

ASYMPTOTIC NORMALITY WITH SMALL RELATIVE ERRORS OF POSTERIOR PROBABILITIES OF HALF-SPACES

BY R. M. DUDLEY¹ AND D. HAUGHTON²

Massachusetts Institute of Technology and Bentley College

Let Θ be a parameter space included in a finite-dimensional Euclidean space and let A be a half-space. Suppose that the maximum likelihood estimate θ_n of θ is not in A (otherwise, replace A by its complement) and let Δ be the maximum log likelihood (at θ_n) minus the maximum log likelihood over the boundary ∂A . It is shown that under some conditions, uniformly over all half-spaces A , either the posterior probability of A is asymptotic to $\Phi(-\sqrt{2\Delta})$ where Φ is the standard normal distribution function, or both the posterior probability and its approximant go to 0 exponentially in n . Sharper approximations depending on the prior are also defined.

1. Introduction. Some examples of half-spaces of interest in parameter spaces are, in a clinical trial of a treatment versus placebo, the half-spaces where the treatment is (a) helpful or (b) harmful. Thus one may not only want to test the hypothesis that the treatment (c) makes no difference, but to assign posterior probabilities to (a), (b) and (c), under conditions as unrestrictive as possible on the choice of prior probabilities [e.g., Dudley and Haughton (2001)]. More generally, we have in mind applications to model selection as in the BIC criterion of Schwarz (1978) and its extensions [Poskitt (1987); Haughton (1988)], specifically to one-sided models and multiple data sets [Dudley and Haughton (1997)].

We will describe our results, omitting some details and conditions until Section 2. Let Φ be the standard normal distribution function and ϕ its density. Let Θ be an open subset of a Euclidean space \mathbf{R}^d and let A vary over half-spaces in \mathbf{R}^d . Let A have boundary hyperplane ∂A . Let $\mathcal{P} = \{P_\phi : \phi \in \Theta\}$ be a family of laws having densities $f(x, \phi)$ with respect to some μ . Given n observations X_1, \dots, X_n i.i.d. with a law P not necessarily in \mathcal{P} , let $\hat{\phi} = \hat{\phi}_n$ be the maximum likelihood estimate of $\phi \in \Theta$. If $\hat{\phi}_n$ is in A , replace A by its complement. Let $\tilde{\phi} = \tilde{\phi}_n$ be the maximum likelihood estimate of ϕ in ∂A . Let $LL(\phi) := LL_n(\phi)$ denote the log likelihood $\sum_{j=1}^n \log f(X_j, \phi)$ and let $\Delta := \Delta_n := LL(\hat{\phi}) - LL(\tilde{\phi})$. Then 2Δ is a likelihood ratio statistic. We will show in Theorem 1, under conditions to be given in Section 2, that for any prior law π on Θ having a continuous, strictly positive density, for $A = A_n$ depending on $n \rightarrow \infty$, if ∂A_n approaches, however slowly,

Received December 1999; revised October 2001.

¹Supported in part by NSF Grants DMS-97-04603 and DMS-01-03821.

²Supported in part by an NSF grant.

AMS 2000 subject classifications. Primary 62F15; secondary 60F99, 62F05.

Key words and phrases. Bernstein–von Mises theorem, gamma tail probabilities, intermediate deviations, Jeffreys prior, Mills' ratio.

the true or pseudo-true (as defined in Section 2, Assumption A₅) parameter point ϕ_0 , then the posterior probability $\pi_n(A_n)$ is asymptotic to $\Phi(-\sqrt{2\Delta_n})$. Or if ∂A_n remains bounded away from ϕ_0 , then both $\pi_n(A_n)$ and $\Phi(-\sqrt{2\Delta_n})$ approach 0 exponentially in n .

The classical normal approximation to posterior laws π_n was by normal laws with their means at the MLE (maximum likelihood estimate) $\hat{\phi}$ and their covariance matrix equal to the inverse of the observed Fisher information matrix, for example, in one dimension [Walker (1969) and Johnson (1970)] and for multidimensional ϕ [Hipp and Michel (1976), Chen (1985) and Pauler, Wakefield and Kass (1999)]. Thus the posterior law becomes close to a standard normal law after a linear transformation. On the other hand, Le Cam (1953) proved asymptotic normality of posteriors without assuming second derivatives of the likelihood functions, so that the Fisher information matrix may not be defined.

