The Annals of Statistics
1994, Vol. 22, No. 4, 1808–1830

INTEGRABLE EXPANSIONS FOR POSTERIOR DISTRIBUTIONS FOR A TWO-PARAMETER EXPONENTIAL FAMILY[1]

BY DONGCHU SUN

University of Missouri, Columbia

Asymptotic expansions of posterior distributions are derived for a two-dimensional exponential family, which includes normal, gamma, inverse gamma and inverse Gaussian distributions. Reparameterization allows us to use a data-dependent transformation, convert the likelihood function to the two-dimensional standard normal density and apply a version of Stein's identity to assess the posterior distributions. Applications are given to characterize optimal noninformative priors in the sense of Stein, to suggest the form of a high-order correction to the distribution function of a sequential likelihood ratio statistic and to provide confidence intervals for one parameter in the presence of other nuisance parameters.

1. Introduction. The question of asymptotic posterior expansions can be dated from Laplace (1847) and has become one of the most widely studied problems in statistical theory and application. Johnson (1970) was among the first authors to investigate pointwise expansions of the posterior distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$ rigorously. Here $\widehat{\theta}_n$ is the maximum likelihood estimator of a parameter θ given n iid observations. Ghosh, Sinha and Joshi (1982) and Bickel, Götze and Van Zwet (1985) gave a uniform variant of Johnson's result for a one-parameter case.

A natural application of posterior expansions is to find Bayesian confidence intervals of θ. Alternatively, Bayesian confidence intervals can be derived from log-likelihood ratio statistics. Let $L_n(\theta)$ denote the likelihood function and $l_n(\theta) = \log(L_n(\theta))$. The likelihood ratio $L_n(\widehat{\theta})/L_n(\theta)$ or equivalently $2\{l_n(\widehat{\theta}_n) - l_n(\theta)\}$ can be studied to assess the truthfulness of the simple null hypothesis $H_0\colon \theta = \theta_0$. Efron (1985) proved the asymptotic normality of the posterior distribution of $\Lambda_n = \{2[l_n(\widehat{\theta}_n) - l_n(\theta)]\}^{1/2}\,\mathrm{sign}(\widehat{\theta}_n - \theta)$, the signed squared root of the log-likelihood ratio statistic, and recommended the use of Λ_n, rather than $\sqrt{n}(\widehat{\theta}_n - \theta)$, to find a Bayesian confidence interval. The use of the signed squared root goes back to Lawley (1956), Woodroofe (1986) and Barndorff-Nielsen (1986), all from a frequentist point of view. Bickel and Ghosh (1990) extended Efron's (1985) result to a multiparameter case and pointed out that the normal approximation to Λ_n is correct to a higher order than the normal approximation to $\sqrt{n}(\widehat{\theta}_n - \theta)$.

A posterior expansion of Λ_n can also be used to characterize priors leading to posterior confidence regions with approximate frequentist validity. In a single-parameter case, Welch and Peers (1963) proved that a one-sided poste-

Received August 1992; revised March 1994.

[1]Partially supported by a research board grant from the University of Missouri System.

AMS 1991 *subject classification.* Primary 62L12.

Key words and phrases. Posterior distributions, Stein's identity, martingale convergence theorem, stopping time, Jeffreys' prior, reference prior, repeated likelihood ratio tests.

rior confidence interval from the Jeffreys (1961) prior, which is proportional to the square root of the determinant of the Fisher information matrix, has the desired coverage probability up to $O(n^{-1})$. In spite of the success in one-parameter problems, the Jeffreys prior frequently runs into serious deficiencies in multiparameter problems [cf. Berger and Bernardo (1992)]. To overcome these deficiencies, Stein (1985) extended the results in Welch and Peers (1963) and Peers (1965) and introduced a method to find a prior by requiring the frequentist coverage probability of the posterior region of a real-valued parametric function to match the normal level with a remainder of $O(n^{-1})$. Tibshirani (1989) generalized this method by using a one-to-one transformation of the parameter vector into a parameter of interest and a nuisance parameter vector orthogonal in the sense of Cox and Reid (1987). Berger and Bernardo (1989, 1992) extended the reference prior approach of Bernardo (1979), giving a general algorithm to derive a reference prior by splitting the parameters into several groups according to their order of inferential importance. Ghosh and Mukerjee (1992) applied Bickel and Ghosh's (1990) result to derive a noninformative prior matching the frequentist and posterior coverage probabilities of a confidence interval of a parameter in the presence of nuisance parameters.

Another important problem is to obtain posterior expansions when a stopping time is involved. Alvo (1977) and Ghosh, Sinha and Joshi (1982) derived posterior expansions for Bayesian sequential estimation in a one-parameter natural exponential family. Unfortunately, this area has received inadequate attention. With the wide use of Bayesian sequential designs and Bayesian sequential tests [cf. Woodroofe (1982, 1989), Lalley (1983) and Hu (1988)], it is desirable to find simple posterior expansions.

The main technique in the articles above is Taylor expansion. For a one-parameter natural exponential family, Woodroofe (1992) noticed that the likelihood function of Λ_n, the signed squared root of the log-likelihood function, is exact normal. He called Λ_n a data-dependent transformation and showed that the remainder term is exactly a conditional expectation. Therefore, martingale theory can be employed to deal with the integrability problem, especially for sequential problems, thus avoiding messy Taylor expansions. Woodroofe's (1992) results can be easily generalized to the special case of a multiparameter exponential family that is the product of several independent one-parameter natural exponential families. However, it is not clear how this can be done for arbitrary multiparameter exponential families [cf. Brown (1986)] such as a multinomial distribution. In this paper, we will consider a two-parameter exponential family of densities on the Borel sets of $\mathbb{R}$,

$$p^*_{(\theta_1,\beta_2)}(x) = \exp\{\theta_1 U_1(x) + \beta_2 U_2(x) - \psi^*(\theta_1,\beta_2)\}, \tag{1}$$

with respect to some σ-finite measure. Bar-Lev and Reiser (1982) first introduced a subfamily of (1) which is characterized by the following two assumptions.

ASSUMPTION A. The parameter β_2 can be represented as: $\beta_2 = -\theta_1 G_2'(\theta_2)$, where $\theta_2 = \mathbb{E}_{(\theta_1,\beta_2)}(U_2(X))$, $G_2'(\theta_2) = dG_2(\theta_2)/d\theta_2$, for some function G_2.

TABLE 1

Examples of a certain two-parameter exponential family of distributions where $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$, $Y_n = \sum_{i=1}^n U_1(X_i) - nG_2((1/n)\sum_{i=1}^n U_2(X_i))$, $\mathcal{N}$ *is the natural parameter space for* (θ_1, β_2) *and* $g(\theta_1) = -\theta_1 + \theta_1 \log(-\theta_1) + \log[\Gamma(-\theta_1)]$

Name	Normal	Inverse Gaussian	Gamma	Inverse gamma
Density	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	$\sqrt{\frac{\alpha}{2\pi x^3}} e^{-\alpha/2x - \mu x/2 + \sqrt{\alpha\mu}}$	$\frac{\alpha^\alpha x^{\alpha-1}}{\Gamma(\alpha)\mu^\alpha} e^{-\alpha x/\mu}$	$\frac{\alpha^\alpha}{\Gamma(\alpha)\mu^\alpha x^{\alpha+1}} e^{-\alpha/\mu x}$
$U_1(x)$	x^2	$\frac{1}{x}$	$-\log(x)$	$-\log(x)$
$U_2(x)$	x	x	x	$\frac{1}{x}$
θ_1	$-\frac{1}{2\sigma^2}$	$\frac{-\alpha}{2}$	$-\alpha$	$-\alpha$
β_2	$\frac{\mu}{\sigma^2}$	$\frac{-\mu}{2}$	$\frac{-\alpha}{\mu}$	$\frac{-\alpha}{\mu}$
θ_2	μ	$\sqrt{\frac{\alpha}{\mu}}$	μ	μ
$\mathcal{N}$	$\mathbb{R}^- \times \mathbb{R}$	$\mathbb{R}^- \times (\mathbb{R}^- \cup \{0\})$	$\mathbb{R}^- \times \mathbb{R}^-$	$\mathbb{R}^- \times \mathbb{R}^-$
Θ	$\mathbb{R}^- \times \mathbb{R}$	$\mathbb{R}^- \times \mathbb{R}^+$	$\mathbb{R}^- \times \mathbb{R}^+$	$\mathbb{R}^- \times \mathbb{R}^+$
$G_1(\theta_1)$	$-\frac{1}{2}\log(-2\theta_1)$	$-\frac{1}{2}\log(-2\theta_1)$	$g(\theta_1)$	$g(\theta_1)$
$G_2(\theta_2)$	θ_2^2	$\frac{1}{\theta_2}$	$-\log(\theta_2)$	$-\log(\theta_2)$
Y_n	$\sum_{i=1}^n (X_i - \bar{X}_n)^2$	$\sum_{i=1}^n \frac{1}{X_i} - \frac{n}{\bar{X}_n}$	$\sum_{i=1}^n \log\left(\frac{\bar{X}_n}{X_i}\right)$	$\sum_{i=1}^n \log\left(\frac{1}{X_i}\right) - \log\left(\frac{1}{n}\sum_{j=1}^n \frac{1}{X_j}\right)$

ASSUMPTION B. $U_2(x)$ is a one-to-one function on the support of p^*.

The normal, inverse Gaussian, gamma and inverse gamma distributions are special cases of (1) satisfying Assumptions A and B (see Table 1). Any monotone function of these random variables also satisfies the two assumptions. One can reparameterize (1) by the mapping $(\theta_1, \beta_2) \to (\theta_1, \theta_2)$. It is easy to verify that θ_1 and θ_2 are orthogonal in the sense of Cox and Reid (1987). In addition, from Theorem 8.4 of Barndorff-Nielsen (1978), this mapping is a homeomorphism with range space $\Theta = \Theta_1 \times \Theta_2$ (i.e., components θ_1 and θ_2 vary independently).

Under Assumptions A and B, one can show [cf. Bar-Lev and Reiser (1982)] that G_2 is infinitely differentiable, G_2'' is positive and the variance of U_2 is $-1/\{\theta_1 G_2''(\theta_2)\} > 0$. Therefore, θ_1 never changes its sign. Without loss of generality, we will assume that $\Theta_1 \subset \mathbb{R}^- \equiv (-\infty, 0)$. With this reparameterization, write

$$p_{\boldsymbol{\theta}}(x) = \exp\{\theta_1 U_1(x) - \theta_1 G_2'(\theta_2) U_2(x) - \psi(\boldsymbol{\theta})\}, \tag{2}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and $\psi(\boldsymbol{\theta}) = -\theta_1[\theta_2 G_2'(\theta_2) - G_2(\theta_2)] + G_1(\theta_1)$ for some differentiable function$G_1(\theta_1)$. We need an additional assumption on the function G_1 for this paper.

ASSUMPTION C. $\sup_{x \le \theta_1} |x| G_1'(x) = \gamma(\theta_1) < \infty$ for any $\theta_1 \in \Theta_1$.

Note that Assumption C holds for all the distributions in Table 1. Basically, it controls the left tail behavior of the underlying distribution function. Similar conditions have been used by Woodroofe (1978) and Bose and Boukai (1993) for various design problems.

The purpose of this paper is to establish the following kinds of results. Some preliminary information is given in Section 2. Motivated by Bickel and Ghosh (1990), a data-dependent transformation (Z_1, Z_2) of (θ_1, θ_2) is defined, that generalizes the signed squared root. An extension of Stein's identity is given to provide a key to the posterior expansion, so that the remainder terms can be written as conditional expectations and treated by martingale convergence theory. Some important inequalities are also derived for later use.

