties of estimators are of great interest. These properties are often studied to understand the behavior of estimators as the sample size increases. The main concern is to determine how the estimators behave in large samples and to assess their efficiency.

Abstract:

The asymptotic properties of estimators play a crucial role in statistical analysis. Understanding these properties helps in assessing the reliability of estimators and in making informed decisions about their use in practical applications. The focus of this paper is on the study of asymptotic properties of certain estimators, particularly those that arise in the context of exponential families. We explore the conditions under which these estimators are consistent and efficient.

Key words and phrases: Asymptotic properties, exponential families, efficiency.

BEEAND CHARKE, AND DONICHI SUN

ASYMPTOTICS OF THE EXPECTED POSTERIOR
Asymptotic bounds for the expected posterior

Our first task will be giving upper and lower bounds on the ratio of two posterior densities, as a consequence, we will be valid on sets whose probability tends to unity. As a consequence, we have

\[ \frac{\theta(x) d(\omega(x))}{\omega(x) d(\omega(x))} \]

for any fixed \( x \).

**Notation**

In this chapter, we present the asymptotic results for the posterior density and for posterior density ratios. The results are obtained in the framework of the Metropolis-Hastings method. In this chapter, we will discuss the asymptotic convergence of the posterior density to the true density.

The structure of this chapter is as follows. In Section 2.1, we formally define the key results. In Section 2.2, we prove the asymptotic convergence of the posterior density to the true density. In Section 2.3, we derive the asymptotic results for the posterior density ratios. In Section 2.4, we give some numerical illustrations of the theory presented in this chapter.

The approach is to use a Markov chain to approximate the posterior density. In this chapter, we do this by using a Markov chain Monte Carlo (MCMC) method.

When the sample size is allowed to increase, the expected posterior density is over the sample space, whereas, inside the Markov chain, the posterior density is a function of the Markov chain. In this case, the expected posterior density is the posterior density of the Markov chain.

\[ \left( 1 - \frac{\mu}{\gamma} \right)^{\frac{\mu}{\gamma}} \left( 1 + \frac{\mu}{\gamma} \right)^{\frac{\mu}{\gamma}} \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} = \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} \int \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} \]

where \( \theta = \text{another quantity depending on the prior and the first two terms} \) and \( \mu = \text{the parameter of interest} \).

The expected posterior density for the parameter, namely:

\[ \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} = \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} \int \frac{\exp(\theta | \omega) d(\omega)}{\exp(\omega | \theta) d(\omega)} \]

where \( \theta = \text{the parameter of interest} \).
will obtain a lower bound on the expected posterior and a corresponding upper bound that consists of two terms. The first is similar to the lower bound and the other is an error term which is typically negligible. To begin this program we introduce some formalities. We assume that the densities \( p \) are twice continuously differentiable with respect to \( \theta \) for almost every \( x \) and that there is a \( \delta \) so that for each \( j \) and \( k \) from 1 to \( d \) we have that

\[
E \sup_{\|\theta - \theta_0\| \leq \delta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(X_1 \mid \theta) \right|^2
\]

is finite and for each \( i \) we have that

\[
E \left| \frac{\partial}{\partial \theta_i} \log p(X_1 \mid \theta_0) \right|
\]

is finite. Expectations here are with respect to the \( n \)-fold product of \( p(\cdot \mid \theta_0) \), where \( \theta_0 \) is a fixed value for \( \theta \). We write \( p_{\theta_0}(\cdot) = p(\cdot \mid \theta_0) \) for convenience, and sometimes write \( E_{\theta_0} \) for emphasis. We assume the family is soundly parameterised in the sense that convergence of a sequence of parameter values is equivalent to the weak convergence of the distributions they index. Denote the standardized score function by

\[
\ell_n(\theta) = \frac{1}{n} \nabla \log p_{\theta}(X^n),
\]

and the empirical Fisher information by the \( d \times d \) matrix

\[
I^*(\theta) = \left( -\frac{1}{n} \sum_{k=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(X_k \mid \theta) \right)_{i,j=1,\ldots,d}
\]

We use a local argument on neighborhoods of the form

\[
N(\theta_0, \alpha) = \{ \theta \mid (\theta - \theta_0)^t I(\theta_0)(\theta - \theta_0) \leq \alpha^2 \},
\]

where \( \alpha > 0 \) and \( I(\theta) \) is the true Fisher information at \( \theta \), assumed to be strictly positive and finite. For \( \varepsilon, \delta > 0 \) and satisfying \( 2\delta/(1 - \varepsilon) < \alpha \) we define 3 sets. The first controls the behavior of the posterior. It is

\[
A_n(\alpha, \varepsilon, \theta_0) = \left\{ X^n \mid W(N(\theta_0, \alpha) \mid X^n) \geq \frac{1}{1 + \varepsilon} \right\},
\]

where \( W(\cdot \mid X^n) \) is the probability corresponding to the posterior density \( w(\cdot \mid X^n) \). The second controls \( I^*(\theta_0) \) on \( N(\theta_0, \alpha) \). It is

\[
B_n(\alpha, \varepsilon, \theta_0) = \{ X^n \mid (1 - \varepsilon)(\theta - \theta_0)^t I(\theta_0)(\theta - \theta_0) \leq (\theta - \theta_0)^t I^*(\theta_0)(\theta - \theta_0) \leq (1 + \varepsilon)(\theta - \theta_0)^t I(\theta_0)(\theta - \theta_0) \}
\]

for \( \theta, \theta' \in N(\theta_0, \alpha) \).

The third controls the score:

\[
C_n(\delta, \theta_0) = \{ X^n \mid \ell_n(\theta_0)^t I^{-1}(\theta_0) \ell_n(\theta_0) \leq \delta^2 \}.
\]
2.2 Bounds on the density ratio

We use a version of Laplace’s method to obtain bounds on \( m(x^n)/p(x^n | \theta) \). Since we want to examine dependence of the expected posterior on the prior it is not enough to approximate \( w(\theta) \) on \( N(\theta_0, \alpha) \) by \( w(\theta_0) \). Indeed, we use a second order Taylor expansion of the prior density. To control the error, we suppose there are functions \( \nabla^2 \bar{w}(\theta_0, \alpha) \) and \( \nabla^2 w(\theta_0, \alpha) \) so that

\[
\lim_{\alpha \to 0} \nabla^2 \bar{w}(\theta_0, \alpha) = \lim_{\alpha \to 0} \nabla^2 w(\theta_0, \alpha) = \nabla^2 w(\theta_0)
\]

and for \( \theta, \theta' \in N(\theta_0, \alpha) \) we have the bound

\[
(\theta - \theta_0)^t \nabla^2 w(\theta_0, \alpha)(\theta - \theta_0) \leq (\theta - \theta_0)^t \nabla^2 w(\theta')(\theta - \theta_0) \leq (\theta - \theta_0)^t \nabla^2 \bar{w}(\theta_0, \alpha)(\theta - \theta_0).
\]

Then, on \( N(\theta_0, \alpha) \) we have upper and lower bounds on the prior density given by

\[
w(\theta) \leq w(\theta_0) + (\theta - \theta_0)^t \nabla w(\theta_0) + \frac{1}{2} (\theta - \theta_0)^t \nabla^2 \bar{w}(\theta_0, \alpha)(\theta - \theta_0),
\]

and

\[
w(\theta) \geq w(\theta_0) + (\theta - \theta_0)^t \nabla w(\theta_0) + \frac{1}{2} (\theta - \theta_0)^t \nabla^2 w(\theta_0, \alpha)(\theta - \theta_0).
\]

We use (2.11a, b) to bound the density ratio by transforming to a quantity in \( (\theta - u) \), where \( u = \theta_0 + \frac{1}{1 - \varepsilon} I^{-1}(\theta_0) \ell_n'(\theta_0) \). To state our bounds requires that we introduce some extra functions which occur in those bounds. For the sake of exposition we also define some related functions which will be used in the course of the proof. For the upper bound on the density ratio we use five functions:

\[
G_1 = w(\theta_0) + \frac{1}{1 - \varepsilon} \ell_n'(\theta_0)^t I^{-1}(\theta_0) \nabla w(\theta_0) + \frac{1}{2(1 - \varepsilon)^2} \ell_n'(\theta_0)^t I^{-1}(\theta_0) \nabla^2 \bar{w} I^{-1}(\theta_0) \ell_n'(\theta_0),
\]

\[
G_2 = (\theta - u)^t \left( \nabla w(\theta_0) + \frac{1}{1 - \varepsilon} \nabla^2 \bar{w} I^{-1}(\theta_0) \ell_n'(\theta_0) \right),
\]

\[
G_3 = \frac{1}{2} (\theta - u)^t \nabla^2 \bar{w}(\theta - u),
\]

\[
G_4 = \nabla w(\theta_0) + \frac{1}{1 - \varepsilon} \nabla^2 \bar{w} I^{-1}(\theta_0) \ell_n'(\theta_0),
\]

\[
H_n = (1 + \varepsilon) e^{(n/2)(1 - \varepsilon)} \ell_n'(\theta_0)^t I^{-1}(\theta_0) \ell_n'(\theta_0).
\]

Note that the sum \( G_1 + G_2 + G_3 \) is the right hand side of (2.11a). It will be seen that \( G_4 \) and \( H_n \) appear explicitly in the upper bound. For the lower bound we use five analogous functions:

\[
\tilde{G}_1 = w(\theta_0) + \frac{1}{1 + \varepsilon} \ell_n'(\theta_0)^t I^{-1}(\theta_0) \nabla w(\theta_0)
\]
\[
\begin{align*}
\tilde{G}_2 &= (\theta - \mu)^\ell \left( \nabla w(\theta_0) + \frac{1}{1 + \epsilon} \nabla^2 w I^{-1}(\theta_0) \ell'_n(\theta_0) \right), \\
\tilde{G}_3 &= \frac{1}{2} (\theta - \mu)^4 \nabla^2 w(\theta - \mu), \\
\tilde{G}_4 &= \nabla w(\theta_0) + \frac{1}{1 + \epsilon} \nabla^2 w I^{-1}(\theta_0) \ell'_n(\theta_0), \\
\tilde{H}_n &= (1 - \epsilon) \epsilon^{(n+2)(1+\epsilon)} \ell'_n(\theta_0) I^{-1}(\theta_0) \ell'_n(\theta_0).
\end{align*}
\]

Analogously, the sum \( \tilde{G}_1 + \tilde{G}_2 + \tilde{G}_3 \) gives the right hand side of (2.11b).