Here, as mentioned, we will instead use nonlinear transformations by way of likelihood ratio statistics, as did, for example, Bickel and Ghosh (1990), who gave a set of precise assumptions, some of which (e.g., that the prior has compact support) are stronger than needed for proving our theorem. (Bickel and Ghosh's Bayesian results may be viewed as technical lemmas for proving frequentist theorems.) Likelihood ratio statistics give much smaller relative errors of the normal approximation in the tails (see Section 10), and the approximation is also rather easy to compute, requiring only the values of the likelihood at the unrestricted MLE $\hat{\phi}$ and at the MLE $\tilde{\phi}$ over the boundary ∂A of the half-space A , and the knowledge of whether $\hat{\phi} \in A$. We do assume second derivatives of the likelihood functions and use them in showing that relative errors are small. Some sharpened approximations depending on the prior, defined in Section 3, are still more accurate in examples (Section 10). After Theorem 1 we give a corollary showing that under our assumptions the same conclusions hold for the first sharpened Laplace approximation as for the simpler likelihood root approximation; stronger assumptions (e.g., beyond mere continuity of the prior density) would apparently be needed to prove the order of improvement that can be given by the sharpened approximations.

With notation as in the abstract, let $B_\Delta := \Phi(\sqrt{2\Delta})/\Phi(-\sqrt{2\Delta})$ if the MLE $\hat{\theta}_n$ is in the half-space A , or $\Phi(-\sqrt{2\Delta})/\Phi(\sqrt{2\Delta})$ otherwise. Then our result as stated in the abstract, or in more detail in Theorem 1, shows that B_Δ gives an asymptotic approximation to the Bayes factor $\pi_n(A)/(1 - \pi_n(A))$ for a half-space relative to its complementary half-space, for a wide class of proper priors, unless $\pi_n(A)$ or $1 - \pi_n(A)$ is exponentially small. Yet B_Δ does not depend on the prior. Thus B_Δ might be viewed as a "default Bayes factor" if no prior is given. Berger and Mortera (1999) treat alternative possibilities for such factors when priors may be improper.

Let $a_n \ll b_n$ mean that $a_n = o(b_n)$ as $n \rightarrow \infty$. For a fixed half-space A , let d_n be the Euclidean distance from $\hat{\phi}_n$ to ∂A . Let us say we have a case of Laplace deviations if $d_n = O(n^{-1/2})$, of intermediate deviations if $n^{-1/2} \ll d_n \ll 1$, and

large deviations if d_n converges to a positive constant d . Let us further say that deviations are positive if $\hat{\phi}_n \notin A$ and negative if $\hat{\phi}_n \in A$.

Asymptotic normality of the posterior distribution holds classically for Laplace deviations [e.g., Hipp and Michel (1976)] and, as the present paper shows, for positive intermediate (and Laplace) deviations in terms of the likelihood root. In the analysis literature, asymptotics for integrals of functions $e^{-ah(x)}$ over half-lines with boundary point near the minimum of h as $a \rightarrow +\infty$ had been treated in terms of normal distribution functions by Bleistein [(1966), Section 5], Wong (1973), Skinner (1980) (in the multidimensional case), Temme (1982) and Wong [(1989), Section VII.3]. In Temme (1982), $h = h(x, a)$ can depend on a , as for posterior distributions with $a = n$, and asymptotic expansions giving small relative errors are proved under some (analyticity) assumptions.

In a statistical problem one presumably does not know in advance (before observations are taken) whether the deviation will be Laplace, intermediate or large. If one uses the asymptotic normality approximation we propose then one has small relative errors for posterior probabilities of Laplace or intermediate deviations.