Asymptotic expansions for the posterior expectation of functions of (Z_1, Z_2) as n tends to ∞ are given in Section 3.1. Second-order and higher-order expansions are given in Sections 3.2 and 3.3, respectively, as a sequence of stopping times tends to ∞. Although Bickel and Ghosh's (1990) results are general for a fixed sample size, it is not easy to verify the conditions for which the expansions are valid. In contrast, the results here hold for the distribution family (2) even when a stopping time is involved. Bartlett (1937) showed that the χ^2 distribution is a far better approximation to the distribution of $\Lambda_n / \mathbb{E}_\theta \Lambda_n$ than to Λ_n itself. Box (1949) and Lawley (1956) generalized the Bartlett correction. A general concept, expansion by rescaling, was established by Efron (1985) and Woodroofe (1992) for a univariate case and Bickel and Ghosh (1990) for a general case with fixed sample size. In Section 3.4, asymptotic integrable expansions by rescaling are also studied for polynomial functions and symmetric functions of (Z_1, Z_2).

Three applications of the main results are given in Section 4. First, second-order expansions from subsection 3.2 are used to characterize some optimal noninformative priors in the sense of Stein (1985). The expression of the optimal noninformative prior is derived and the relation with Tibshirani's and Berger and Bernardo's reference priors is discussed. It is shown that the Jeffreys prior could be improved by taking the stopping time into account [cf. Ye (1993) for a one-parameter case]. Second, expansions by rescaling are applied to suggest the form of a high-order correction to the distribution function of a sequential likelihood ratio statistic. Our correction to the sampling distribution of the sequential log-likelihood ratio statistic for the normal population with unknown mean and variance greatly improves the chi-square approximation, especially for small sample sizes. Finally, inferences about θ_1 in the presence of the nuisance parameter θ_2 after sequential experiments are investigated.

Due to space limitations, some proofs are given in the Appendix and many proofs are outlined. Readers who desire a more complete exposition can contact the author.

2. Preliminaries.

2.1. *Data-dependent transformation.* Consider a Bayesian model in which the prior density of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is ξ, and $X_1, X_2, \ldots$ are conditionally iid with density (2) for given $\boldsymbol{\theta} \in \Theta$. We will denote probability and expectation in the Bayesian model by $\mathbb{P}^{\xi}$ and $\mathbb{E}^{\xi}$, conditional probability and expectation given $\boldsymbol{\theta}$ by $\mathbb{P}_{\boldsymbol{\theta}}$ and $\mathbb{E}_{\boldsymbol{\theta}}$ and conditional expectation given $X_1, X_2, \ldots, X_n$ by $\mathbb{E}_n^{\xi}$. For simplicity, $\boldsymbol{\theta}$ will be treated as either a random variable or its observation. The log-likelihood function of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ given $X_1, \ldots, X_n$ is

$$l_n(\boldsymbol{\theta}) = \theta_1 T_{n1} - \theta_1 G_2'(\theta_2) T_{n2} - n\psi(\boldsymbol{\theta}),$$

where $T_{nj} = \sum_{i=1}^n U_j(X_i)$. Let $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_{n1}, \widehat{\theta}_{n2})$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$. Note that $\widehat{\theta}_{n2} = T_{n2}/n$, and $\widehat{\theta}_{n1}$ satisfies $G_1'(\widehat{\theta}_{n1}) = \overline{Y}_n$, where $\overline{Y}_n = \{T_{n1} - nG_2(T_{n2}/n)\}/n$. It follows from Bar-Lev and Reiser (1982) that both G_1 and G_2 in (2) are strictly convex and infinitely differentiable. The log-likelihood ratio test statistic is then

$$l_n(\widehat{\boldsymbol{\theta}}_n) - l_n(\boldsymbol{\theta}) = nI(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}) \equiv nI_1(\widehat{\theta}_{n1}, \theta_1) - n\theta_1 I_2(\widehat{\theta}_{n2}, \theta_2), \tag{3}$$

where

$$I_1(\omega_1, \theta_1) = G_1(\theta_1) - G_1(\omega_1) - G_1'(\omega_1)(\theta_1 - \omega_1), \qquad \omega_1, \theta_1 \in \Theta_1, \tag{4}$$

$$I_2(\omega_2, \theta_2) = G_2(\omega_2) - G_2(\theta_2) - G_2'(\theta_2)(\omega_2 - \theta_2), \qquad \omega_2, \theta_2 \in \Theta_2. \tag{5}$$

The convexity of G_j implies that $I_j(\omega_j, \theta_j)$ is always nonnegative and equals 0 if and only if $\omega_j = \theta_j$. Therefore $I(\boldsymbol{\omega}, \boldsymbol{\theta}) \geq 0$, and equality holds if and only if $\boldsymbol{\omega} = \boldsymbol{\theta}$. [$I(\boldsymbol{\omega}; \boldsymbol{\theta})$ is the Kullback–Leibler divergence between $p_{\boldsymbol{\omega}}$ and $p_{\boldsymbol{\theta}}$ defined by $\mathbb{E}_{\boldsymbol{\omega}} \log\{p_{\boldsymbol{\omega}}(X_1)/p_{\boldsymbol{\theta}}(X_1)\}$.]

The data-dependent transformation is defined by

$$\mathbf{Z} \equiv \mathbf{Z}_n \equiv (Z_{n1}, Z_{n2}) = \begin{pmatrix} \sqrt{2nI_1(\widehat{\theta}_{n1}, \theta_1)}\, \text{sign}(\theta_1 - \widehat{\theta}_{n1}) \\ \sqrt{-2n\theta_1 I_2(\widehat{\theta}_{n2}, \theta_2)}\, \text{sign}(\theta_2 - \widehat{\theta}_{n2}) \end{pmatrix}^{\tau}. \tag{6}$$

Denote the partial derivatives of I_k with respect to θ_k by $I_{k,j}(\omega_k, \theta_k) = \partial^j I_k(\omega_k, \theta_k)/\partial\theta_k^j$. Since

$$\det\left(\frac{\partial(Z_{n1}, Z_{n2})}{\partial(\theta_1, \theta_2)}\right) = \det\begin{pmatrix} \sqrt{n}|I_{1,1}(\widehat{\theta}_{n1}, \theta_1)| \Big/ \sqrt{2I_1(\widehat{\theta}_{n1}, \theta_1)} & 0 \\ * & \sqrt{-n\theta_1}|I_{2,1}(\widehat{\theta}_{n2}, \theta_2)| \Big/ \sqrt{2I_2(\widehat{\theta}_{n2}, \theta_2)} \end{pmatrix}$$

is positive, the transformation from $\boldsymbol{\theta}$ to $\mathbf{Z}_n$ is one-to-one and onto. Let S_n be the range of $\mathbf{Z}_n$ and define $J_k(\omega_k, \theta_k) = \sqrt{2I_k(\omega_k, \theta_k)}/|I_{k,1}(\omega_k, \theta_k)|$ and $J(\omega_1, \omega_2; \theta_1, \theta_2) = (1/\sqrt{-\theta_1})\, J_1(\omega_1, \theta_1) J_2(\omega_2, \theta_2)$. Then the conditional density of $\mathbf{Z}_n$ given $X_1, \ldots, X_n$ is

$$\zeta_n(\mathbf{z}) \propto \xi(\boldsymbol{\theta}) J(\widehat{\theta}_{n1}, \widehat{\theta}_{n2}; \theta_1, \theta_2) \exp\left\{-\frac{1}{2}\mathbf{z}\mathbf{z}^{\tau}\right\} 1_{S_n}(\mathbf{z}). \tag{7}$$

Here 1_A denotes the indicator function of an event A. The partial derivatives of J_k with respect to θ_k will be used:

$$J_{k,1}(\omega_k, \theta_k) = \frac{1}{\sqrt{2I_k}}\{1 - I_{k,2}J_k^2\}\operatorname{sign}(\theta_k - \omega_k), \tag{8}$$

$$J_{k,2}(\omega_k, \theta_k) = -\left\{\frac{|I_{k,1}|}{2I_k}J_{k,01} + \frac{1}{\sqrt{2I_k}}\left[I_{k,3}J_k^2 + 2I_{k,2}J_kJ_{k,1}\right]\right\}\operatorname{sign}(\theta_k - \omega_k) \tag{9}$$

for $\omega_k \neq \theta_k \in \Theta_k, k = 1,2$. The value of J_k and its partial derivatives on the diagonal may be obtained from l'Hospital's rule as

$$\begin{aligned}
J_k(\omega_k, \omega_k) &= \frac{1}{\sqrt{G_k''(\omega_k)}}, \qquad k = 1,2,\\
J_{1,1}(\omega_1, \omega_1) &= -\frac{G_1'''(\omega_1)}{3G_1''(\omega_1)^{3/2}},\\
J_{1,2}(\omega_1, \omega_1) &= \frac{11G_1'''(\omega_1)^2}{36G_1''(\omega_1)^{5/2}} - \frac{G_1^{\text{iv}}(\omega_1)}{4G_1''(\omega_1)^{3/2}},\\
J_{2,1}(\omega_2, \omega_2) &= -\frac{2G_2'''(\omega_2)}{3G_2''(\omega_2)^{3/2}},\\
J_{2,2}(\omega_2, \omega_2) &= \frac{11G_2'''(\omega_2)^2}{9G_2''(\omega_2)^{5/2}} - \frac{3G_2^{\text{iv}}(\omega_2)}{4G_2''(\omega_2)^{3/2}}.
\end{aligned}$$

2.2. *Inequalities.*

LEMMA 2.1. *Define* $A_n = \{\overline{Y}_n \in G_1'(\Theta_1)\}, B_n = \{\overline{T}_{n2} \in \Theta_2\}$ *and* $A_nB_n = A_n \cap B_n$. *For any* $x > 0, m = 2,3,\dots,\theta \in \Theta$, *we have* $\mathbb{P}_\theta(\sup_{n \geq m} I(\widehat{\theta}_n, \theta)1_{A_nB_n} \geq x) \leq [e^{m\gamma(\theta_1)} + 4]\exp\{-(m-1)x/2\}$.

PROOF. See the Appendix.

LEMMA 2.2. *For any* $x \geq 1, \theta \in \Theta$ *and a positive integer* $n, \mathbb{P}_\theta(\max_{k \leq n} \|\mathbf{Z}_k\|1_{A_kB_k} \geq x) \leq [e^{m\gamma(\theta_1)} + 4](1 + \log_2 n)\exp\{-x^2/4\}$, *where* $\log_2$ *denotes logarithm to the base* 2.

PROOF. Let M_n be the smallest integer which exceeds $\log_2 n$. We have

$$\mathbb{P}_\theta\Big(\max_{k \leq n} \|\mathbf{Z}_k\|1_{A_kB_k} \geq x\Big) \leq \sum_{m=1}^{M_n} \mathbb{P}_\theta\{A_kB_k \text{ and } I(\widehat{\theta}_k, \theta) > x^2/2^{m+1}, \exists k \geq 2^{m-1}\}.$$

The result follows from Lemma 2.1. □

For any fixed prior density ξ, let $\mathcal{W}^\xi$ denote the class of all functions $W: \Theta \times \Theta \to \mathbb{R}$ for which

$$\mathbb{E}^\xi\left[\sup_{n \geq m} |W(\widehat{\theta}_n, \theta)|1_{A_nB_n}\right] < \infty \tag{10}$$

for some $m = 2, 3, \ldots$. Observe that $\mathcal{W}^{\xi}$ is a linear space which contains all constant functions.