Now we can state and prove our bounds on the density ratio.

**Proposition 2.1.** (a) Upper bound on \( m_n(X^n)/p(X^n \mid \theta_0) \): Suppose \( \delta/(1 - \epsilon) < \alpha/2 \) and the prior density admits a second order Taylor expansion as in (2.11a). Then on the set \( A_n \cap B_n \cap C_n \) we obtain the upper bound

\[
(2.14a) \quad \frac{m_n(X^n)}{p(X^n \mid \theta_0)} \leq \frac{H_n(2\pi)^{d/2}}{|n(1 - \epsilon)I(\theta_0)|^{1/2}} \left( G_1 + \frac{\text{tr} \nabla^2 w I^{-1}(\theta_0)}{2n(1 - \epsilon)} + C_1(\theta_0) e^{-n(1/16)\alpha^2(1-\epsilon)} \right),
\]

where

\[
C_1(\theta_0) = 2^{d/2} \left( |G_1| + \left| G_1 + \frac{2G_4 I^{-1}(\theta_0) G_4}{n(1 - \epsilon)} \right|^{1/2} + \sqrt{3 \text{tr} \nabla^2 w I^{-1}(\theta_0) \nabla^2 w I^{-1}(\theta_0)} \right) \frac{n(1 - \epsilon)}{n(1 - \epsilon)}.
\]

Here \( \text{tr} \) denotes the trace function of a matrix.

(b) Lower bound on \( m_n(X^n)/p(X^n \mid \theta_0) \): Suppose \( \delta/(1 + \epsilon) < \alpha/2 \) and the prior density admits a second order Taylor expansion as in (2.11b). Then on the set \( B_n \cap C_n \) we have the lower bound

\[
(2.14b) \quad \frac{m_n(X^n)}{p(X^n \mid \theta_0)} \geq \frac{\tilde{H}_n(2\pi)^{d/2}}{|n(1 + \epsilon)I(\theta_0)|^{1/2}} \left( \tilde{G}_1 + \frac{\text{tr} \nabla^2 w I^{-1}(\theta_0)}{2n(1 + \epsilon)} + C_2(\theta_0) e^{-n(1/16)\alpha^2(1+\epsilon)} \right),
\]

where

\[
C_2(\theta_0) = 2^{d/2} \left( |\tilde{G}_1| + \left| \tilde{G}_1 + \frac{2\tilde{G}_4 I^{-1}(\theta_0) \tilde{G}_4}{n(1 + \epsilon)} \right|^{1/2} + \sqrt{3 \text{tr} \nabla^2 w I^{-1}(\theta_0) \nabla^2 w I^{-1}(\theta_0)} \right) \frac{n(1 + \epsilon)}{n(1 + \epsilon)}.
\]

**Remark.** In the proof of Theorem 2.1 in the Appendix A, it is shown that under the assumptions already stated at the beginning of this section the probabilities of the sets \( A_n \), \( B_n \) and \( C_n \) are \( o(1/n) \). Thus, from (2.14a, b) we will be able to obtain a characterization of \( m_n(X^n)/p(X^n \mid \theta_0) \) which is exact enough to approximate the expected posterior asymptotically.

**Proof.** See the Appendix A. \( \Box \)
2.3 Bounds on the posterior

Next, we use the bounds in Proposition 2.1 to obtain bounds on the expected posterior, $E_{\theta_0} w(\theta_0 \mid X^n)$. First, we give a lower bound, then we give an upper bound of two parts. The first part of the upper bound is the same as the lower bound. The second part of the upper bound is an additional error term. The main factor in it is

$$J_n(\theta_0) = \int_{(B_n \cap C_n)^c} w(\theta_0 \mid x^n)p(x^n \mid \theta_0)dx^n.$$ 

For convenience we write $\beta = 2^{d/2}(2\pi)^{-d/2}|I(\theta_0)|^{1/2}$.

We are unable to demonstrate that $J_n$ goes to zero in general. However, in Proposition 2.2 we give conditions under which $J_n$ will go to zero when $p_0$ is a member of a parametric family in exponential form. We conjecture that it does in fact go to zero under much weaker conditions but have not been able to identify them.

To state Theorem 2.1, let

$$L_1(\theta_0) = \frac{\nabla w(\theta_0)^I I^{-1}(\theta_0) \nabla w(\theta_0)}{2w^2(\theta_0)} - \frac{3 \text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{4w(\theta_0)}$$

$$- 2^{d/2} E[p_{2,\theta_0}(Z)e^{-Z'Z/2}]$$

$$- \frac{2^{d/2}}{w(\theta_0)} \nabla w^4(\theta_0) I^{-1/2}(\theta_0) E[Zp_{1,\theta_0}(Z)e^{-Z'Z/2}],$$

where $Z$ is $\text{Normal}(0, I_{d \times d})$ and the polynomials $p_{1,\theta_0}(Z)$ and $p_{2,\theta_0}(Z)$ are given in terms of the cumulants of the distribution defined by the density $p_{\theta_0}$. In the unidimensional parameter case they are the Hermite polynomials, see Feller (1971). The general definition for the multidimensional case is given in Bhattacharya and Rao (1986).

**Theorem 2.1.** Assume the conditions stated at the beginning of this section and in Proposition 2.1. Suppose $X_i$ is continuous with positive definite Fisher information matrix $I(\theta)$ and has finite fourth moments. Also, assume that the characteristic function of $X_1$ is an element of $L^p$ for some $p \geq 1$. Then we have the following bounds. Lower bound on the expected posterior:

$$\lim_{n \to \infty} n \left\{ \frac{2^{d/2}(2\pi)^{d/2}}{n^{d/2}|I(\theta_0)|^{1/2}} \int w(\theta_0 \mid x^n)p(x^n \mid \theta_0)dx^n - 1 \right\} \geq L_1(\theta_0).$$

Upper bound on the expected posterior:

$$\lim_{n \to \infty} n \left\{ \frac{2^{d/2}(2\pi)^{d/2}}{n^{d/2}|I(\theta_0)|^{1/2}} \int w(\theta_0 \mid x^n)p(x^n \mid \theta_0)dx^n - 1 \right\} \leq L_1(\theta_0) + \beta \lim_{n \to \infty} \frac{J_n(\theta_0)}{n^{d/2-1}},$$

where the first limit on the right hand side of (2.15b) is over appropriately chosen sequences of $\alpha_n$, $\epsilon_n$, $\delta_n \to 0$, satisfying $2\delta_n/(1 - \epsilon_n) < \alpha_n$. 
PROOF. See the Appendix A. □

Our next result gives sufficient conditions for \( J_n(\theta) \to 0 \) for appropriately chosen sequences of \( \alpha_n, \varepsilon_n, \delta_n \to 0 \). Suppose the parametric family \( p_\theta \) is of exponential family with \( d \) parameters in the natural parameterization. That is, write

\[
(2.16) \quad p(x \mid \theta) = e^{\theta' x - \psi(\theta)}.
\]

We show that the extra error term in the upper bound (2.15b) is bounded above by \( e^{-\varepsilon n} \) for some positive \( c \). Note that in an exponential family we can assume that the set \( B_n \) is the entire sample space since the second derivative of the log likelihood is not a function of the data \( x^n \). Furthermore, the set \( C_n \) reduces to the set on which the maximum likelihood estimator is not close to the true value of the parameter. Thus our task is to bound quantities of the form

\[
E_{\hat{\theta}_0}[w(\theta_0 \mid X^n)1_{|\theta - \theta_0| > \varepsilon}]
\]

where \( \hat{\theta} \) is the MLE of \( \theta_0 \).