Pauler, Wakefield and Kass (1999) also indicate some cases where posterior probabilities of subsets A which may be half-spaces, or orthants, are useful. They give in their Section 4 a “boundary Laplace approximation” which involves the normal probability of A based on the observed Fisher information matrix at $\hat{\phi}$ times the prior density at the maximum $\tilde{\phi}$ of the likelihood on A . Pauler, Wakefield and Kass (1999) prove under a set of assumptions that as $n \rightarrow \infty$ the (relative) error of their approximation approaches 0. Their approximation has the advantages of applying when A is not necessarily a half-space and the data are not necessarily i.i.d. On the other hand, their assumption C1, that the approximating normal probabilities are bounded away from 0, is quite restrictive relative to our situation.

We will treat in Section 9 a multivariate normal location family and present in Section 10 some numerical results for beta distributions.

2. Statements of assumptions and the main theorem. First we give some notation. Let $P_n := \frac{1}{n}(\delta_{X_1} + \dots + \delta_{X_n})$, with $\delta_x(A) := \mathbb{1}_A(x)$. The log likelihood $LL_n(\phi)$ then equals $LL_n(\phi) = n \int \log f(x, \phi) dP_n(x)$, with $-\infty \leq LL_n(\phi) < +\infty$ for all ϕ . We will need the following assumptions.

(A₁) The set $\mathcal{P} = \{P_\phi : \phi \in \Theta\}$ is a family of laws dominated by a σ -finite measure μ on some sample space (χ, \mathcal{B}) . Let $f(x, \phi)$ be the density $(dP_\phi/d\mu)(x)$, $x \in \chi$, $\phi \in \Theta$. We take $0 \leq f(x, \phi) < \infty$.

(A₂) Θ is an open set in a Euclidean space \mathbf{R}^d .

(A₃) The observations X_1, X_2, \dots are i.i.d. random variables with values in χ and with some law P (not necessarily in the family \mathcal{P}).

(A₄) A prior probability $\pi_0(\phi) d\phi$ is given on Θ with a continuous density $\pi_0(\phi) > 0$ for all ϕ , so $\int_{\Theta} \pi_0(\phi) d\phi = 1$.

(A₅) There is a $\phi_0 \in \Theta$, to be called the pseudo-true value of ϕ , such that for every neighborhood \mathcal{N} of ϕ_0 there is a $\kappa > 0$ such that almost surely for n large enough, $\sup_{\phi \notin \mathcal{N}} LL_n(\phi) < \sup_{\phi \in \mathcal{N}} LL_n(\phi) - n\kappa$.

(A₆) For ϕ in a small enough neighborhood W of ϕ_0 , the function $f(x, \cdot)$ is strictly positive and C^2 in ϕ for P -almost all x , and the P -Fisher information matrix $E(\phi) := \{E_{ij}(\phi)\}_{i,j=1}^d := \{E_P(-\partial^2 \log f(\cdot, \phi) / \partial \phi_i \partial \phi_j)\}_{i,j=1}^d$ exists and is finite, strictly positive definite and continuous in ϕ .

(A₇) For some neighborhood W of ϕ_0 , the class $\mathcal{F}_W := \{-\partial^2 \log f(\cdot, \phi) / \partial \phi_i \partial \phi_j\}$ of functions for $\phi \in W$, $i, j = 1, \dots, d$, is a Glivenko–Cantelli class for P ; in other words we have $\sup_{g \in \mathcal{F}_W} |\int g d(P_n - P)| \rightarrow 0$ almost surely as $n \rightarrow \infty$.

REMARKS. Sufficient conditions for (A₅) to hold can be found from sufficient conditions for consistency of approximate maximum likelihood estimators; see Huber (1967), and for further extensions, Dudley (1998). Berk (1966) and Poskitt (1987) have considered the limiting behavior of posterior distributions for observations with a law P not in a parametric family.

Both (A₆) and (A₇), if true for one neighborhood W of ϕ_0 , are also true for all smaller neighborhoods, so we can take W to be the same in both. Since P will usually be unknown to the statistician, one will need to check that assumptions (A₆) and (A₇) hold for all P in a class including at least all P_ϕ and preferably all P in some large class. Assumptions (A₆) and (A₇) hold for an exponential family in standard form, where $f(x, \phi) = e^{x \cdot \phi - b(\phi)}$, since the second derivatives of $\log f$ are constant in x .