LEMMA 2.3. *Assume that ξ has compact support and w is a real function on Θ such that $\mathbb{E}^{\xi}|\omega(\boldsymbol{\theta})| < \infty$. For any fixed m_0, define $W(\omega;\boldsymbol{\theta}) = w(\boldsymbol{\theta})\exp\{m_0 I(\omega, \boldsymbol{\theta})\}$, $\omega, \boldsymbol{\theta} \in \Theta$. Then $W \in \mathcal{W}^{\xi}$.*

PROOF. The conclusion is true if $m_0 \le 0$. For $m_0 > 0$, choose an integer $m \ge 4(m_0 + 1)$,

$$
\begin{aligned}
&\mathbb{E}^{\xi}\left[\sup_{n\ge m} W(\hat{\theta}_n, \theta)1_{A_n B_n}\right] \\
&\quad \le \mathbb{E}^{\xi}\left[|\omega(\theta)| \int_0^{\infty} \mathbb{P}_{\boldsymbol{\theta}}\left(\sup_{n\ge m} \exp\{m_0 I(\widehat{\theta}_n, \theta)\} > s\right) ds\right] \\
&\quad \le \mathbb{E}^{\xi}\left(|\omega(\theta)|\left\{\int_{10}^{\infty} \mathbb{P}_{\theta}\left[\sup_{n\ge m} I(\widehat{\theta}_n, \theta) > \frac{\log(s)}{m_0}\right] ds + 10\right\}\right) \\
&\quad \le \mathbb{E}^{\xi}\left(|w(\theta)|\left\{\left[e^{m\gamma(\theta_1)} + 4\right] \int_{10}^{\infty} \exp\left[-\frac{m\log(s)}{2m_0}\right] ds + 10\right\}\right) \\
&\quad \le \mathbb{E}^{\xi}\left(|w(\theta)|\left\{\frac{e^{m\gamma(\theta_1)} + 4}{10} + 10\right\}\right).
\end{aligned}
$$

Here the third inequality follows from Lemma 2.1. Since $\gamma(\theta_1)$ is bounded on the compact support of ξ, the right side is bounded by $C\mathbb{E}^{\xi}|w(\theta)|$, for some constant C, which is finite by assumption. □

2.3. *Stein's identity.* Let $\mathcal{H}$ denote the collection of measurable functions $h: \mathbb{R}^2 \to \mathbb{R}$ of polynomial growth; let $\mathcal{H}_p = \{h \in \mathcal{H}: |h(\mathbf{z})|/(1 + |z_1|^p + |z_2|^p) \le 1, \forall\, \mathbf{z} \in \mathbb{R}^2\}$, $\widetilde{\mathcal{H}}_p = \{h: h/c \in \mathcal{H}_p \text{ for some } c > 0\}$. Thus $\mathcal{H} = \cup_{p\ge 0}\widetilde{\mathcal{H}}_p$. Let $\|\mathbf{z}\|$ be the Euclidean norm of a vector $\mathbf{z}$. If $|h(\mathbf{z})| \le c(1 + \|\mathbf{z}\|^p)$ for some $0 < c < \infty$, then $h \in \widetilde{\mathcal{H}}_p$. Let Φ_1 and Φ denote one- and two-dimensional standard normal distributions, respectively. Let ϕ_1 and ϕ be the corresponding densities and let

$$
\begin{aligned}
\Phi h &= \int_{\mathbb{R}^2} h(\mathbf{z})\phi(\mathbf{z})\,d\mathbf{z}, \qquad \Phi_1^h(z_1) = \int_{-\infty}^{\infty} h(z_1, z_2)\phi_1(z_2)\,dz_2, \\
\mathbf{V}^h(\mathbf{z}) &= \left(V_1^h(\mathbf{z}), V_2^h(\mathbf{z})\right) \\
&= \begin{pmatrix} [\phi_1(z_1)]^{-1} \displaystyle\int_{z_1}^{\infty} \{\Phi_1^h(y_1) - \Phi h\}\phi_1(y_1)\,dy_1 \\ [\phi_1(z_2)]^{-1} \displaystyle\int_{z_2}^{\infty} \{h(z_1, y_2) - \Phi_1^h(z_1)\}\phi_1(y_2)\,dy_2 \end{pmatrix}^{\tau}.
\end{aligned}
$$

For example, if $h(\mathbf{z}) = z_1, z_2$ or $z_1^2 + z_2^2$, $\mathbf{V}^h(\mathbf{z}) = (1, 0), (0, 1)$ or $\mathbf{z}$, respectively. Note that V_1^h, as a function on $\mathbb{R}^2$, is a constant in its last variable. The transformation from h to $\mathbf{V}^h$ is a linear operator from $\mathcal{H}_p$ into $\mathcal{H}_p \times \mathcal{H}_p$ and satisfies

$$\Phi\mathbf{V}^h = \int_{\mathbb{R}^2} \mathbf{z}h(\mathbf{z})\Phi(d\mathbf{z}) \quad \text{and} \quad \Phi(\mathbf{V} \circ \mathbf{V}^h) = \frac{1}{2}\int_{\mathbb{R}^2} \begin{pmatrix} z_1^2 - 1 & 0 \\ 2z_1 z_2 & z_2^2 - 1 \end{pmatrix} h(\mathbf{z})\Phi(d\mathbf{z}). \tag{11}$$

Here and in the following, $\mathbf{V} \circ \mathbf{V}$ is the composition of $\mathbf{V}$ with itself.

LEMMA 2.4. *Given any nonnegative integer p, there is a constant C_p so that for any $h \in \mathcal{H}_p$,*

$$\|\mathbf{V}^h(\mathbf{z})\| \le C_p\left(1 + |z_1|^p + |z_2|^p\right) \qquad \forall\, \mathbf{z} \in \mathbb{R}^2.$$

PROOF. The assertion for $p = 0$ is proved by Stein (1986). The proof for $p \ge 1$ is similar. □

STEIN'S IDENTITY. *Let Γ be a finite signed measure of the form $d\Gamma = f\,d\Phi$, where f is in integrable function with respect to Φ on $\mathbb{R}^2$ and is absolutely continuous on every compact subset of $\mathbb{R}^2$. Denote $\Gamma h = \int_{\mathbb{R}^2} h\,d\Gamma$, when $h \in \mathcal{H}$ and the integral exists, and define $\nabla f = (f_{10}(\mathbf{z}), f_{01}(\mathbf{z}))$, where*

$$f_{jk}(\mathbf{z}) = \frac{\partial^{j+k}}{\partial z_1^j \partial z_2^k} f(\mathbf{z}).$$

If for any nonnegative integer p,

$$\int_{\mathbb{R}^2} \left(|z_1|^p + |z_2|^p\right)\left[|f_{10}(\mathbf{z})| + |f_{10}(\mathbf{z})|\right]\Phi(d\mathbf{z}) < \infty, \tag{12}$$

then

$$\Gamma h - \Gamma 1 \cdot \Phi h = \int_{\mathbb{R}^2} \mathbf{V}^h [\nabla f]^\tau \Phi(d\mathbf{z}) \qquad \forall\, h \in \mathcal{H}_p.$$

PROOF. For any $h \in \mathcal{H}_p$, Lemma 2.4 implies that $V_j^h \in \mathcal{H}_p$, and $(|z_1|^p + |z_2|^p)[|f_{10}(\mathbf{z})| + |f_{10}(\mathbf{z})|]$ is integrable with respect to Φ from (12). The rest of the proof follows from integration by parts and is similar to Stein (1986) or Woodroofe (1992). □

3. Asymptotic expansions.

3.1. *Basic consequences.* Let Ξ_0 denote the class of all absolutely continuous densities with compact support, and let AC be the class of all absolutely continuous functions on Θ. For $\alpha > 1$, define two subclasses of Ξ_0 by

$$\Xi_1^\alpha = \left\{ \xi \in \Xi_0 \colon \mathbb{E}^\xi \left\{ \left| \frac{\xi_{10}}{\xi} \right|^\alpha + \left| \frac{\xi_{01}}{\xi} \right|^\alpha \right\} < \infty \right\} \tag{13}$$

and

$$\Xi_2^\alpha = \left\{ \xi \in \Xi_1 \colon \xi_{10}, \xi_{01} \in \mathrm{AC}, \mathbb{E}^\xi \left\{ \left| \frac{\xi_{20}}{\xi} \right|^\alpha + \left| \frac{\xi_{11}}{\xi} \right|^\alpha + \left| \frac{\xi_{02}}{\xi} \right|^\alpha \right\} < \infty \right\}, \tag{14}$$

where $\xi_{jk}(\boldsymbol{\theta}) = \partial^{j+k}\xi(\boldsymbol{\theta})/\partial\theta_1^j\,\partial\theta_1^k$. Let $\tilde{\xi}(\boldsymbol{\theta}) = \xi(\boldsymbol{\theta})/\sqrt{-\theta_1}$, so $\tilde{\xi}_{0k}/\tilde{\xi} = \xi_{0k}/\xi$ for $k = 0, 1, \ldots$. If $\xi \in \Xi_1$, then $\mathbb{E}^\xi\{|\tilde{\xi}_{10}/\xi|^\alpha + |\tilde{\xi}_{01}/\xi|^\alpha\} < \infty$, and if $\xi \in \Xi_2$, then $\mathbb{E}^\xi\{|\tilde{\xi}_{20}/\xi|^\alpha + |\tilde{\xi}_{11}/\xi|^\alpha + |\tilde{\xi}_{02}/\xi|^\alpha\} < \infty$. For $\boldsymbol{\omega} \neq \boldsymbol{\theta} \in \boldsymbol{\Theta}$, define

$$\mathbf{K}_1^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) = \left(K_{1,1}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}), K_{1,2}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta})\right) = \frac{1}{\tilde{\xi}}\left(\partial\tilde{\xi}J_1/\partial\theta_1, \partial\tilde{\xi}J_2/\partial\theta_2\right),$$

$$\mathbf{K}_2^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) = \begin{pmatrix} K_{2,11}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) & K_{2,12}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) \\ K_{2,12}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) & K_{2,22}^\xi(\boldsymbol{\omega}, \boldsymbol{\theta}) \end{pmatrix} = \frac{1}{\tilde{\xi}} \begin{pmatrix} \partial^2\tilde{\xi}J_1/\partial\theta_1^2 & \partial^2\tilde{\xi}J_1J_2/\partial\theta_1\,\partial\theta_2 \\ \partial^2\tilde{\xi}J_1J_2/\partial\theta_1\,\partial\theta_2 & \partial^2\tilde{\xi}J_2/\partial\theta_2^2 \end{pmatrix}.$$