**Proposition 2.2.** Suppose that \( p_\theta \) is an exponential family with the natural parametrization. Then, if the natural parameter space has nonvoid interior, we have that there is a \( \xi = \xi(\varepsilon) > 0 \) so that

\[
(2.17) \quad E_{\hat{\theta}_0}[w(\theta_0 \mid X^n)1_{|\theta - \theta_0| > \varepsilon}] \leq e^{-\xi n}.
\]

**Proof.** We have that

\[
p(x^n \mid \theta) = \exp \left\{ \theta \cdot \sum_{i=1}^{n} x_i + n\psi(\theta) \right\},
\]

where \( \theta = (\theta_1, \ldots, \theta_d) \) and \( x_i = (x_{i1}, \ldots, x_{id}) \) for \( i = 1, \ldots, n \). Consequently, for fixed \( n \) we obtain

\[
(2.18) \quad \int \frac{w(\theta) p(x^n \mid \theta)}{p(x^n \mid \theta_0)} d\theta 
\geq \int \hat{E}(\theta) \exp \left\{ (\theta - \theta_0) \sum_{i=1}^{n} x_i + n[\psi(\theta) - \psi(\theta_0)] \right\} d\theta,
\]

where \( \hat{E} \equiv \hat{E}(\varepsilon', \theta_0, \sum_{i=1}^{n} x_i) \subset \Omega \) is defined as follows. For \( \varepsilon' > 0 \), we require \( |\theta_i - \theta_{0i}| < \varepsilon' \) and for given \( \sum_{i=1}^{n} x_{ij} \), we require \( (\theta_i - \theta_{0i}) \sum_{i=1}^{n} x_{ij} \geq 0 \), where \( \theta_0 = (\theta_{01}, \ldots, \theta_{0d}) \). Since \( \hat{E}(\varepsilon', \theta_0, \sum_{i=1}^{n} x_{i}) \) depends on \( \sum_{i=1}^{n} x_{i} \) only on the signs of its \( d \) entries, there are \( 2^d \) possible sets. Denote them by \( E_1(\varepsilon', \theta_0), \ldots, E_{2^d}(\varepsilon', \theta_0) \).

Now we can bound the right hand side of (2.18) from below by

\[
(2.19) \quad \inf_{|\theta - \theta_0| < \varepsilon'} \min_{1 \leq i \leq 2^d} \int_{E_i(\varepsilon', \theta_0)} \exp\{n[\psi(\theta) - \psi(\theta_0)]\} d\theta.
\]
Since \( w \) is continuous and \( \psi \) is convex, we know that for given \( \eta > 0 \), and \( \delta > 0 \), there is \( \varepsilon' > 0 \) so small that

\[
(2.20) \quad \psi(\theta) - \psi(\theta_0) \geq -\eta \quad \text{and} \quad w(\theta) - w(\theta_0) \geq -\delta,
\]

if \( \max_{1 \leq i \leq d} |\theta_i - \theta_{0i}| < \varepsilon' \). Thus (2.19) and the left hand side of (2.18) is bounded below by

\[
(2.21) \quad [w(\theta_0) - \delta](\varepsilon')^d e^{-n\eta}.
\]

We use (2.21) to obtain the proposition. The right hand side of (2.17) is

\[
(2.22) \quad w(\theta_0) \int_{|\theta - \theta_0| \geq \varepsilon} \left\{ \int w(\theta) \frac{p(x^n | \theta)}{p(x^n | \theta_0)} d\theta \right\}^{-1} p(x^n | \theta_0) dx^n
\]

\[
\leq \frac{w(\theta_0) e^{n\eta} P_{\theta_0}(|\theta - \theta_0| \geq \varepsilon)}{|w(\theta_0) - \delta|(\varepsilon')^d}.
\]

Since the natural parameter space is convex and has nonvoid interior, \( p_\theta \) has a moment generating function which is finite on an neighborhood of zero. So, for given \( \varepsilon > 0 \), there is a \( c(\varepsilon) > 0 \), so that \( P_{\theta_0}(|\theta - \theta_0| \geq \varepsilon) \leq \exp(-nc(\varepsilon)) \). Now for given \( \varepsilon > 0 \), it is enough to choose \( \varepsilon' > 0 \) and \( \delta \) so small that (2.49) is satisfied for some \( \eta > 0 \) and \( \delta > 0 \) which also satisfies \( c(\varepsilon) - \eta > 0 \).

3. Asymptotic bounds for the expected square of the posterior

In this section we use adapt the proof of Theorem 2.1 to establish an asymptotic expansion for the expected square of the posterior. First, we define analogs of the quantities appearing in Theorem 2.1 and Proposition 2.2. Let

\[
(3.1a) \quad L_2(\theta_0) = \frac{\nabla w(\theta_0)^t I^{-1}(\theta_0) \nabla w(\theta_0)}{w^2(\theta_0)} - \frac{4 \text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{3w(\theta_0)}
\]

\[
- 3^{d/2} E|Z_{2, \theta_0}(Z)e^{-Z^tZ}|
\]

\[
- \frac{2(3^{d/2})}{w(\theta_0)} \nabla w(\theta_0) I^{-1/2}(\theta_0) E[Z_{0, \theta_0}(Z)e^{-Z^tZ}],
\]

where \( Z \) is Normal\((0, I_{d \times d})\) distributed. Also, write

\[
(3.1b) \quad \tilde{\beta} = \frac{|I(\theta_0)|}{3^{d/2}(2\pi)^d} \quad \text{and} \quad \tilde{J}_n(\theta_0) = \int_{(B_n \cap C_{\alpha})^c} \frac{|w(\theta_0 | x^n)|^2 p(x^n | \theta_0) dx^n}.
\]

Now we have the following theorem, and, as before, we control the error term \( \tilde{\beta} J_n(\theta_0) \) in a separate result.

**Theorem 3.1.** Under the assumptions of Theorem 2.1, we have the following bounds:

\[
(3.2a) \quad \lim_{n \to \infty} n \left\{ \frac{3^{d/2}(2\pi)^d}{n^d |I(\theta_0)|} \int [w(\theta_0 | x^n)|^2 p(x^n | \theta_0) dx^n - 1 \right\} \geq L_2(\theta_0)
\]

and
\[(3.2b) \quad \lim_{n \to \infty} n \left\{ \frac{3^{d/2}(2\pi)^d}{n^d |I(\theta_0)|} \int [w(\theta_0 | x^n)]^2 p(x^n | \theta_0) dx^n - 1 \right\} \leq L_2(\theta_0) + \beta \lim_{n \to \infty} \frac{\bar{J}_n(\theta_0)}{n^{d-1}}, \]

where the first limit on the right hand side of (3.2b) is over appropriately chosen sequences of \(\alpha_n, \varepsilon_n, \delta_n \to 0\), satisfying \(2\delta_n/(1 - \varepsilon_n) < \alpha_n\).

**Proof.** The proof is a straightforward modification of the proof of Theorem 2.1; it is given in the Appendix B. \(\square\)

The next result gives conditions under which \(\bar{J}_n(\theta) \to 0\) when \(p_\theta\) is of exponential family as in (2.16). As before, the sets \(B_n\) and \(C_n\) reduce to the set on which the maximum likelihood estimator is far from the true value of the parameter so our task is to bound \(E_{\theta_0} w(\theta_0 | X^n)^2 \chi_{[|\hat{\theta} - \theta_0| > \varepsilon]}\), where \(\hat{\theta}\) is the MLE of \(\theta_0\).

**Proposition 3.1.** Under the conditions of Proposition 2.2, we have that there is a \(\xi = \xi(\varepsilon) > 0\) so that

\[(3.3) \quad E_{\theta_0} w(\theta_0 | X^n)^2 \chi_{[|\hat{\theta} - \theta_0| > \varepsilon]} \leq e^{-\xi n}.\]

**Proof.** This is a straightforward modification of the proof of Proposition 2.2. \(\square\)

4. Implications of the results

The original motivation for Theorem 2.1 arose from the search for noninformative priors. Bernardo (1979) proposed the use of what he called reference priors, obtained by maximizing the relative entropy distance between a prior and its corresponding posterior over a large class of possible priors. Clarke and Sun (1997) performed an analogous optimization using the Chi-squared distance in place of the relative entropy. In this latter context, the \(O(1/n)\) terms in the expansion of Theorem 2.1 generate a functional which can be optimized so as to yield Chi-squared reference priors proportional to the inverse of the Fisher information. Since this application has been pursued at length elsewhere, we turn instead to implications of Theorems 2.1 and 3.1 more generally.
4.1 Heuristic derivation of confidence bands for the expected posterior

In Sections 2 and 3 formal asymptotic expansions for $E_{\theta_0}w(\theta_0 \mid X^n)$ and $E_{\theta_0}[w(\theta_0 \mid X^n)]^2$ were given for exponential families. These results lead to conjectures about the form of the expansions when the argument of the posterior density is not the true value of the parameter. Indeed, this case can be reduced to the case where the argument is $\theta_0$. Consider the identities

\begin{align}
(4.1a) \quad E_{\theta_0}[w(\theta \mid X^n)] &= \frac{w(\theta)}{w(\theta_0)} E_{\theta_0} \left[ w(\theta_0 \mid X^n) \exp \left( -n \left\{ \frac{1}{n} \log \left( \frac{p(x^n \mid \theta_0)}{p(x^n \mid \theta)} \right) \right\} \right) \right], \\
(4.1b) \quad E_{\theta_0}[w(\theta \mid X^n)]^2 &= \left( \frac{w(\theta)}{w(\theta_0)} \right)^2 E_{\theta_0} \left[ w(\theta_0 \mid X^n)^2 \exp \left( -2n \left\{ \frac{1}{n} \log \left( \frac{p(x^n \mid \theta_0)}{p(x^n \mid \theta)} \right) \right\} \right) \right].
\end{align}