REMARK 3.1. The value of $\mathbf{K}_j$ on the diagonal may be obtained from l'Hospital's rule as

$$K_{1,1}^\xi(\boldsymbol{\theta}, \boldsymbol{\theta}) = \frac{\tilde{\xi}_{10}}{\tilde{\xi}} \frac{1}{[G_1''(\theta_1)]^{1/2}} - \frac{G_1'''(\theta_1)}{3[G_1''(\theta_1)]^{3/2}},$$

$$K_{1,2}^\xi(\boldsymbol{\theta}, \boldsymbol{\theta}) = \frac{\tilde{\xi}_{01}}{\tilde{\xi}} \frac{1}{[G_2''(\theta_2)]^{1/2}} - \frac{2G_2'''(\theta_2)}{3[G_2''(\theta_2)]^{3/2}},$$

$$K_{2,11}^\xi(\boldsymbol{\theta}, \boldsymbol{\theta}) = \frac{\tilde{\xi}_{20}}{\tilde{\xi}} \frac{1}{G_1''} - \frac{\tilde{\xi}_{10}}{\tilde{\xi}} \frac{G_1'''}{[G_1'']^2} + \frac{5[G_1''']^2}{12[G_1'']^3} - \frac{G_1^{\mathrm{iv}}}{4[G_1'']^2},$$

$$\begin{aligned} K_{2,12}^\xi(\boldsymbol{\theta}, \boldsymbol{\theta}) = {} & \frac{\tilde{\xi}_{11}}{\tilde{\xi}} \frac{1}{[G_1''G_2'']^{1/2}} - \frac{\tilde{\xi}_{10}}{\tilde{\xi}} \frac{2G_2'''}{3\{G_1''[G_2'']^3\}^{1/2}} \\ & - \frac{\tilde{\xi}_{01}}{\tilde{\xi}} \frac{G_1'''}{3\{[G_1'']^3 G_2''\}^{1/2}} + \frac{2G_1'''G_2'''}{9[G_1''G_2'']^{3/2}}, \end{aligned}$$

$$K_{2,22}^\xi(\boldsymbol{\theta}, \boldsymbol{\theta}) = \frac{\tilde{\xi}_{02}}{\tilde{\xi}} \frac{1}{G_2''} - \frac{\tilde{\xi}_{01}}{\tilde{\xi}} \frac{2G_2'''}{[G_2'']^2} + \frac{5[G_2''']^2}{3[G_2'']^3} - \frac{3G_2^{\mathrm{iv}}}{4[G_2'']^2}.$$

Recalling (7), the posterior distribution of $\mathbf{Z}_n$ given $X_1, \ldots, X_n$, say Γ_n, is of the form $d\Gamma_n = f_n\,d\Phi$, where $f_n(\mathbf{z}) = c_n\xi(\boldsymbol{\theta})J(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})1_{S_n}(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^2$, for some $0 < c_n = c_n(X_1, \ldots, X_n) < \infty$.

LEMMA 3.1. *Assume that* A_nB_n *occurs. If* $\xi \in \Xi^\alpha$, $\alpha > 1$, *then* f_n *is absolutely continuous with*

$$\nabla f_n(\mathbf{z}) = \frac{1}{\sqrt{n}} f_n(\mathbf{z}) \mathbf{K}_1^\xi(\widehat{\theta}_n, \theta), \qquad \mathbf{z} \in \mathbb{R}^2, \tag{15}$$

and if $\xi \in \Xi_2^\alpha$, $\alpha > 1$, *then* f_n *is twice continuously differentiable for which*

$$\nabla^2 f_n(\mathbf{z}) = \frac{1}{n} f_n(\mathbf{z}) \mathbf{K}_2^\xi(\widehat{\theta}_n, \theta), \qquad \mathbf{z} \in \mathbb{R}^2. \tag{16}$$

PROOF. This follows from the assumptions immediately. □

LEMMA 3.2. *If* $\xi \in \Xi_1^\alpha$, $\alpha > 1$, *then* $|K_{1,1}^\xi|^\alpha + |K_{1,2}^\xi|^\alpha \in \mathcal{W}_\xi$; *if* $\xi \in \Xi_2^\alpha$, $\alpha > 1$, *then* $|K_{1,11}^\xi|^\alpha + |K_{1,12}^\xi|^\alpha + |K_{1,22}^\xi|^\alpha \in \mathcal{W}_\xi$.

PROOF. This follows from the compactness of the support of ξ and the smoothness of G_1 and G_2. □

THEOREM 3.1. *If* $\xi \in \Xi_1^\alpha$ *for some* $\alpha > 1$, *there is an* $m \geq 2$ *such that, for all* $h \in \mathcal{H}$,

$$\mathbb{E}_n^\xi\{h(\mathbf{Z}_n)\} = \Phi h + \frac{1}{\sqrt{n}} \mathbb{E}_n^\xi\left\{\mathbf{K}_1^\xi(\widehat{\theta}_n, \theta)[\mathbf{V}^h(\mathbf{Z}_n)]^\tau\right\}, \tag{17}$$

a.s. ($\mathbb{P}^\xi$) *on* A_nB_n *for all* $n \geq m$. *If* $\xi \in \Xi_2^\alpha$ *for some* $\alpha > 1$, *there is an* $m \geq 2$ *such that*

$$\begin{aligned}\mathbb{E}_n^\xi\{h(\mathbf{Z}_n)\} &= \Phi h + \frac{1}{\sqrt{n}} (\Phi \mathbf{V}^h) \mathbb{E}_n^\xi\{\mathbf{K}_1^\xi(\widehat{\theta}_n, \theta)\}^\tau \\ &\quad + \frac{1}{n} \operatorname{tr}\left\{\mathbb{E}_n^\xi\left\{\mathbf{K}_2^\xi(\widehat{\theta}_n, \theta)[\mathbf{V} \circ \mathbf{V}^h(\mathbf{Z}_n)]\right\}\right\}\end{aligned} \tag{18}$$

for all $h \in \mathcal{H}$, *a.s.* ($\mathbb{P}^\xi$) *on* A_nB_n *for all* $n \geq m$.

PROOF. If $\xi \in \Xi_1^\alpha$, then $|K_{1,1}^\xi|^\alpha + |K_{1,2}^\xi|^\alpha \in \mathcal{W}^\xi$ from Lemma 3.2. Let $m \geq 2$ satisfy (10). If $n \geq m$ and A_nB_n occurs, then for all $0 \leq p < \infty$,

$$\sqrt{n} \int_{\mathbb{R}^2} \left(|z_1|^p + |z_2|^p\right) \left|\frac{\partial}{\partial z_1} f_n(\mathbf{z})\right| \Phi(d\mathbf{z}) = \mathbb{E}_n^\xi\left\{\left(|Z_{n1}|^p + |Z_{n2}|^p\right) \left|K_{1,1}^\xi(\widehat{\theta}_n, \theta)\right|\right\},$$

which is finite w.p.1 by Hölder's inequality. It is also finite w.p.1 if $(\partial/\partial z_1)$ $f_n(\mathbf{z})$ is replaced by $(\partial/\partial z_2) f_n(\mathbf{z})$. Then (17) follows from the Stein identity and Lemma 3.1. The proof of (18) is similar. □

3.2. *Second-order expansions.* In order to obtain a second-order expansion as a sequence of stopping times goes to ∞, we need the following lemma.

LEMMA 3.3. *Assume that* $\xi \in \Xi_1^{p+1}$, $1 \le p < \infty$. *Then*

$$\left\{ \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \sqrt{n} \left| \mathbb{E}_n^{\xi}\left[h(\mathbf{Z}_n)\right] - \Phi h \right| 1_{A_n B_n} \right\}_{n \ge m}$$

and

$$\left\{ \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \left| \mathbb{E}_n^{\xi}\left[h(\mathbf{Z}_n)\right] \right| 1_{A_n B_n} \right\}_{n \ge m}$$

are both uniformly integrable.

PROOF. See the Appendix.

For the following, $\{t_a : a \ge 1\}$ is an increasing family of stopping times with respect to the σ-field generated by $\{X_1, \dots, X_n\}$. For clarity, we will let $t = t_a$.

THEOREM 3.2. *Assume that* $\xi \in \Xi_1^{p+1}$ *for some* $p \ge 1$. *If the stopping time* t *satisfies*

$$\mathbb{P}_{\boldsymbol{\theta}}\{A_t B_t\} = 1 \qquad \forall\, a \ge 1,\ \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{19}$$

$$a/t \to \rho^2(\boldsymbol{\theta}) \quad \textit{in } \mathbb{P}^{\xi}\textit{-probability as } a \to \infty, \tag{20}$$

and

$$\lim_{a \to \infty} a^q \mathbb{P}^{\xi}\{t \le \eta a\} = 0 \qquad \textit{for some } \eta > 0,\ q > 1/2. \tag{21}$$

Then

$$\lim_{a \to \infty} \mathbb{E}^{\xi}\left\{ \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \left| \sqrt{a}\left(\mathbb{E}_t^{\xi}\left[h(\mathbf{Z}_t)\right] - \Phi h - \frac{1}{\sqrt{a}}(\Phi \mathbf{V}^h)\mathbb{E}_t^{\xi}\left[\rho(\boldsymbol{\theta})\mathbf{K}_1^{\xi}(\boldsymbol{\theta}, \boldsymbol{\theta})\right]^{\tau}\right) \right| \right\} = 0.$$

PROOF. Note that $\mathbb{E}^{\xi}|h(\mathbf{Z}_t)|^p < \infty$ for any $h \in \mathcal{H}_p$ and $a \ge 1$ by Lemma 3.3. Let $R_a(\xi, h)$ denote the quantity within the absolute value. Then

$$\mathbb{E}^{\xi}\left\{ 1_{\{t \le \eta a\}} \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} |R_a(\xi, h)| \right\} \to 0 \quad \text{as } a \to \infty$$

by Hölder's inequality, Lemma 2.2 and assumptions (19)–(21). For the integral over $\{t > \eta a\}$, we may write

$$\operatorname*{ess\,sup}_{h \in \mathcal{H}_p} |R_a(\xi, h)| 1_{\{t > \eta a\}} \le \sum_{j=1}^{3} R_{j,a},$$

where

$$R_{1,a} = \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \mathbb{E}_t^{\xi} \left| \left[\mathbf{K}_1^{\xi}(\widehat{\theta}_t, \theta) - \mathbb{E}_t^{\xi}\left[\mathbf{K}_1^{\xi}(\theta,\theta)\right] \left[\mathbf{V}^h(\mathbf{Z}_t)\right]^{\tau} \right| 1_{\{t > \eta a\}} / \sqrt{\eta},$$

$$R_{2,a} = \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \mathbb{E}_t^{\xi} \left| \mathbb{E}_t^{\xi}\left[\mathbf{K}_1^{\xi}(\theta,\theta)\right] \mathbb{E}_t^{\xi}\left[\mathbf{V}^h(\mathbf{Z}_t) - \Phi \mathbf{V}^h\right]^{\tau} \right| 1_{\{t > \eta a\}} / \sqrt{\eta}$$

and

$$R_{3,a} = \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \left| \Phi \mathbf{V}^h \mathbb{E}_t^{\xi}\left[\sqrt{a/t} - \rho(\theta)\right]\left[\mathbf{K}_1^{\xi}(\theta,\theta)\right]^{\tau} \right| 1_{\{t > \eta a\}} / \sqrt{\eta}.$$

It follows from Hölder's inequality and Lemma 2.4 that $\mathbb{E}^{\xi}(R_{1,a}) \to 0$. It is clear that $R_{2,a} \to 0$, w.p.1. By Lemma 3.3 and Hölder's inequality, $R_{2,a}$, $a \geq 1$, is uniformly integrable. Therefore, $\mathbb{E}^{\xi}(R_{2,a}) \to 0$. From (20) and (21), we have $|\sqrt{a/t} - \rho(\theta)| 1_{\{t > \eta a\}} \leq |\sqrt{a/t} - \rho(\theta)| \to 0$ in probability and $|\sqrt{a/t} - \rho(\theta)| 1_{\{t > \eta a\}} \leq 1/\sqrt{\eta} + \rho(\theta)$, which is essentially bounded. Therefore,

$$\mathbb{E}^{\xi}(R_{3,a}) \leq C(p) \sum_{j=1}^{2} \mathbb{E}^{\xi}\left(|K_{1,j}(\theta,\theta)[\sqrt{a/t} - \rho(\theta)]| 1_{\{t > \eta a\}} \right) \to 0$$

by the dominated convergence theorem. □

COROLLARY 3.1. *Assume that* $\xi \in \Xi_0^{p+1}$ *for some* $p \geq 1$, (19) *and* (20) *hold and for every compact* $\Theta_0 \subset \Theta$, *there is an* $\eta = \eta(\Theta_0)$ *for which*

$$\lim_{a \to \infty} a^q \int_{\Theta_0} \mathbb{P}_{\theta}(t_a \leq \eta)\, d\theta = 0 \quad \textit{for some } q > 1/2. \tag{22}$$

Assume further that ρ *is absolutely continuous on all compact subsets of* Θ *and define*

$$\kappa_{1,1}(\theta) = \frac{G_1'''(\theta_1)}{6\left[G_1''(\theta_1)\right]^{3/2}} - \frac{\frac{\partial}{\partial \theta_1}\left[\rho(\theta)\sqrt{-\theta_1}\right]}{\left[G_1''(\theta_1)\right]^{1/2} \rho(\theta)\sqrt{-\theta_1}},$$

$$\kappa_{1,2}(\theta) = -\frac{G_2'''(\theta_2)}{6\left[G_2''(\theta_2)\right]^{3/2}} - \frac{\frac{\partial}{\partial \theta_2}\rho(\theta)}{\left[G_2''(\theta_2)\right]^{1/2} \rho(\theta)}$$

and $\widetilde{\kappa}_{1,j}^{\xi} = \mathbb{E}^{\xi}[\rho(\theta)\kappa_{1,j}(\theta)]$, $j = 1, 2$. *Then we have*

(a)

$$\lim_{a \to \infty} \mathbb{E}^{\xi}\left\{ \operatorname*{ess\,sup}_{h \in \mathcal{H}_p} \sqrt{a} \left| \mathbb{E}_t^{\xi}[h(\mathbf{Z}_t)] - \Phi h - \frac{1}{\sqrt{a}}(\Phi \mathbf{V}^h)\left(\widetilde{\kappa}_{1,1}^{\xi}, \widetilde{\kappa}_{1,2}^{\xi}\right)^{\tau} \right| \right\} = 0,$$

(b)

$$\lim_{a \to \infty} \sup_{h \in \mathcal{H}_p} \sqrt{a} \left| \mathbb{E}^{\xi}[h(\mathbf{Z}_t)] - \Phi h - \frac{1}{\sqrt{a}}(\Phi \mathbf{V}^h)\left(\widetilde{\kappa}_{1,1}^{\xi}, \widetilde{\kappa}_{1,2}^{\xi}\right)^{\tau} \right| = 0,$$

(c)

$$\lim_{a\to\infty}\sup_{h\in\mathcal{H}_p}\left|\mathbb{E}^{\xi}\sqrt{a}\left[F_{a,\theta}h-\Phi_{a,\theta}^{(1)}h\right]\right|=0,$$

where $F_{a,\theta}(\mathbf{z})=\mathbb{P}_\theta(\mathbf{Z}_t\le\mathbf{z})$ *and* $\Phi_{a,\theta}^{(1)}h=\Phi h+(\rho(\theta)/\sqrt{a})(\kappa_{1,1}(\theta),\kappa_{1,2}(\theta))(\Phi\mathbf{V}^h)^\tau$.

PROOF. Note that condition (22) implies (21) for every ξ with compact support. Assertion (c) is just a restatement of (b). For assertions (a) and (b), in view of Theorem 3.2, it is enough to prove that $\mathbb{E}^{\xi}[\rho(\theta)K_{1,j}^{\xi}(\theta,\theta)]=\tilde{\kappa}_{1,j}^{\xi}$, $j=1,2$. The technique, based on integration by parts, is due to Stein (1985) and has been used by Woodroofe (1992) and Ghosh and Mukerjee (1992). □

3.3. *Higher-order expansions.*

THEOREM 3.3. *Assume that* $\xi\in\Xi_2^{p+1}$, $p\ge 2$, *and* (19), (20) *and* (22) *hold for some* $q>1$. *Then*

$$\mathbb{E}^{\xi}\left\{\operatorname*{ess\,sup}_{h\in\mathcal{H}_p}|S_a(\xi,h)|\right\}\to 0\quad \textit{as } a\to\infty,$$

where

$$S_a(\xi,h)=a\biggl(\mathbb{E}_t^{\xi}\left[h(\mathbf{Z}_t)\right]-\Phi h-\frac{1}{\sqrt{t}}(\Phi\mathbf{V}^h)\mathbb{E}_t^{\xi}\left[\mathbf{K}_1^{\xi}(\theta,\theta)\right]^\tau \\ -\frac{1}{a}\operatorname{tr}\left\{\Phi(\mathbf{V}\circ\mathbf{V}^h)\mathbb{E}_t^{\xi}\left[\rho^2(\theta)\mathbf{K}_2^{\xi}(\theta,\theta)\right]\right\}\biggr).$$

PROOF. A proof similar to that of Theorem 3.2 shows

$$\mathbb{E}^{\xi}\left\{1_{\{t\le\eta a\}}\operatorname*{ess\,sup}_{h\in\mathcal{H}_p}|S_a(\xi,h)|\right\}\to 0.$$

To estimate the integral over $\{t>\eta a\}$, we need the decomposition $S_a(\xi,h)=R_a^*+R_a^{**}$, where $R_a^*=(a/t)\operatorname{tr}\mathbb{E}_t^{\xi}\{\mathbf{K}_2^{\xi}(\hat{\theta}_t,\theta)\mathbf{V}\circ\mathbf{V}^h(\mathbf{Z}_t)\}-\sqrt{a/t}\operatorname{tr}\Phi(\mathbf{V}\circ\mathbf{V}^h)\mathbb{E}_t^{\xi}[\rho(\theta)\mathbf{K}_2^{\xi}(\theta,\theta)]$ and $R_a^{**}=(a/\sqrt{t})\mathbb{E}_t^{\xi}[\mathbf{K}_1^{\xi}(\hat{\theta}_t,\theta)-\mathbf{K}_1^{\xi}(\theta,\theta)](\Phi\mathbf{V}^h)^\tau$. The analysis of R_a^* over $\{t>\eta a\}$ is similar to R_a in Theorem 3.2. For R_a^{**}, we use the facts that $K_{1,1}^{\xi}(\omega,\theta)$ does not depend on ω_1 and that ξ has a compact support. □

COROLLARY 3.2. *Assume that the conditions of Theorem* 3.3. *hold and that* $\rho(\cdot)$ *is twice absolutely continuous with*

$$\lim_{a\to\infty}\sqrt{a}\int_{\Theta_0}\left|\mathbb{E}_\theta\left(\sqrt{\frac{a}{t}}\right)-\rho(\theta)\right|d\theta=0$$

for every compact set Θ_0 in Θ. Define

$$\kappa_{2,11}(\boldsymbol{\theta}) = \frac{\frac{\partial^2}{\partial\theta_1^2}\left[\rho^2\sqrt{-\theta_1}\right]}{G_1''\rho^2\sqrt{-\theta_1}} - \frac{G_1'''}{G_1''^2}\frac{\frac{\partial}{\partial\theta_1}\left[\rho^2\sqrt{-\theta_1}\right]}{\rho^2\sqrt{-\theta_1}} + \frac{5G_1'''^2}{12G_1''^3} - \frac{G_1^{\mathrm{iv}}}{4G_1''^2},$$

$$\kappa_{2,12}(\boldsymbol{\theta}) = \frac{\frac{\partial^2}{\partial\theta_1\partial\theta_2}\left[\rho^2\sqrt{-\theta_1}\right]}{\sqrt{G_1''G_2''}\rho^2\sqrt{-\theta_1}} + \frac{G_2'''\frac{\partial}{\partial\theta_1}\left[\rho^2\sqrt{-\theta_1}\right]}{6\sqrt{G_1''G_2''^3}\rho^2\sqrt{-\theta_1}} - \frac{G_1'''\frac{\partial}{\partial\theta_2}\left[\rho^2\sqrt{-\theta_1}\right]}{6\sqrt{G_1''^3G_2''}\rho^2\sqrt{-\theta_1}} - \frac{19G_1'''G_2'''}{36[G_1''G_2'']^{3/2}},$$

$$\kappa_{2,22}(\boldsymbol{\theta}) = \frac{\frac{\partial^2}{\partial\theta_2^2}\rho^2}{G_2''\rho^2} - \frac{G_2'''^2}{3G_2''^3} + \frac{G_2^{\mathrm{iv}}}{4G_2''^2}$$

and

$$\tilde{\kappa}^{\xi}_{2,ij} = \mathbb{E}^{\xi}\left[\rho^2(\boldsymbol{\theta})\kappa_{2,ij}(\boldsymbol{\theta})\right].$$

Then

(a)

$$\lim_{a\to\infty}\sup_{h\in\mathcal{H}_p} a\left|\mathbb{E}^{\xi}\left[h(\mathbf{Z}_t)\right] - \Phi h - \frac{1}{\sqrt{a}}\Phi\mathbf{V}^h\begin{pmatrix}\tilde{\kappa}^{\xi}_{1,1}\\ \tilde{\kappa}^{\xi}_{1,2}\end{pmatrix} - \frac{1}{a}\operatorname{tr}\left\{\Phi(\mathbf{V}\circ\mathbf{V}^h)\begin{pmatrix}\tilde{\kappa}^{\xi}_{2,11} & \tilde{\kappa}^{\xi}_{2,12}\\ \tilde{\kappa}^{\xi}_{2,12} & \tilde{\kappa}^{\xi}_{2,22}\end{pmatrix}\right\}\right| = 0,$$

(b)

$$\lim_{a\to\infty}\sup_{h\in\mathcal{H}_p}\left|\mathbb{E}^{\xi}a\left[F_{a,\theta}h - \Phi^{(2)}_{a,\theta}h\right]\right| = 0,$$

where $F_{a,\theta}(\mathbf{z}) = \mathbb{P}_{\theta}(\mathbf{Z}_t \le \mathbf{z})$ and

$$\Phi^{(2)}_{a,\theta}h = \Phi h + \frac{\rho(\boldsymbol{\theta})}{\sqrt{a}}\left(\kappa_{1,1}(\boldsymbol{\theta}),\kappa_{1,2}(\boldsymbol{\theta})\right)(\Phi\mathbf{V}^h)^{\tau} + \frac{\rho(\boldsymbol{\theta})^2}{a}\operatorname{tr}\left\{\begin{pmatrix}\kappa_{2,11}(\boldsymbol{\theta}) & \kappa_{2,12}(\boldsymbol{\theta})\\ \kappa_{2,12}(\boldsymbol{\theta}) & \kappa_{2,22}(\boldsymbol{\theta})\end{pmatrix}\Phi(\mathbf{V}\circ\mathbf{V}^h)\right\}.$$

PROOF. For part (a), it is enough to show

$$\sup_{h\in\mathcal{H}_p}\left|\sqrt{a}\left\{\mathbb{E}^{\xi}\sqrt{a/t}\,\mathbb{E}^{\xi}_t\left[\mathbf{K}^{\xi}_1(\boldsymbol{\theta},\boldsymbol{\theta})\right]^{\tau} - \Phi\mathbf{V}^h\left(\tilde{\kappa}^{\xi}_{1,1},\tilde{\kappa}^{\xi}_{1,2}\right)^{\tau}\right\}\right| \to 0,$$

which follows from the additional assumption. Part (b) is a restatement of part (a). □

3.4. *Asymptotic expansions by rescaling.* One of the significant features of transformation (6) is that the correction terms in the asymptotic expression may be described by rescaling.