By the law of large numbers, the exponent $n^{-1} \log[p(x^n \mid \theta_0)/p(x^n \mid \theta)]$ converges in various modes to the relative entropy between $p_{\theta_0}$ and $p_{\theta}$, which we write as

$$D(\theta_0 \parallel \theta) = \int p(x^n \mid \theta_0) \log \frac{p(x^n \mid \theta_0)}{p(x^n \mid \theta)} dx^n.$$ 

Making this substitution heuristically in (4.1a,b) gives

\begin{align}
(4.2a) \quad E_{\theta_0}w(\theta \mid X^n) &\approx \left[ \frac{w(\theta)}{w(\theta_0)} \right] e^{-nD(\theta_0 \parallel \theta)} E_{\theta_0}[w(\theta_0 \mid X^n)], \\
(4.2b) \quad E_{\theta_0}w(\theta \mid X^n)^2 &\approx \left[ \frac{w(\theta)}{w(\theta_0)} \right]^2 e^{-2nD(\theta_0 \parallel \theta)} E_{\theta_0}[w(\theta_0 \mid X^n)]^2 
\end{align}

for large $n$. Now, using Theorem 2.1 and Theorem 3.1 we have

\begin{align}
(4.3a) \quad E_{\theta_0}w(\theta \mid X^n) &\approx \left[ \frac{w(\theta)}{w(\theta_0)} \right] e^{-nD(\theta_0 \parallel \theta)} \sqrt{\frac{nI(\theta_0)}{4\pi}}, \\
(4.3b) \quad E_{\theta_0}w(\theta \mid X^n)^2 &\approx \left[ \frac{w(\theta)}{w(\theta_0)} \right]^2 e^{-2nD(\theta_0 \parallel \theta)} \frac{nI(\theta_0)}{\sqrt{32\pi}},
\end{align}

leaving out the terms of order $O(1/n)$ and smaller. For the unidimensional case, $d = 1$ and we note that by expressing the variance as the mean square minus the square mean we have

$$\text{Var}_{\theta_0}(w(\theta \mid X^n)) \approx \left[ \frac{w(\theta)}{w(\theta_0)} \right]^2 e^{-2nD(\theta_0 \parallel \theta)} \left( \frac{1}{\sqrt{3}} - \frac{1}{2} \right) \frac{nI(\theta_0)}{2\pi}.$$ 

If the details justifying (4.2a, b) can be supplied then one can obtain a confidence band for the expected posterior $E_{\theta_0}w(\theta \mid X^n)$. Note that the expected value
considers all the data that might have been got whereas the posterior, \( w(\theta \mid X^n) \),
only uses the actual set of data obtained.

It is interesting to note that the left hand side of (4.3a) is a probability density
for \( \theta \): it is positive and integrates to unity. The right hand side is nearly a density
also. The discrepancy arises from the neglect of terms that are \( O(1/n) \) and smaller.
Indeed, \( D(\theta_0 \parallel \theta) \) can be Taylor expanded at \( \theta_0 \) to see that for \( \theta \) near \( \theta_0 \)

\[
D(\theta_0 \parallel \theta) \approx \frac{1}{2} (\theta - \theta_0)^2 I(\theta_0)
\]

because \( D(\theta_0 \parallel \theta_0) = D'(\theta_0 \parallel \theta_0) = 0 \). By Laplace’s method, see De Bruijn (1961),
it is straightforward to verify that

\[
\int w(\theta)e^{-nD(\theta_0\parallel\theta)}d\theta \approx w(\theta_0)\sqrt{\frac{2\pi}{nI(\theta_0)}}
\]
as \( n \to \infty \). Comparison of (4.3a) and (4.5) suggests defining

\[
g_n(\theta_0, \theta) = \frac{w(\theta)}{w(\theta_0)} \sqrt{\frac{nI(\theta_0)}{\sqrt{2\pi}}e^{-nD(\theta_0\parallel\theta)}}.
\]

Now, (4.3a, b) become

\[
E_{\theta_0}[w(\theta \mid X^n)] \approx \frac{1}{\sqrt{2}} g_n(\theta_0, \theta)
\]

and

\[
E_{\theta_0}[w(\theta \mid X^n)^2] \approx \frac{1}{\sqrt{3}} g_n(\theta_0, \theta)^2.
\]

That is, the near-density \( g_n \) encapsulates in an asymptotic sense the expectation of
the posterior, apart from constant factors independent of the prior and parametric family. These factors \( 1/\sqrt{2} \) and \( 1/\sqrt{3} \) may arise from the fact that most location
estimators, such as the maximum likelihood estimator are asymptotically Chi-square, a fact which was neglected in the application of Laplace’s method since the expansion was around \( \theta_0 \) rather than around an estimator for \( \theta \).

Now consider trying to estimate \( E_{\theta_0}[w(\theta \mid X^n)] \), the expected posterior when
\( \theta_0 \) is true, by \( w(\theta \mid X^n) \). Markov’s inequality gives

\[
P_{\theta_0}(|w(\theta \mid X^n) - E_{\theta_0}[w(\theta \mid X^n)]| \geq \epsilon) \leq \frac{\text{Var}_{\theta_0}[w(\theta \mid X^n)]}{\epsilon^2}.
\]

So, it is seen from (4.4) that

\[
P_{\theta_0}(|w(\theta \mid X^n) - E_{\theta_0}[w(\theta \mid X^n)]| \leq \epsilon) \geq 1 - \alpha
\]
can be achieved for fixed confidence level \( 1 - \alpha \) by choosing

\[
[\epsilon_n(\theta_0, \theta)]^2 = (1/\sqrt{3} - 1/2)[g_n(\theta_0, \theta)]^2.
\]
That is, with this choice of of $\varepsilon_n(\theta_0, \theta)$,

$$w(\theta \mid x^n) \pm \frac{1}{\sqrt{1 - \alpha}} \left( \frac{1}{\sqrt{3}} - \frac{1}{2} \right)^{1/2} \varphi_n(\theta_0, \theta)$$

provides an approximate, asymptotic 1-\(\alpha\) confidence band for \(E_{\theta_0}[w(\theta \mid X^n)]\) if one replaces \(\theta_0\) in (4.8) by an estimator such as the maximum likelihood estimator. Such a confidence band indicates the reliability of inferences from the posterior density by giving a range of nearby densities, based on data near \(x^n\) that might have been obtained. (In principle, one might get a better confidence band if it were established that the posterior density, for each \(\theta\) was asymptotically normal with mean \(E_{\theta_0}[w(\theta \mid X^n)]\) and variance \(\text{Var}_{\theta_0}[w(\theta \mid X^n)]\) times some function of \(n\), in the usual frequentist sense.)

### 4.2 A frequentist refinement of Bayesian testing

Consider testing \(H_0 : \theta \in \Omega_0\) versus \(H_1 : \theta \in \Omega_1\), where both hypotheses are the closures of open bounded sets in the parameter space. A Bayes test is performed by calculating \(W(\Omega_0 \mid X^n)\), or, equivalently, the posterior odds ratio provided \(\Omega_0 = \Omega_0^c\) where the complement is taken within the parameter space. It is well known that this procedure is Bayes optimal under generalized zero-one loss, see Casella and Berger (1990). A frequentist test, by contrast, specifies a rejection region say \(A\), with the property that \(\sup_{\theta \in \Omega_0} P_\theta(A) \leq \alpha\), where \(\alpha\) is an upper bound on the probability of type I error. For frequentist optimality, one ensures \(P_\theta(A)\) is as large as possible for \(\theta \in \Omega_1\).

Kass and Raftery (1995) credit Good with the proposal that Bayes and frequentist testing can be partially reconciled by using the posterior probability as a test statistic. To see how this might work, note if the posterior probability is used to define a rejection region then one would accept \(H_0\) when \(W(\Omega_0 \mid x^n) \geq \beta\) where \(\beta\) is a threshold value. A frequentist would bound the power of this test by noting that, for \(\theta \in \Omega_1\),

$$P_\theta(W(\Omega_0 \mid X^n) \geq \beta) \leq \frac{1}{\beta^2} E_\theta \left\{ \int_{\Omega_0} w(\theta' \mid X^n) d\theta' \right\}^2,$$

by Markov's inequality. Multiplying and dividing \(w(\theta' \mid X^n)\) by \(\sqrt{w(\theta')}\) and applying the Cauchy-Schwartz inequality gives that the expectation in (4.9) is bounded by

$$\int_{\Omega_0} \frac{w(\theta')}{\sqrt{w(\theta')}} E_\theta \left\{ \int_{\Omega_0} \frac{w(\theta' \mid X^n)^2}{w(\theta')} d\theta' \right\}.$$

Noting that the first factor in (4.10) is bounded by unity, Fubini's theorem permits the interchange of expectation and integration so that (4.3b) gives

$$K_n I(\theta) \frac{1}{[\beta w(\theta)]^2} \int_{\Omega_0} w(\theta') e^{-2nD(\theta || \theta')} d\theta',$$
as an approximate upper bound for (4.9), where $K$ is a positive constant that can be determined. It is seen that as $n \to \infty$, the right hand side of (4.11) is of order $O(e^{-\gamma n})$ for some $\gamma > 0$. Apart from the value of $\gamma$ this is the same rate as is achieved by the Neyman-Pearson test in the simple versus simple case, so it cannot be improved substantially.