THEOREM 3.4. *Assume that the conditions of Corollary* 3.1 *hold. For* $\mathbf{z} \in \mathbb{R}^2$, $\boldsymbol{\theta} \in \Theta$ *and* $a \geq 1$, *define*

$$\Phi^{(3)}_{a,\theta}(\mathbf{z}) = \Phi\left(\mathbf{z} - \frac{\rho(\boldsymbol{\theta})}{\sqrt{a}}\left[\kappa_{1,1}(\boldsymbol{\theta}), \kappa_{1,2}(\boldsymbol{\theta})\right]\right).$$

Then

$$\lim_{a \to \infty} \sup_{h \in \mathcal{H}_p} \sqrt{a}\left|\mathbb{E}^{\xi}\left[F_{a,\theta}h - \Phi^{(3)}_{a,\theta}h\right]\right| = 0.$$

PROOF. The proof is similar to the proof of the next theorem. □

As mentioned in the Introduction, the symmetric functions of $\mathbf{Z}_t$ play an important role in many cases. For example, the log-likelihood ratio test statistic, $\Lambda_t(\boldsymbol{\theta}) = l_t(\widehat{\boldsymbol{\theta}}_t) - l_t(\boldsymbol{\theta}) = tI(\widehat{\boldsymbol{\theta}}_t, \boldsymbol{\theta}) = \frac{1}{2}(Z_{t1}^2 + Z_{t2}^2)$, can be used to test the null hypothesis H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In general, a subset of symmetric functions in $\mathcal{H}_p$ is defined by

$$\mathcal{H}_p^* = \{h \in \mathcal{H}_p\colon h(z_1, z_2) = h(-z_1, z_2) \\ = h(z_1, -z_2) = h(-z_1, -z_2) \quad \forall\, (z_1, z_2) \in \mathbb{R}^2\}.$$

Note that for any $h \in \mathcal{H}_p^*$, $\Phi\mathbf{V}^h = 0$, Φh is the approximation for both $\mathbb{E}_t^{\xi}[h(\mathbf{Z}_t)]$ and $\mathbb{E}^{\xi}[h(\mathbf{Z}_t)]$ of the second-order accuracy for any $\xi \in \Xi_1^{p+1}$, $h \in \mathcal{H}_p^*$. Furthermore, it is also possible to have a high-order rescaled expansion.

THEOREM 3.5. *Assume that the conditions of Corollary* 3.2 *hold. In addition, suppose* $\kappa_{2,11}(\boldsymbol{\theta}) \geq 0, j = 1, 2$. *Define*

$$\Phi^{(4)}_{a,\theta}(\mathbf{z}) = \Phi\left(\mathbf{z} - \frac{\rho(\boldsymbol{\theta})}{\sqrt{a}}\left(\sqrt{\kappa_{2,11}(\boldsymbol{\theta})}, \sqrt{\kappa_{2,22}(\boldsymbol{\theta})}\right)\right)$$

for $\mathbf{z} \in \mathbb{R}^2, \boldsymbol{\theta} \in \Theta$ *and* $a \geq 1$. *Then* $\lim_{a \to \infty} \sup_{h \in \mathcal{H}_p^*} |a\mathbb{E}^{\xi}[F_{a,\theta}h - \Phi^{(4)}_{a,\theta}h]| = 0$.

PROOF. First, the absolute value is no more than $S_1 + S_2$, where

$$S_1 = a\left|\mathbb{E}^{\xi}\left[h(\mathbf{Z}_t)\right] - \Phi h - \frac{1}{2a}\sum_{j=1}^{2}\tilde{\kappa}^{\xi}_{2,jj}\int_{\mathbb{R}^2}\left(z_j^2 - 1\right)h(\mathbf{z})\Phi(d\mathbf{z})\right|$$

and

$$S_2 = a\left|\mathbb{E}^{\xi}\Phi^{(4)}_{a,\theta}h - \Phi h - \frac{1}{2a}\sum_{j=1}^{2}\rho^2(\boldsymbol{\theta})\kappa_{2,jj}(\boldsymbol{\theta})\int_{\mathbb{R}^2}\left(z_j^2 - 1\right)h(\mathbf{z})\Phi(d\mathbf{z})\right|.$$

Here we use the facts that $\int_{\mathbb{R}^2} \mathbf{z}h(\mathbf{z})\Phi(d\mathbf{z}) = (0,0)$ and $\int_{\mathbb{R}^2} z_1 z_2 h(\mathbf{z})\Phi(d\mathbf{z}) = 0$ for any $h \in \mathcal{H}_p^*$. It follows from (11) and the argument of Corollary 3.2 that $S_1 \to 0$ as $a \to \infty$ uniformly for $h \in \mathcal{H}_p^*$. By a simple calculation, the integrand in S_2 is bounded by

$$\begin{aligned}(23)\qquad &\frac{1}{\sqrt{a}} \int_{\mathbb{R}^2} \sum_{k=0}^{3} \binom{3}{k} \rho^3(\theta)\left[\kappa_{2,11}(\theta)\right]^{k/2}\left[\kappa_{2,22}(\theta)\right]^{(3-k)/2} \\ &\times \left| \frac{\partial^3}{\partial z_1^k \, \partial z_2^{3-k}} \phi(\mathbf{z}^*) \right| \left(1 + |z_1|^p + |z_2|^p\right) d\mathbf{z},\end{aligned}$$

where

$$\|\mathbf{z}^* - \mathbf{z}\| \le \frac{\rho(\theta)}{\sqrt{a}} \sqrt{\kappa_{2,11}(\theta) + \kappa_{2,22}(\theta)}.$$

But (23) is independent of $h \in \mathcal{H}_p^*$, approaches 0 as $a \to \infty$, for each θ, and is bounded by a constant multiple of $1 + Q(\theta)$. Here $Q(\theta)$ depends only on $1 + \|\nabla\rho(\theta)\| + \|\nabla^2\rho(\theta)\|$, which is bounded on the compact support of ξ. So $S_2 \to 0$ uniformly in $h \in \mathcal{H}_p^*$ by the dominated convergence theorem. □

4. Applications.

4.1. *Matching posterior and frequentist expansions.* It would be desirable to choose a prior ξ such that the frequentist expansion $\mathbb{E}_\theta[h(\mathbf{Z}_t)] = \Phi_2 h + O(a^{-1})$, $\forall h \in \mathcal{H}_p$, holds uniformly on compact sets of θ. In general, no such a prior is available. [See Ghosh and Mukerjee (1992).] However, we are able to find a prior so that the posterior and frequentist expansions for any $h \in \mathcal{H}$ match up to the second order, that is,

$$(24)\qquad \sqrt{a}\left\{\mathbb{E}_t^\xi\left[(\mathbf{Z}_t)\right] - \mathbb{E}_\theta\left[h(\mathbf{Z}_t)\right]\right\} \to 0 \quad \text{as } a \to \infty,$$

in P_θ-probability. In fact, it follows from Theorem 3.2 that

$$\mathbb{E}_t^\xi\left[h(\mathbf{Z}_t)\right] = \Phi h + \frac{1}{\sqrt{a}}\Phi_2 \mathbf{V}^h \mathbb{E}_t^\xi\left[\rho(\theta)\mathbf{K}^\xi(\theta)\right]^\tau + O\left(\frac{1}{a}\right)$$

for any $h \in \mathcal{H}$. Furthermore, if the frequentist expectation of $h(\mathbf{Z}_t)$ ($\mathbb{E}_\theta[h(\mathbf{Z}_t)]$) can be uniformly expressed up to the second order $O(a^{-1})$, then

$$\mathbb{E}_\theta\left[h(\mathbf{Z}_t)\right] = \Phi h + \frac{1}{\sqrt{a}}\Phi_2 \mathbf{V}^h \rho(\theta)\left(\kappa_{1,1}(\theta), \kappa_{1,2}(\theta)\right)^\tau + O\left(\frac{1}{a}\right).$$

This leads to the following two partial differential equations for ξ:

$$(25)\qquad K_{1,j}^\xi(\theta) - \kappa_{1,j}(\theta) \equiv 0, \qquad j = 1, 2.$$

THEOREM 4.1. *The optimal noninformative prior of $\boldsymbol{\theta}$ matching the posterior and frequentist expansions of $h(\mathbf{Z})\,\forall\, h \in \mathcal{H}$ has the form*

$$\xi(\boldsymbol{\theta}) \propto \frac{1}{\rho(\boldsymbol{\theta})}\sqrt{G_1''(\theta_1)G_2''(\theta_2)}.$$

PROOF. Note that (25) is equivalent to

$$\frac{\frac{\partial}{\partial\theta_k}\tilde{\xi}}{\tilde{\xi}} = \frac{G_k'''(\theta_k)}{2G_k''(\theta_k)} - \frac{\frac{\partial}{\partial\theta_k}\left[\rho(\boldsymbol{\theta})\sqrt{-\theta_1}\,\right]}{\rho(\boldsymbol{\theta})\sqrt{-\theta_k}}, \qquad k = 1,2.$$

Apart from a constant, the solution to the equations is

$$\log\bigl(\xi(\boldsymbol{\theta})/\sqrt{-\theta_1}\bigr) = \tfrac{1}{2}\log\bigl(G_1''(\theta_1)G_2''(\theta_2)\bigr) - \log\bigl(\rho(\boldsymbol{\theta})\sqrt{-\theta_1}\bigr).$$

This completes the proof of the theorem. □

REMARK 4.1. Consider special stopping times t for which $\rho(\boldsymbol{\theta}) \equiv$ constant. A typical example is a fixed sample size problem. Theorem 4.1 implies that the priors matching posterior and frequentist expansions of $h(\mathbf{Z}_n)$ are proportional to $\sqrt{G_1''(\theta_1)G_2''(\theta_2)}$. In a forthcoming work, we will show that this prior agrees with Tibshirani's and Berger and Bernardo's reference priors for θ_1 in the presence of nuisance parameter θ_2, or θ_2 in the presence of nuisance parameter θ_1. One can show that the Jeffreys prior for the case is $\sqrt{-\theta_1 G_1''(\theta_1)G_2''(\theta_2)}$, different from the optimal noninformative prior.