Despite its Bayes optimality, a Bayes test is incomplete in the sense that rejecting $H_0$ when $W(\Omega_0 | X^n)$ is too small does not take into account data sets near $X^n$ that might have given a different value for the posterior probability. In short, the Bayesian neglects the sampling distribution of $W(\Omega_0 | X^n)$. To remedy this, we suggest that $H_0$ should only be accepted when the posterior exceeds a threshold determined by use of the sampling distribution, to wit, accept $H_0$ when

\[ W(\Omega_0 | X^n) \geq \sup_{\theta \in \Omega_0} \left\{ E_\theta W(\Omega_0 | X^n) + 3\sqrt{\text{Var}_\theta[W(\Omega_0 | X^n)]} \right\} \tag{4.12} \]

and reject $H_0$ when

\[ W(\Omega_0 | X^n) \leq \inf_{\theta \in \Omega_0} \left\{ E_\theta W(\Omega_0 | X^n) - 3\sqrt{\text{Var}_\theta[W(\Omega_0 | X^n)]} \right\}. \tag{4.13} \]

Expression (4.3a) gives a heuristic asymptotic expansion for the expected posterior probability. We have been unable to derive an analogous expansion for the variance of the posterior probability. In between (4.12) and (4.13) the data cannot be regarded as sufficiently conclusive as to justify either $H_0$ or $H_1$.

In effect, this procedure uses a frequentist criterion, power, with frequentist near optimality in the sense of the power function to set a threshold for a Bayes test. As a consequence, this test is simultaneously Bayes optimal and frequentist ‘good’. The decision to accept or reject the null is based on a quantity equivalent to the posterior odds ratio which minimizes the posterior risk and has a well defined rejection region with power function bounded below by a function of the form $1 - O(e^{-\gamma n})$, as $n \to \infty$.

Finally, we examine the case that $H_0$ is a dimensional reduction from $H_1$. That is, for example, $H_0 : \mu_1, \ldots, \mu_p-1 \in \mathbb{R}, \mu_p = \mu_0$, and $H_1 : \mu_1, \ldots, \mu_p \in \mathbb{R}$. Here $\mu_0$ is a fixed constant. The usual Bayes factor would be

\[ W(\Omega_0 | X^n)/W(\Omega_0^c | X^n). \tag{4.14} \]

It is difficult even to conjecture an asymptotic form for the mean and variance of (4.14). We can nevertheless partially identify the behavior of the numerator or denominator separately. This does not use the present results, but does help to understand how one might use the posterior probability of a hypothesis as a test statistic.

Parallel to (4.12) and (4.13), we would like to find forms for $E_\theta W(\Omega_0 | X^n)$ and $\text{Var}_\mu[W(\Omega_0 | X^n)]$ where $\mu = (\mu_1, \ldots, \mu_p) \in \mathbb{R}^p$ assumed to be true. The denominator may be dependent statistically, but dealing with the numerator is enough for testing purposes.

Write the overall prior on $\mu$ as

\[ w(\mu) = w(\bar{\mu}^{p-1})[\omega w_d(\mu_p) + (1 - \alpha) w_c(\mu_p)], \]
where $0 < \alpha < 1$ is a fixed constant, $\tilde{\mu}^{p-1} = (\mu_1, \ldots, \mu_{p-1})$, $w_c$ is a continuous prior for $\mu_p$ with support $R$, and $w_d$ assigns mass 1 to $\mu_p = \mu_0$. Write $m_d(X^n)$ and $m_c(X^n)$ to mean the mixture of the likelihoods $p(X^n \mid \mu)$ with respect to $w_d(\mu) = w(\tilde{\mu}^{p-1})w_d(\mu_p)$ and $w_c(\mu) = w(\tilde{\mu}^{p-1})w_c(\mu_p)$. Now, the overall mixture is $m(X^n) = \alpha m_d(X^n) + (1 - \alpha)m_c(X^n)$ and we have

$$w(\mu \mid X^n) = \lambda(X^n)w_d(\mu \mid X^n) + [1 - \lambda(X^n)]w_c(\mu \mid X^n),$$

where $w_d(\mu \mid X^n)$ and $w_c(\mu \mid X^n)$ are the posteriors formed from $w_d(\mu)$ and $w_c(\mu)$, respectively, and $\lambda(X^n) = \alpha m_d(X^n)/m(X^n)$.

Since $W_d(\Omega \mid X^n) = 1$ and $W_c(\Omega_0 \mid X^n) = 0$, integrating in (4.15) gives

$$W(\Omega_0 \mid X^n) = \lambda(X^n) = \left[1 + \left(\frac{1 - \alpha}{\alpha}\right)\frac{m_c(X^n)}{m_d(X^n)}\right]^{-1}.$$

The main quantity in (4.16) is the density ratio in the denominator. Suppose we fix an element of $\Omega_0$ defined by $\tilde{\mu}^{p-1}$ and $\mu_p = \mu_0$. Then $m_c(X^n)/m_d(X^n)$ is

$$m_c(X^n) = \int \frac{w_c(\mu)p(X^n \mid \mu)d\mu}{p(X^n \mid \tilde{\mu}^{p-1}, \mu_0)} = \frac{\int w(\tilde{\mu}^{p-1})p(X^n \mid \tilde{\mu}^{p-1}, \mu_0)d\tilde{\mu}^{p-1}}{p(X^n \mid \tilde{\mu}^{p-1}, \mu_0)}.$$

The two density ratios in (4.17) have behavior partially given by Theorem 2.1, Clarke and Barron (1990). This result gives conditions under which

$$\log \left\{ \frac{p(X^n \mid \mu)}{m(X^n)} \right\} = \frac{d}{2} \log \left(\frac{n}{2\pi}\right) + \frac{1}{2} S J^{-1}(\mu) S - \log[w(\mu)]$$

$$+ \frac{1}{2} \log |J(\mu)| + o_p(1),$$

where $d$ is the dimension of $\mu$, $S = \frac{1}{\sqrt{n}} \nabla \log p(X^n \mid \mu)$, and $J$ is the Fisher information matrix. Using (4.18) for $d = p$ and $d = p - 1$ gives approximations for the two densities in (4.17). Substituting those approximations into (4.17) and then (4.16) gives a term of order $O_p(n^{-1/2})$. Thus, the expectation of (4.16) can be approximated as

$$E_\mu[W(\Omega_0 \mid X^n)] \approx \left[1 + \left(\frac{1 - \alpha}{\alpha}\right) o_p(n^{-1/2})\right]^{-1} \rightarrow 1.$$
in place of (4.17). Again, using (4.18) in the first and third factors of (4.20) (cases \( d = p, p - 1 \) as before) and using

\[
(4.21) \quad \frac{p(X^n \mid \tilde{\mu}^{p-1}, \mu_p')}{p(X^n \mid \tilde{\mu}^{p-1}, \mu_0)} = \exp\left[n \hat{D}(\tilde{\mu}^{p-1}, \mu_p' \parallel \tilde{\mu}^{p-1}, \mu_0)\right],
\]

where \( \hat{D} \) is the empirical relative entropy between distributions with the indicated parameter values, the middle factor of (4.20) gives an asymptotic approximation for use in (4.16). This gives the expression

\[
(4.22) \quad E_{(\tilde{\mu}^{p-1}, \mu_p')}[W(\Omega_0 \mid X^n)] \approx \left(1 + \left(\frac{1 - \alpha}{\alpha}\right) 0_p(e^n \hat{D} n^{-1/2})\right)^{-1} \rightarrow 0.
\]

for any \((\tilde{\mu}^{p-1}, \mu_p')\).

In closing, we note that a more careful analog would give more refined asymptotic forms for (4.19) and (4.22), and might give similarly refined forms for the variances of \( W(\Omega_0 \mid X^n) \). Note that the analysis here is in probability only, even though (4.18) holds in an \( L^1 \) sense also. Extending the present heuristics would justify the decision rule suggested by (4.12) and (4.13).

Acknowledgements

The authors thank an anonymous referee for suggestions that improved an earlier version of this manuscript.

Appendix A: Proofs from Section 2

In this Appendix we have recorded proofs the two main technical results in Section 2, namely Proposition 2.1 and Theorem 2.1.