4.2. *A generalized sequential test.* One application of Theorem 3.5 is to approximate the sampling distribution of the log-likelihood ratio test statistic. For example, let $X_1, X_2, \ldots$ be iid. $N(\mu, \sigma^2)$, where both $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$ are unknown, and define

$$t = t_a = \min\left(b_2 a, \inf\left\{n \geq b_1 a\colon \sum_{i=1}^{n} X_i^2 - n - n\log(\hat{\sigma}_n^2) > 2a\right\}\right),$$

where $0 < b_1 < b_2 < \infty$ are two prespecified numbers, $\hat{\sigma}_n^2 = (1/n)\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$ and $\overline{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then

$$2\Lambda_t = \|\mathbf{Z}_t\|^2 = \frac{1}{\sigma^2}\sum_{i=1}^{t}(X_i - \mu)^2 + t\log(\sigma^2) - t - t\log(\hat{\sigma}_t^2) \quad \text{on } \sum_{j=1}^{t}(X_j - \overline{X}_t)^2 > 0.$$

From Table 1, $\boldsymbol{\theta} = (-1/2\sigma^2, \mu)$. Theorem 8.3 of Woodroofe (1982) implies that

$$\frac{a}{t_a} \to \begin{cases} b_2, & \text{if } \rho^2(\boldsymbol{\theta}) < 1/b_2, \\ \rho^2(\boldsymbol{\theta}), & \text{if } 1/b_2 < \rho^2(\boldsymbol{\theta}) < 1/b_1, \\ b_1, & \text{if } \rho^2(\boldsymbol{\theta}) > 1/b_1, \end{cases}$$

in $\mathbb{P}_{\theta}$-probability, as $a \to \infty$, where $\rho^2(\boldsymbol{\theta}) = I(\boldsymbol{\theta}, (-0.5, 0)) = \{(\mu^2+1)/\sigma^2 + \log(\sigma^2) -1\}/2$. Let $h_u(\mathbf{z}) = I_{(\|\mathbf{z}\|^2 \le 2u)}$ for $\mathbf{z} \in \mathbb{R}^2$. Then Theorem 3.5 suggested the approximation $\mathbb{P}_{\theta}(\Lambda_t \le u) \approx \Phi_{a,\theta}^{(4)} h_u \approx \delta(u; a, \boldsymbol{\theta})$, where

$$\delta(u; a, \boldsymbol{\theta}) = \Phi h_u + \frac{1}{2a} \sum_{j=1}^{2} \rho^2(\boldsymbol{\theta}) \kappa_{2,jj}(\boldsymbol{\theta}) \int_{\mathbb{R}^2} (z_j^2 - 1) h_u(\mathbf{z}) \Phi(d\mathbf{z}).$$

It is easy to verify that

$$\rho^2(\boldsymbol{\theta})\kappa_{2,11}(\boldsymbol{\theta}) = \frac{\mu^2+1}{\sigma^2} + \frac{11}{6}\rho^2(\boldsymbol{\theta}), \qquad \rho^2(\boldsymbol{\theta})\kappa_{2,22}(\boldsymbol{\theta}) = -\theta_1 = \frac{1}{2\sigma^2}.$$

Also, $\Phi h_u = \mathbb{P}(\chi_2^2 \le 2u) = 1 - e^{-u}$ and $\int_{\mathbb{R}^2} (z_j^2 - 1) h_u(\mathbf{z}) \Phi(d\mathbf{z}) = -ue^{-u}$ for $j = 1, 2$ and $u \ge 0$, where χ_2^2 is a chi-square random variable with 2 degrees of freedom. Therefore,

$$\delta(u; a, \boldsymbol{\theta}) = 1 - e^{-u} - \frac{ue^{-u}}{2a}\left\{\frac{\mu^2+1.5}{\sigma^2} + \frac{11}{6}\rho^2(\boldsymbol{\theta})\right\}, \qquad u \ge 0. \tag{26}$$

Figure 1 shows Monte Carlo estimates of $\mathbb{P}_{\theta}(\Lambda_t \le u)$ for $a = 8$, $b_1 = 0.5$, $b_2 = 50$, against u in [0, 10] for various combinations $(\mu, \sigma^2) = (0.5, 1.25), (0.5, 0.8), (1, 1.25)$ and $(1, 0.8)$, together with directed χ_2^2 approximation $\Phi h_u = 1 - e^{-u}, u \ge 0$, and the corrected approximation $\delta(u; a, \theta)$ given by (26). The χ_2^2-approximation seriously overestimates the probabilities. The corrected term in (26) is always negative and the corrected approximations are closer to Monte Carlo estimates in all cases as shown in Figure 1.

4.3. *Inferences about* θ_1. In this section, we will apply Theorem 3.5 to the problem of sequential tests and confidence intervals for θ_1 in the presence of the nuisance parameter θ_2. Suppose that we want to test H_0: $\theta_1 = \theta_{10}$ under a stopping rule t. The log-likelihood test statistic after obtaining the first t observations is then $\Lambda_t = l(\widehat{\theta}_{t1}, \widehat{\theta}_{t2}) - l(\theta_1, \widehat{\theta}_{t2})$. It turns out that $2\Lambda_t = Z_{t1}^2$. Let $h_u(\mathbf{z}) = I_{(z_{t1}^2 \le 2u)}$. If the stopping time $t = t_a$, $a > 0$, satisfies the conditions of Theorem 3.5, the following approximation can be used:

$$\mathbb{P}_{\theta}(\Lambda_t \le u) = \Phi h_u - \frac{ue^{-u}}{2a}\rho^2(\boldsymbol{\theta})\kappa_{2,11}(\boldsymbol{\theta}) + o(a^{-1}).$$

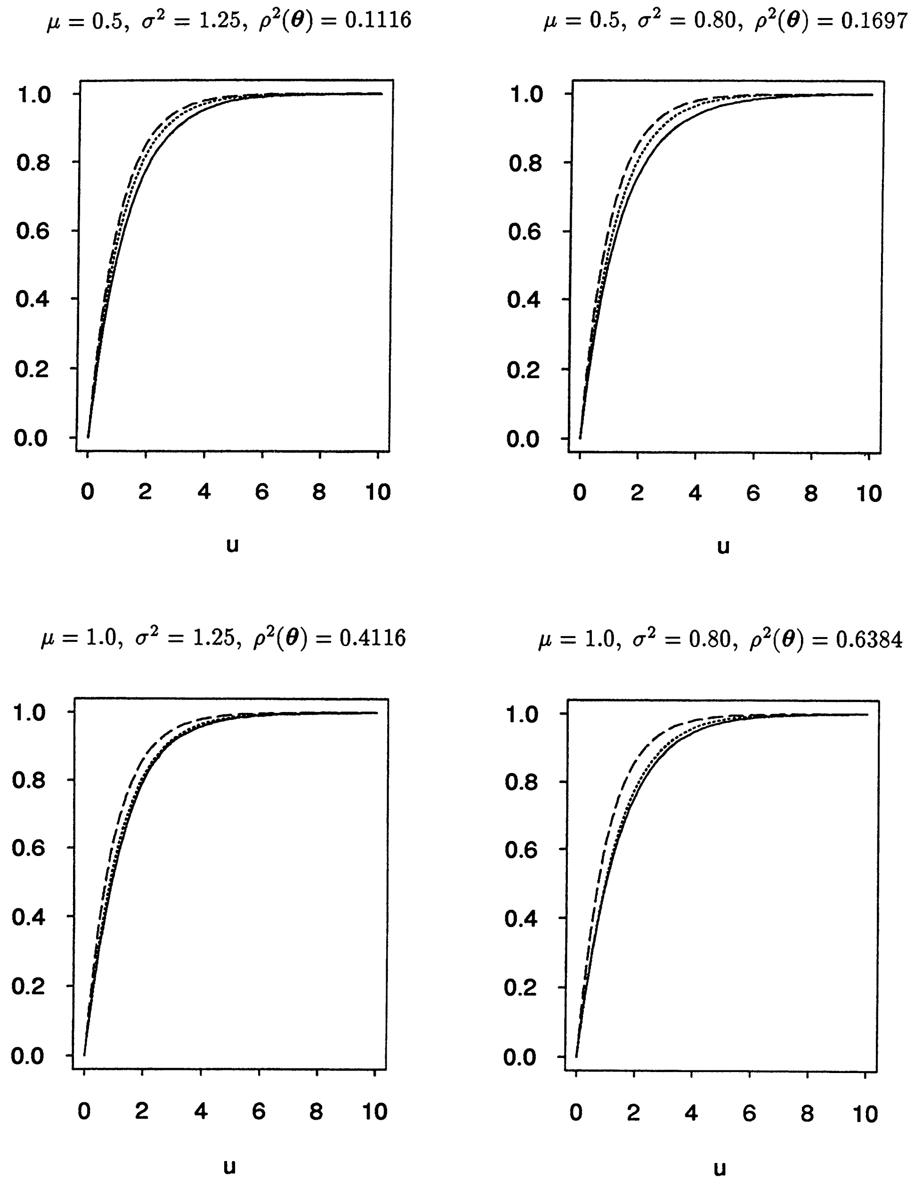
Here we use the facts that $\int_{\mathbb{R}^2}(z_1^2 - 1)h_u(\mathbf{z})\Phi(d\mathbf{z}) = -2\sqrt{u/\pi}e^{-u}$ and $\int_{\mathbb{R}^2}(z_2^2 - 1) \times h_u(\mathbf{z})\Phi(d\mathbf{z}) = 0$.

Bose and Boukai (1993) considered the stopping time

$$t_a = \inf\{n \ge 2: |\widehat{\theta}_{n1}| > a^2/n^2\} = \inf\left\{n \ge 2: Y_n < nG_1'(-a^2/n^2)\right\}$$

for $a > 0$ and proved that

$$\lim_{a \to \infty} \frac{a}{t} = \rho^2(\boldsymbol{\theta}) = \sqrt{-\theta_1}, \quad \text{a.s. in } \mathbb{P}_{\theta}\text{-probability}.$$



——— Simulations of $\mathbb{P}_{\boldsymbol{\theta}}(\Lambda_t \le u)$ based on 40,000 replications

– – – – $\Phi h_u = 1 - e^{-u}$

·········· $\delta(u, a, \boldsymbol{\theta}) = 1 - e^{-u} - \frac{ue^{-u}}{2a}\left\{\frac{\mu^2+1.5}{\sigma^2} + \frac{11}{6}\rho^2(\boldsymbol{\theta})\right\}$

FIG. 1. *Simulations of* $\mathbb{P}_{\boldsymbol{\theta}}(\Lambda_t \le u)$ *and approximations for* $a = 8$.

After observing $(X_1, \ldots, X_t)$, we may use the following confidence interval for θ_1:

$$\{\theta_1: l(\widehat{\theta}_{t1}, \widehat{\theta}_{t2}) - l(\theta_1, \widehat{\theta}_{t2}) \le u\} = \{\theta_1: tI_1(\widehat{\theta}_{t1}, \theta_1) \le u\} = \{\theta_1: Z_{t1}^2 \le 2u\},$$

where $I_1(\cdot, \cdot)$ is given by (4). The confidence coefficient for this interval is then

$$\mathbb{P}(\chi_1^2 \le u) - \frac{ue^{-u}}{2a}\sqrt{-\theta_1}\left\{\frac{G_1'''}{\theta_1 G_1''^2} + \frac{5G_1'''^2}{12G_1''^3} - \frac{G_1^{\text{iv}}}{4G_1''^2}\right\} + o(a^{-1}).$$

Here χ_1^2 is a χ^2 random variable with 1 degree of freedom. In particular, for the normal distribution $N(\mu, \sigma^2)$, it can be shown that

$$\frac{G_1'''}{\theta_1 G_1''^2} + \frac{5G_1'''^2}{12G_1''^3} - \frac{G_1^{\text{iv}}}{4G_1''^2} = -\frac{11}{3}.$$

APPENDIX

In order to prove Lemma 2.1, we need the following lemma.

LEMMA A.1. *For all $\boldsymbol{\theta} \in \Theta$ and $m = 2, 3, \ldots,$*

$$\mathbb{P}_{\boldsymbol{\theta}}(A_n \text{ and } \widehat{\theta}_{n1} \ge \omega_1, \exists\, n \ge m) \le \exp\{-mI_1(\omega_1, \theta_1)\} \quad \text{if } \omega_1 > \theta_1,$$

$$\mathbb{P}_{\boldsymbol{\theta}}(A_n \text{ and } \widehat{\theta}_{n1} \le \omega_1, \exists\, n \ge m) \le \exp\{-mI_1(\omega_1, \theta_1) + G_1(m\theta_1) - G_1(m\omega_1)\} \quad \text{if } \omega_1 < \theta_1,$$

$$\mathbb{P}_{\boldsymbol{\theta}}(B_n \text{ and } \widehat{\theta}_{n2} \ge \omega_2, \exists\, n \ge m) \le \exp\{m\theta_1 I_2(\omega_2, \theta_2)\} \quad \text{if } \omega_2 > \theta_2,$$

$$\mathbb{P}_{\boldsymbol{\theta}}(B_n \text{ and } \widehat{\theta}_{n2} \le \omega_2, \exists\, n \ge m) \le \exp\{m\theta_1 I_2(\omega_2, \theta_2)\} \quad \text{if } \omega_2 < \theta_2.$$

PROOF. First, $\mathbb{P}_{\boldsymbol{\theta}}(A_n \text{ and } \widehat{\theta}_{n1} \ge \omega_1, \exists\, n \ge m) = \mathbb{P}_{\boldsymbol{\theta}}(\sup_{n \ge m} \overline{Y}_n \ge G_1'(\omega_1))$, where $\overline{Y}_n = (1/n)\Sigma_{i=1}^n V_i - \xi_n \equiv \overline{V}_n - \xi_n$, $V_i = U_1(X_i) - G_2'(\mu_2)U_2(X_i) + [G_2'(\mu_2)\mu_2 - G_2(\mu_2)]$ and $\xi_n = (T_{n2}/n - \mu_2)^2 G_2''(\tau_n)/2$. The first assertion follows from the facts that $\mathbb{E}_{\boldsymbol{\theta}} \exp(sV_1) = \exp\{G_1(s + \theta_1) - G_1(\theta_1)\}$ and that $\exp\{m(\omega_1 - \theta_1)\overline{V}_n\}$, $n \ge m$, is a reverse submartingale. From Bar-Lev and Reiser (1982),

$$\mathbb{E}_{\boldsymbol{\theta}} \exp(sY_n) = \exp\left\{n\left[G_1(s + \theta_1) - G_1(\theta_1)\right] - \left[G_1\left(n(s + \theta_1)\right) - G_1(n\theta_1)\right]\right\}.$$

The second assertion follows from the fact that $\exp\{m(\omega_1 - \theta_1)\overline{V}_n\}$, $n \ge m$, is a reverse submartingale. The third assertion follows from the facts that $\{\exp(-m\theta_1[G_2'(\omega_2) - G_2'(\theta_2)]T_{n2}/n),\ n \ge m\}$ is a reverse submartingale and that $\mathbb{E}_{\boldsymbol{\theta}}\{\exp[sU_2(X_1)]\} = \exp\{\theta_1[G_2(\widetilde{\theta}_2) - G_2(\theta_2) + \theta_2 G_2'(\theta_2) - \widetilde{\theta}_2 G_2'(\widetilde{\theta}_2)]\}$, where $\widetilde{\theta}_2$ is the unique solution of the equation $G_2'(\widetilde{\theta}_2) = G_2'(\theta_2) - s/\theta_1$. The last assertion follows similarly. □

PROOF OF LEMMA 2.1. For any fixed $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta$,

$$\begin{aligned}\mathbb{P}_\theta\Big(\sup_{n\ge m} I(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}) 1_{A_n B_n} \ge x\Big) &\le \mathbb{P}_\theta\Big(\sup_{n\ge m} I_1(\widehat{\theta}_{n1}, \theta_1) 1_{A_n} \ge \frac{x}{2}\Big)\\ &\quad + \mathbb{P}_\theta\Big(\sup_{n\ge m} I_2(\widehat{\theta}_{n2}, \theta_2) 1_{B_n} \ge \frac{x}{-2\theta_1}\Big).\end{aligned} \tag{27}$$

If G_1 is bounded, the assertion holds. Otherwise,

$$\begin{aligned}-I_1(\omega_1, \theta_1) + G_1(m\theta_1) - G_1(m\omega_1) &= (m-1)\big[\theta_1 G_1'(\theta_1^*) - \omega_1 G_1'(\omega_1^*)\big]\\ &\quad + \theta_1 G_1'(\omega_1) - \omega_1 G_1'(\omega_1)\end{aligned} \tag{28}$$

for some $\theta_1^* \in [m\theta_1, \theta_1]$ and $\omega_1^* \in [m\omega_1, \omega_1]$. Since G_1' is positive and strictly increasing, (28) is no larger than $-m\omega_1 G_1'(\omega_1) \le m\gamma(\theta_1)$ for all $\omega_1 < \theta_1$ by Assumption C. Therefore, (27) is bounded by $e^{m\gamma(\theta_1)}\exp\{-(m-1)x/2\} + 3\exp\{-mx/2\}$. The proof follows from Lemma A.1 and the monotonicity of I_j. □

PROOF OF LEMMA 3.3. Let m be in (10) for $W = |K_{1,1}^\xi|^{p+1} + |K_{1,2}^\xi|^{p+1}$. If $h \in \mathcal{H}_p$, then for $n \ge m$,

$$\begin{aligned}\sqrt{n}\big|\mathbb{E}_n^\xi[h(\mathbf{Z}_n)] - \Phi h\big| &\le \sum_{j=1}^{2}\Big\{\mathbb{E}_n^\xi\big|K_{1,j}^\xi(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})\big|^{p+1}\Big\}^{1/(p+1)}\\ &\quad \times \Big\{\mathbb{E}_n^\xi\Big[\big|V_j^h(\mathbf{Z}_n)\big|^{(p+1)/p}\Big]\Big\}^{p/(p+1)},\end{aligned}$$

a.e. on $A_n B_n$ by Theorem 3.1 and Hölder's inequality. Let $C_j(p)$ denote positive constants depending only on p. From Lemma 2.4, $\mathbb{E}_n^\xi[|V_j^h(\mathbf{Z}_n)|^{(p+1)/p}] \le C_2(p)\mathbb{E}_n^\xi[1+|Z_{n1}|^{p+1}+|Z_{n2}|^{p+1}]$ a.e. on $A_n B_n$. Let $g_p(\mathbf{z}) = 1+|z_1|^p$. Then $V_1^{g_{p+1}}(\mathbf{z}) \le C_3(p) g_p(\mathbf{z})$ and $V_2^{g_{p+1}}(\mathbf{z}) = 0$. Therefore,

$$\begin{aligned}\sqrt{n}\big|\mathbb{E}_n^\xi[g_{p+1}(\mathbf{Z}_n)] - \Phi g_{p+1}\big| &\le C_4(p)\Big\{\mathbb{E}_n^\xi\Big[\big|K_{1,1}^\xi(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})\big|^{p+1}\Big]\Big\}^{1/(p+1)}\\ &\quad \times \Big\{\mathbb{E}_n^\xi\big[|g_{p+1}(\mathbf{Z}_n)|^{(p+1)/p}\big]\Big\}^{p/(p+1)}.\end{aligned}$$

Observe that if $0 < b,\ c < \infty$, $1 \le p < \infty$ and $0 \le x \le b + cx^{p/(p+1)}$, then $x \le pb + c^{p+1}$. This inequality shows that

$$\mathbb{E}_n^\xi|g_{p+1}(\mathbf{Z}_n)| \le p\Phi g_{p+1} + \big(C_4(p)/\sqrt{n}\big)^{p+1}\mathbb{E}_n^\xi\big[|K_{1,1}^\xi(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})|^{p+1}\big].$$

Similarly, let $\widetilde{g}_p(z) = 1 + |z_2|^p$. Then

$$\mathbb{E}_n^\xi|\widetilde{g}_{p+1}(\mathbf{Z}_n)| \le p\Phi\widetilde{g}_{p+1} + \big(C_5(p)/\sqrt{n}\big)^{p+1}\mathbb{E}_n^\xi\Big[\big|K_{1,2}^\xi(\widehat{\boldsymbol{\theta}}_n, \boldsymbol{\theta})\big|^{p+1}\Big].$$

Now the first assertion follows immediately. The second assertion can be proved similarly. □

Acknowledgments. Most of the work was done when I was visiting the University of Michigan. I would like to express my deep appreciation and sincerest thanks to Professor Michael B. Woodroofe for suggesting the problem and his invaluable help and contributions during the course of this work. I am indebted to Professors James O. Berger, Robert W. Keener and Steven P. Lalley for their useful comments and discussions. I would also like to acknowledge Professor Paul L. Speckman for his helpful suggestions and proofreading of this paper. I also thank the editors and a referee for insightful comments and suggestions which resulted in a significant improvement over a previous version of the paper.

REFERENCES

ALVO, M. (1977). Bayesian sequential estimation. *Ann. Statist.* **5** 955–968.

BAR-LEV, S. K. and REISER, B. (1982). An exponential subfamily which admits UMPU tests based on a single test statistic. *Ann. Statist.* **10** 979–989.

BARNDORFF-NIELSEN, O. E. (1978). *Information and the Exponential Families in Statistical Theory*. Wiley, New York.

BARNDORFF-NIELSEN, O. E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A* **160** 268–282.

BERGER, J. O. and BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84** 200–207.

BERGER, J. O. and BERNARDO, J. M. (1992). On the development of the reference prior method (with discussion). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 35–60. Oxford Univ. Press.

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 113–147.

BICKEL, P. J. and GHOSH J. K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction—a Bayesian argument. *Ann. Statist.* **18** 1070–1090.

BICKEL, P. J., GÖTZE, F. and VAN ZWET, W. R. (1985). A simple analysis of third order efficiency of estimates. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. Le Cam and R. A. Olshen, eds.) **2** 749–768. Wadsworth, Belmont, CA.

BOSE, A. and BOUKAI, B. (1993). Sequential estimation results for a two-parameter exponential family of distributions. *Ann. Statist.* **21** 484–502.

BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–346.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families*. IMS, Hayward, CA.

COX, D. R. and REID, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49** 1–39.

EFRON, B. (1985). Bootstrap confidence intervals for a class of parametric problems. *Biometrika* **72** 45–58.

GHOSH, J. K. and MUKERJEE, R. (1992). Noninformative priors (with discussion). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 195–210. Oxford Univ. Press.

GHOSH, J. K., SINHA, B. K. and JOSHI, S. M. (1982). Expansions for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics 3* (S. S. Gupta and J. O. Berger, eds.) **1** 403–456. Academic, New York.

HU, I. (1988). Repeated significance tests for exponential families. *Ann. Statist.* **16** 1643–1666.

JEFFREYS, H. (1936, 1961). *Theory of Probability*. Oxford Univ. Press.

JOHNSON, R. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–864.

LALLEY, S. P. (1983). Repeated likelihood ratio tests for curved exponential families. *Z. Wahrsch. Verw. Gebiete* **62** 293–321.

LAPLACE, P. S. (1847). *Oeuvers* **7.** Imprimerie Royale, Paris.

LAWLEY, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika* **43** 295–303.

PEERS, H. W. (1965). On confidence sets and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. Ser. B* **27** 9–16.

STEIN, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics* (R. Zielinski, ed.) 485–514. PWN-Polish Scientific Publishers, Warsaw.

STEIN, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, CA.

TIBSHIRANI, R. (1989). Noninformative priors for one parameter of many. *Biometrika* **76** 604–608.

WELCH, B. N. and PEERS, B. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **35** 318–329.

WOODROOFE, M. (1978). Large deviations of the likelihood ratio statistic with applications to sequential testing. *Ann. Statist.* **6** 72–84.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

WOODROOFE, M. (1986). Very weak expansions for sequential confidence intervals. *Ann. Statist.* **14** 1049–1067.

WOODROOFE, M. (1989). Very weak expansions for sequentially designed experiments: linear models. *Ann. Statist.* **17** 1087–1102.

WOODROOFE, M. (1992). Integrable expansions for posterior distributions for one parameter exponenital families. *Statist. Sinica* **2** 91–111.

YE, K. (1993). Reference priors when the stopping rule depends on the parameter of interest. *J. Amer. Statist. Assoc.* **88** 360–363.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI
COLUMBIA, MISSOURI 65211