PROOF OF PROPOSITION 2.1. First we obtain (2.14a). Note that as in Clarke and Barron (1990) p. 463 we have on \( A_n \cap B_n \) that

\[
(A.1) \quad \frac{m_n(X^n)}{p_0(X^n)} \leq (1 + \varepsilon)e^{(n/2)(1-\varepsilon)'}(\theta_0) I^{-1}(\theta_0)\ell_n(\theta_0) \int_{N(\theta_0, \alpha)} e^{-(n/2)(1-\varepsilon)^s f(\theta_0)(\theta-u) w(\theta) d\theta},
\]

where \( u = \theta_0 + \frac{1}{1-\varepsilon} I^{-1}(\theta_0)\ell'_n(\theta_0) \). Now we have the left hand side of (A.1) is bounded by

\[
(A.2) \quad \frac{m_n(X^n)}{p(X^n \mid \theta_0)} \leq H_n \left( \sum_{i=1}^{3} \int_{\mathbb{R}^d} e^{-(n(1-\varepsilon)/2)(\theta-u)^f I(\theta_0)(\theta-u)} G_i d\theta \right.
\]

\[
- \sum_{i=1}^{3} \int_{N(\theta_0, \alpha)} e^{-(n(1-\varepsilon)/2)(\theta-u)^f I(\theta_0)(\theta-u)} G_i d\theta \right).
\]
The summands in the first summation of the right hand side of (A.2) can be evaluated. The first gives

\begin{equation}
H_n G_1 (2\pi)^{d/2} \frac{1}{|n(1-\varepsilon)I(\theta_0)|^{1/2}},
\end{equation}

the second is zero, and the third is

\begin{equation}
\frac{H_n}{2} \frac{(2\pi)^{d/2}}{|n(1-\varepsilon)I(\theta_0)|^{1/2}} E[Z^t \nabla^2 \tilde{w} Z],
\end{equation}

where $Z \sim \text{Normal}(0, \frac{1}{n(1-\varepsilon)} I^{-1}(\theta_0))$. We note that the expectation in (A.4) is

\begin{align*}
E[\text{tr}(Z^t \nabla^2 \tilde{w} Z)] &= E[\text{tr}(\nabla^2 \tilde{w} ZZ^t)] \\
&= \text{tr}(\nabla^2 \tilde{w} I^{-1}(\theta_0) \frac{1}{n(1-\varepsilon)}) \\
&= \frac{1}{n(1-\varepsilon)} \text{tr}[\nabla^2 \tilde{w} I^{-1}(\theta_0)].
\end{align*}

So, the third term in the first summation on the right hand side of (A.2) is

\begin{equation}
\frac{H_n}{2} \frac{(2\pi)^{d/2}}{|n(1-\varepsilon)I(\theta_0)|^{1/2}} \frac{1}{n(1-\varepsilon)} \text{tr}(\nabla^2 \tilde{w} I^{-1}(\theta_0)).
\end{equation}

Next, we control the terms in the second summation on the right hand side of (A.2). The norm is defined by $I(\theta_0)$; $\theta$ is in $N(\theta_0, \alpha)^\varepsilon$; and we have that $\delta/(1-\varepsilon) < \alpha/2$. Consequently, we have

\begin{equation}
\|\theta - u\| \geq \|\theta - \theta_0\| - \frac{1}{1-\varepsilon} \|I^{-1}(\theta_0) \ell_n(\theta_0)\| \geq \frac{\alpha}{2},
\end{equation}

on $C_n$.

The first term in the second sum in (A.2) is controlled as follows. By (A.6), we have that

\begin{equation}
e^{-(n/2)(1-\varepsilon)(\theta - u)^t I(\theta_0) (\theta - u)} \leq e^{-(n/4)(1-\varepsilon)(\theta - u)^t I(\theta_0) (\theta - u)} e^{-(n/16)\alpha^2(1-\varepsilon)}.
\end{equation}

Now, we can bound the first term in the second sum by using this last inequality and enlarging the domain of integration. The result is

\begin{equation}
H_n |G_1| \frac{(2\pi)^{d/2} 2^{d/2}}{|n(1-\varepsilon)I(\theta_0)|^{1/2}} e^{-(n/16)\alpha^2(1-\varepsilon)}.
\end{equation}

By enlarging the domain of integration and applying the Cauchy-Schwartz inequality, the second term in the second sum in (A.2) is bounded above by

\begin{align*}
H_n e^{-(n/16)(1-\varepsilon)\alpha^2} &\int_{N_\varepsilon} \left| \left( \nabla w(\theta) + \frac{\nabla^2 \tilde{w} I^{-1}(\theta_0) \ell_n'(\theta_0)}{1-\varepsilon} \right)^t (\theta - u) \right| \\
&\cdot e^{-(n/4)(1-\varepsilon)(\theta - u)^t I(\theta_0) (\theta - u)} d\theta \\
&\leq H_n e^{-(n/16)(1-\varepsilon)\alpha^2} \frac{(2\pi)^{d/2} 2^{d/2}}{|n(1-\varepsilon)I(\theta_0)|^{1/2}} \sqrt{EZ^t G_4 G_4^t Z},
\end{align*}
where $Z \sim \text{Normal}(0, \frac{n^2}{n(1-\varepsilon)} I^{-1}(\theta_0))$. The expectation in (A.9) can be evaluated: Rearranging under trace gives

$$E Z' G_4 G_4^t Z = \frac{2}{n(1-\varepsilon)} \text{tr}[G_4 G_4^t I^{-1}(\theta_0)] = \frac{2}{n(1-\varepsilon)} G_4^t I^{-1}(\theta_0) G_4.$$  

Using (A.10) in (A.9) gives the bound

$$H_n e^{-(n/16)(1-\varepsilon)^2} \frac{2}{n(1-\varepsilon)} \text{tr}\left[\frac{(2\pi)^d/2}{n(1-\varepsilon)} I(\theta_0)^{1/2} \sqrt{E(Z^t \nabla \tilde{w}^t Z)^2}\right].$$  

To finish obtaining the upper bound note that the third summand in the second summation in (A.2) is bounded above by

$$e^{-(n/16)(1-\varepsilon)^2} \frac{2}{n(1-\varepsilon)} \frac{2}{n(1-\varepsilon)} \text{tr}\left[\frac{(2\pi)^d/2}{n(1-\varepsilon)} I(\theta_0)^{1/2} \sqrt{E(Z^t \nabla \tilde{w}^t Z)^2}\right].$$

where $Z \sim \text{Normal}(0, \frac{2}{n(1-\varepsilon)} I^{-1}(\theta_0))$. To evaluate the expectation in (A.12) note that for any $d \times d$ matrix $T$,

$$E(Z^t T Z) = \text{tr}\left[\frac{2}{n(1-\varepsilon)} T I^{-1}(\theta_0)\right],$$

and

$$\text{Var}(Z^t T Z) = 2 \text{tr}\left[\frac{4}{n^2(1-\varepsilon)^2} T I^{-1}(\theta_0) T I^{-1}(\theta_0)\right].$$

So, (A.12) is bounded from above by

$$H_n e^{-(n/16)(1-\varepsilon)^2} \sqrt{3} \frac{2}{n(1-\varepsilon)} \frac{2}{n(1-\varepsilon)} \text{tr}\left[\frac{(2\pi)^d/2}{n(1-\varepsilon)} I(\theta_0)^{1/2} \sqrt{E(Z^t \nabla \tilde{w}^t Z)^2}\right].$$

Since all terms in the second sum in (A.2) (namely, (A.8), (A.11) and (A.13)) are of lower order than the terms in the first sum in (A.2) (namely, (A.3), zero and (A.5)) we have that the upper bound (2.14a) holds with $C_1(\theta_0)$ as defined.

Next, we derive (2.14b). From Clarke and Barron (1990) p. 463, we observe that

$$\frac{m_n(X^n)}{p(X^n | \theta_0)} \geq e^{(n/2)(1+\varepsilon)\ell_n'(\theta_0) I(\theta_0) \ell_n'(\theta_0)} \int_{N(\theta_0, \alpha)} \cdot e^{-(1+\varepsilon)n/2}(\theta - \mu)^t I(\theta_0)(\theta - \mu) w(\theta) d\theta,$$

on $B_n \cap C_n$, where $u$ is now defined by $u = \theta_0 + \frac{1}{1+\varepsilon} I^{-1}(\theta_0) \ell_n'(\theta_0)$. The right hand side of (A.14) is bounded from below by

$$\frac{m_n(X^n)}{p(X^n | \theta_0)} \geq \tilde{H}_n \left( \sum_{i=1}^3 \int_{\mathbb{R}^d} \tilde{G}_i e^{-(1+\varepsilon)n/2}(\theta - \mu)^t I(\theta_0)(\theta - \mu) d\theta \right) - \sum_{i=1}^3 \int_{N(\theta_0, \alpha)} \tilde{G}_i e^{-(1+\varepsilon)n/2}(\theta - \mu)^t I(\theta_0)(\theta - \mu) d\theta.$$
Analogously to the first summation in (A.2), the summands in the first summation in (A.29) can be evaluated explicitly. They give, respectively,

\begin{equation}
\hat{H}_n \hat{G}_1 \frac{(2\pi)^{d/2}}{|n(1+\varepsilon)I(\theta_0)|^{1/2}},
\end{equation}

zero, and

\begin{equation}
\hat{H}_n \frac{(2\pi)^{d/2} \text{tr}(\nabla^2 w I^{-1}(\theta_0))}{2(1+\varepsilon)n |n(1+\varepsilon)I(\theta_0)|^{1/2}}.
\end{equation}

Next we deal with the terms in the second summation in (A.15). Note that for \( \theta \in N(\theta_0, \alpha^2) \) and \( \delta/(1+\varepsilon) < \alpha/2 \), we have, on \( C_n \), that \( \|\theta - \theta_0\| \geq \alpha/2 \). So, the first summand is bounded above by

\begin{equation}
\hat{H}_n |\hat{G}_1| \frac{(2\pi)^{d/2} 2^{d/2}}{|n(1+\varepsilon)I(\theta_0)|^{1/2}} e^{-(n/16)\alpha^2 (1+\varepsilon)},
\end{equation}

the second is bounded above by

\begin{equation}
\hat{H}_n \frac{(2\pi)^{d/2} 2^{d/2}}{|n(1+\varepsilon)I(\theta_0)|^{1/2}} \left[ \frac{2}{n(1+\varepsilon)} \hat{G}_4 I^{-1}(\theta_0) \hat{G}_4 \right] e^{-(n/16)(1+\varepsilon)\alpha^2},
\end{equation}

and, the third is bounded above by

\begin{equation}
\hat{H}_n \frac{(2\pi)^{d/2} 2^{d/2}}{|n(1+\varepsilon)I(\theta_0)|^{1/2}} \sqrt{\text{tr} \nabla^2 w I^{-1}(\theta_0) \nabla^2 w I^{-1}(\theta_0)} e^{-(n/16)(1+\varepsilon)\alpha^2}.
\end{equation}

Assembling (A.16) to (A.20) gives (2.14b). \( \square \)

**Proof of Theorem 2.1.** First observe that for \( \varepsilon \in (0, 1) \) and \( \alpha, \delta > 0 \) such that \( \delta/(1-\varepsilon) \leq \alpha/2 \) we have

\begin{equation}
\frac{1}{n^{d/2}} \int w(\theta_0 \mid x^n) p(x^n \mid \theta_0) dx^n \geq \frac{w(\theta_0)}{n^{d/2}} \int_{A_n \cap B_n \cap C_n} \frac{p(x^n \mid \theta_0)}{m(x^n)} p(x^n \mid \theta_0) dx^n.
\end{equation}

Using (2.14a) we lower bound the right hand side of (A.21) by

\begin{equation}
\frac{w(\theta_0)(1-\varepsilon)^{d/2}|I(\theta_0)|^{1/2}}{(2\pi)^{d/2}} \int_{A_n \cap B_n \cap C_n} \frac{p(x^n \mid \theta_0) dx^n}{H_n \left[ G_1 + \frac{\text{tr}(\nabla^2 w I^{-1}(\theta_0))}{2(1-\varepsilon)n} \right] e^{-n(1-\varepsilon)\alpha^2/16}}
\end{equation}

\begin{equation}
= \frac{(1-\varepsilon)^{d/2}|I(\theta_0)|^{1/2}}{(1+\varepsilon)(2\pi)^{d/2}} E_{\theta_0} \{ 1_{A_n \cap B_n \cap C_n} \frac{e^{-(1/2(1-\varepsilon))z_n^T z_n G^{-1}_n}}{G_n} \}.
\end{equation}
where $Z_n = \sqrt{n} I(\theta_0)^{-1/2} L_n(\theta_0)$ and

$$
G_5 = 1 + \frac{\nabla w(\theta_0)^I \bar{w} I^{-1/2}(\theta_0) Z_n}{\sqrt{n}(1-\varepsilon) w(\theta_0)} + \frac{Z_n^I \bar{w} I^{-1/2}(\theta_0) Z_n}{2n(1-\varepsilon)^2 w(\theta_0)} + \frac{\text{tr} \nabla^2 \bar{w} I^{-1}(\theta_0) C_1(\theta_0) e^{-n\alpha^2(1-\varepsilon)/16}}{2n(1-\varepsilon) w(\theta_0)}.
$$

By the restriction to $C_n$, $C_1(\theta_0)$ is bounded so the last term in the denominator in the expectation can be neglected. Since the other three nontrivial terms in the denominator sum to a small number we can apply the second order Taylor expansion of $(1 + x)^{-1}$. Now the argument of the expectation is

$$
1_{A_n \cap B_n \cap C_n} \left[ 1 - \frac{\nabla w(\theta_0)^I \bar{w} I^{-1/2}(\theta_0) Z_n}{(1-\varepsilon)^2 n w(\theta_0)} - \frac{Z_n^I \bar{w} I^{-1/2}(\theta_0) Z_n}{2(1-\varepsilon)^2 n w(\theta_0)} + \frac{\text{tr} \nabla^2 \bar{w} I^{-1}(\theta_0)}{2n(1-\varepsilon) w(\theta_0)} + \frac{(\nabla w(\theta_0)^I \bar{w} I^{-1/2}(\theta_0) Z_n)^2}{(1-\varepsilon)^2 n w(\theta_0)^2} + o\left(\frac{1}{n}\right) \right] e^{-Z_n^1 Z_n/(2(1-\varepsilon))}.
$$

Note that by results in Clarke and Barron (1990), $P_\theta(A_\varepsilon)$, $P_\theta(B_\varepsilon)$, and $P_\theta(C_\varepsilon)$ are all $o(1/n)$, so the indicator function does not affect the limiting behaviour of any of the terms in (A.23). Indeed, if one examines expression (6.3) in Clarke and Barron (1990), one sees that $P_\theta(A_\varepsilon) = o(1/n)$ follows from including the indicator function in the second expression in (6.3) so that the left hand side of (6.3) is $o(1/n)$ and this can be used at the end of the proof of Proposition 6.3. That $P_\theta(B_\varepsilon)$ and $P_\theta(C_\varepsilon)$ are $o(1/n)$ follows from (4.16) and (4.17) in Clarke and Barron (1990).

Next, we obtain the limiting behavior of the six terms in (A.23). The first term in (A.23) can be written as

$$
E_{\theta_0}[e^{-Z_n^1 Z_n/(2(1-\varepsilon))}] = E_{\theta_0}[1_{A_n \cap B_n \cap C_n}] e^{-Z_n^1 Z_n/(2(1-\varepsilon))}.
$$

The second term in (A.24) is $o(1/n)$. For the first term in (A.24) we use a local limit theorem from Bhattacharya and Rao (1986) (see Theorem 19.2, p. 192 and Sections 6 and 7 for definitions of quantities) so as to approximate it to order $o(1/n)$. Recall that the density $f_{n, \theta_0}$, of $Z_n$, has mean zero and variance matrix equal to the identity matrix in $d$ dimensions. Consequently, $f_{n, \theta_0}$ admits an expansion of the form

$$
f_{n, \theta_0}(z) = \varphi(z) - \varphi(z) \sum_{k=1}^r \frac{p_{k, \theta_0}(z)}{\eta^{k/2}} + o\left(\frac{1}{n^{r/2}}\right)
$$

uniformly in $z = (z_1, \ldots, z_d)$, where $\varphi$ is the standard normal density in $d$ dimensions and $p_{k, \theta_0}$ is a polynomial in $d$ arguments. We use (A.25) for $r = 2$ since we
want a $o(1/n)$ error term. The result is

\begin{equation}
(A.26) \ E[e^{-Z_n^*Z_n/(2(1-\varepsilon))}] = E[e^{-Z_n^*Z_n/(2(1-\varepsilon))}] - E \left[ e^{-Z_n^*Z_n/(2(1-\varepsilon))} \frac{p_{1,\theta_0}(Z)}{\sqrt{n}} \right] \\
- E \left[ e^{-Z_n^*Z_n/(2(1-\varepsilon))} \frac{p_{2,\theta_0}(Z)}{n} \right] + o \left( \frac{1}{n} \right),
\end{equation}

where $Z \sim Normal(0, I_{d \times d})$. The first term on the right hand side of (A.26) is $(1 + \frac{1}{\sqrt{n}})^{-d/2}$ by straightforward calculations since $Z^*Z$ is $\chi_2^2$. The second term on the right hand side of (A.26) is zero since $p_{1,\theta_0}(Z)$ is an odd function, see Bhattacharyya and Rao (1986), Expression (7.20). This leaves the third term which is difficult to evaluate in general. Expression (A.23) now gives

\begin{equation}
(A.27) \ n \left[ Ee^{-Z_n^*Z_n/(2(1-\varepsilon))} - \left( 1 + \frac{1}{1 - \varepsilon} \right)^{-d/2} \right] \\
= -E e^{-Z_n^*Z_n/(2(1-\varepsilon))}p_{2,\theta_0}(Z) + o(1).
\end{equation}

For the second term in (A.23) we use (A.25) with $r = 1$, since there is already a $\sqrt{n}$ in the denominator. We have the vector equation

\begin{equation}
(A.28) \ E[Z_ne^{-Z_n^*Z_n/(2(1-\varepsilon))}] = E[Z_n e^{-Z_n^*Z_n/(2(1-\varepsilon))}] \\
- E \left[ Z \frac{p_{1,\theta_0}(Z)}{\sqrt{n}} e^{-Z_n^*Z_n/(2(1-\varepsilon))} \right] + o \left( \frac{1}{\sqrt{n}} \right),
\end{equation}

in which the first term of the right hand side is zero. Thus the limiting behavior of the second term in (A.23) is

\begin{equation}
(A.29) \ \frac{\nabla w(\theta_0)I^{-1/2}(\theta_0)}{(1 - \varepsilon)\sqrt{n}w(\theta_0)} E[Z_ne^{-Z_n^*Z_n/(2(1-\varepsilon))}]
\end{equation}

\begin{equation}
= \frac{\nabla w(\theta_0)I^{-1/2}(\theta_0)}{n(1 - \varepsilon)w(\theta_0)} E[Zp_{1,\theta_0}(Z)e^{-Z_n^*Z_n/(2(1-\varepsilon))}] + o \left( \frac{1}{n} \right).
\end{equation}

The third, fourth and fifth terms are easier since they are already of order $O(1/n)$, with smaller order error terms. Thus, it is enough to use the asymptotic normality of $Z_n$ for all of them. For the third term we obtain

\begin{equation}
(A.30) \ E \left[ 1_{A_n \cap B_n \cap C_n} \frac{Z_n^2 I^{-1/2}(\theta_0)\nabla^2 w I^{-1/2}(\theta_0) Z_n e^{-Z_n^*Z_n/(2(1-\varepsilon))}}{2(1 - \varepsilon)w(\theta_0)} \right]
\end{equation}

\begin{equation}
= \frac{E[Z^2 I^{-1/2}(\theta_0)\nabla^2 w I^{-1/2}(\theta_0) Z e^{-Z^*Z/(2(1-\varepsilon))}]}{2(1 - \varepsilon)w(\theta_0)} \frac{1}{\text{tr}[\nabla^2 w I^{-1}(\theta_0)]}
\end{equation}

\begin{equation}
= \frac{2(1 - \varepsilon)^2 w(\theta_0)}{2(1 - \varepsilon)^2 w(\theta_0)} \left( 1 + \frac{1}{1 - \varepsilon} \right)^{d/2+1}.
\end{equation}
for the fourth term we obtain
\begin{equation}
\frac{\text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{2(1 - \varepsilon) w(\theta_0)} E_{\theta_0} \left[ 1_{A_n \cap B_n \cap C_n} e^{-2 Z_n^2 / (2(1 - \varepsilon))} \right] \\
\to \frac{\text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{2(1 - \varepsilon) w(\theta_0) \left( 1 + \frac{1}{1 - \varepsilon} \right)^{d/2}}
\end{equation}
and for the fifth term we obtain
\begin{equation}
E \left[ 1_{A_n \cap B_n \cap C_n} e^{-2 Z_n^2 / (2(1 - \varepsilon))} \frac{[\nabla w(\theta_0) I^{-1/2}(\theta_0) Z_n]^2}{(1 - \varepsilon)^2 w(\theta_0)^2} \right] \\
\to \frac{\nabla w(\theta_0) I^{-1/2}(\theta_0) \nabla w(\theta_0)}{(1 - \varepsilon)^2 w(\theta_0)^2 \left( 1 + \frac{1}{1 - \varepsilon} \right)^{d/2+1}}
\end{equation}
Putting together (A.27), (A.29) through (A.32) and letting \( n \) increase gives the stated lower bound.

The upper bound follows by writing the expected posterior as a sum of two integrals, one over \( B_n \cap C_n \) which can be controlled as in the lower bound, and one over \( (B_n \cap C_n)^c \) which gives the term involving \( J_\mu \). \( \Box \)

Appendix B: Proofs from Section 3

Here we give the proof of Theorem 3.1. We will only prove the lower bound (3.2a). The proof of (3.2b) is similar. First observe that for \( \varepsilon \in (0, 1) \) and \( \alpha, \delta > 0 \) such that \( \delta/(1 - \varepsilon) \leq \alpha/2 \) we have
\begin{equation}
\frac{1}{n^d} \int u^2(\theta_0 | x^n) p(x^n | \theta_0) dx^n \\
\geq \frac{w^2(\theta_0)}{n^d} \int_{A_n \cap B_n \cap C_n} \left\{ \frac{p(x^n | \theta_0)}{m(x^n)} \right\}^2 p(x^n | \theta_0) dx^n.
\end{equation}
Using (2.14a) we lower bound the right hand side of (B.1) by
\begin{align*}
\frac{w^2(\theta_0)(1 - \varepsilon)^d I(\theta_0)}{(2\pi)^d} \\
\cdot \frac{1}{H_\mu^2} \frac{p(x^n | \theta_0) dx^n}{G_1 + \frac{\text{tr} \nabla^2 m^{-1}(\theta_0)}{2n(1 - \varepsilon)} + C_1(\theta_0) e^{-n(1 - \varepsilon)\alpha^2 / 16}} \\
= \frac{(1 - \varepsilon)^d I(\theta_0)}{(1 + \varepsilon)(2\pi)^d} E_{\theta_0} \left[ 1_{A_n \cap B_n \cap C_n} G_5^{-2} e^{-Z_n^2 / (1 - \varepsilon)} \right],
\end{align*}
where \( Z_n = \sqrt{n} I(\theta_0)^{-1/2} \ell_n(\theta_0) \) and \( G_5 \) is given by (A.22). By the restriction to \( C_n \), \( C_1(\theta_0) \) is bounded so the last term in the denominator in the expectation can
be neglected. Since the other three nontrivial terms in the denominator sum to a small number we can apply the second order Taylor expansion of \((1 + x)^{-2}\). Now the argument of the expectation is

\[
E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}]
\]

As before, \(P_{\theta}(A_n^c), P_{\theta}(B_n^c)\) and \(P_{\theta}(C_n^c)\) are all \(o(1/n)\) so the indicator function does not affect the limiting behaviour of any of the terms in (B.2). Let us consider the limiting behavior of the six terms in (B.2).

The first term in (B.2) can be written as

\[
E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E_{\theta_0}[e^{-Z_n^T Z_n/(1-\varepsilon)}]
\]

The second term in (B.3) is \(o(1/n)\). For the first term in (B.3), we use the expansion (A.24) with \(r = 2\) for density \(f_{n,\theta_0}\), of \(Z_n\). The result is

\[
E[e^{-Z_n^T Z_n/(1-\varepsilon)}] = E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}]
\]

where \(Z \sim \text{Normal}(0, I_{d \times d})\). The first term on the right hand side of (B.3) is \((1 + \frac{2}{1-\varepsilon})^{-d/2}\) by straightforward calculations since \(Z^T Z = \chi_d^2\). The second term on the right hand side of (B.3) is zero since \(p_{1,\theta_0}(Z)\) is an odd function, see Bhattacharyya and Rao (1986), Expression (7.20). This leaves the third term which is difficult to evaluate in general. Expression (B.3) now gives

\[
E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}] - E[e^{-Z_n^T Z_n/(1-\varepsilon)}] = o(1).
\]

For the second term in (B.2) we use (A.24) with \(r = 1\), since there is already a \(\sqrt{n}\) in the denominator. We have the vector equation

\[
E[Z_n e^{-2Z_n^T Z_n/(1-\varepsilon)}] = E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}]
\]

in which the first term of the right hand side is zero. Thus the limiting behavior of the second term in (B.2) is

\[
E[Z_n e^{-2Z_n^T Z_n/(1-\varepsilon)}] = E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}] - E[Z e^{-Z^T Z/(1-\varepsilon)}]
\]
The third, fourth and fifth terms are easier since they are already of order \( O(1/n) \), with smaller order error terms. Thus, it is enough to use the asymptotic normality of \( Z_n \) for all of them. For the third term we obtain

\[
(B.8) \quad E \left[ 1_{A_n \cap B_n \cap C_n} \frac{Z_n^* I^{-1/2}(\theta_0) \nabla^2 \bar{w} I^{-1/2}(\theta_0) Z_n e^{-Z_n^* Z_n/(1-\varepsilon)}}{(1-\varepsilon)^2 w(\theta_0)} \right] \\
\quad \to \frac{1}{(1-\varepsilon)^2 w(\theta_0)} E[Z_n^* I^{-1/2}(\theta_0) \nabla^2 \bar{w} I^{-1/2}(\theta_0) Z_n e^{-Z_n^* Z_n/(1-\varepsilon)}] \\
\quad = \frac{\text{tr} \nabla^2 \bar{w} I^{-1}(\theta_0)}{(1-\varepsilon)^2 w(\theta_0)} \left( 1 + \frac{2}{1-\varepsilon} \right)^{d/2+1},
\]

for the fourth term we obtain

\[
(B.9) \quad \frac{\text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{(1-\varepsilon) w(\theta_0)} E_{\theta_0} \left[ 1_{A_n \cap B_n \cap C_n} e^{-Z_n^* Z_n/(1-\varepsilon)} \right] \\
\quad \to \frac{\text{tr} \nabla^2 w(\theta_0) I^{-1}(\theta_0)}{(1-\varepsilon) w(\theta_0)} \left( 1 + \frac{2}{1-\varepsilon} \right)^{d/2+1},
\]

and for the fifth term we obtain

\[
(B.10) \quad E \left[ 1_{A_n \cap B_n \cap C_n} e^{-Z_n^* Z_n/(1-\varepsilon)} \frac{3|\nabla w(\theta_0)^T I^{-1/2}(\theta_0) Z_n|^2}{(1-\varepsilon)^2 w(\theta_0)^2} \right] \\
\quad \to \frac{3 \nabla w(\theta_0)^T I^{-1/2}(\theta_0) \nabla w(\theta_0)}{(1-\varepsilon)^2 w(\theta_0)^2} \left( 1 + \frac{2}{1-\varepsilon} \right)^{d/2+1}.
\]

Putting together (B.5) through (B.10) gives the stated lower bound.

REFERENCES