Talagrand (1987) gave a characterization of Glivenko–Cantelli classes for a given P . Dudley (1998) applies Talagrand’s theorem to extend the Glivenko–Cantelli property to transformed classes; see also van der Vaart and Wellner (2000). Sufficient conditions are known for a class \mathcal{F} to be a Glivenko–Cantelli class for all P , and a criterion is known up to measurability for \mathcal{F} to be a Glivenko–Cantelli class uniformly in P [Dudley, Giné and Zinn (1991)]. See also Alon, Ben-David, Cesa-Bianchi and Haussler (1997).

Note that (A₆) and (A₇) are preserved by nonsingular linear transformations of ϕ_1, \dots, ϕ_d . Let I denote the $d \times d$ identity matrix.

Let A be a half-space $\{\phi \in \Theta : \phi \cdot v_A \geq M\}$ for some $v_A \in \mathbf{R}^d$, $|v_A| = 1$ and $M \in \mathbf{R}$. Then ∂A is the boundary hyperplane $\partial A := \{\phi \in \Theta : \phi \cdot v_A = M\}$. Let

$$\Delta := \Delta_{A,n} := \left(\sup_{\phi \in \Theta} - \sup_{\phi \in \partial A} \right) LL_n(\phi),$$

where if the suprema are attained at unique points these are $\hat{\phi}_n$ (unrestricted MLE) and $\tilde{\phi}_n$ (MLE in ∂A), respectively. Here 2Δ is known as the (usual) likelihood ratio (test) statistic, and $\pm(2\Delta)^{1/2}$ as a likelihood root [defined, e.g., as m_1 for $p = 1$, $q = 0$ in Lawley (1956), Section 4].

When a statement is made “a.s. for n large enough” it will mean that it holds with probability 1 for $n \geq n_i$ where n_i is a random variable, *not depending on A* , but possibly depending on other quantities, as will be made explicit in the proofs.

THEOREM 1. *Suppose assumptions (A₁)–(A₇) hold, for a neighborhood W of ϕ_0 in (A₆) and (A₇). Let $\pi_0(\phi) d\phi$ be a prior law on Θ satisfying (A₄). For $x = (X_1, \dots, X_n)$ let $\pi_{x,n}$ be the posterior distribution on Θ . Then almost surely $\hat{\phi}_n$ exists for n large enough.*

For any $\varepsilon > 0$ there exists $\kappa' > 0$ such that almost surely for n large enough, for all half-spaces A , if $\hat{\phi}_n \notin A$, then either:

- (a) $\Phi(-\sqrt{2\Delta})/(1 + \varepsilon) \leq \pi_{x,n}(A) \leq (1 + \varepsilon)\Phi(-\sqrt{2\Delta})$, or
- (b) both $\pi_{x,n}(A) \leq 3e^{-n\kappa'}$ and $\Phi(-\sqrt{2\Delta}) \leq e^{-n\kappa'}$.

Or, if $\hat{\phi}_n \in A$, then the same holds for $B = \overline{A^c}$ in place of A .

Specifically, if U is a neighborhood of ϕ_0 , then for some $\alpha > 0$, $\pi_{x,n}(U^c) \leq 3e^{-n\alpha}$ almost surely for n large enough, and for small enough $\kappa' > 0$ (b) holds for any half-space $A \subset U^c$. Or, if A_n is a sequence of half-spaces such that $d(A_n, \phi_0) := \inf\{|\phi - \phi_0| : \phi \in A_n\} \rightarrow 0$ as $n \rightarrow \infty$, and $\hat{\phi}_n \notin A_n$ for each n , then almost surely $\pi_{x,n}(A_n)/\Phi(-\sqrt{2\Delta_{A_n,n}}) \rightarrow 1$ as $n \rightarrow \infty$.

The proof will be given in Sections 4 through 8. The main steps in the proof are as follows: Assume that $\hat{\phi}_n \notin A$, the case $\hat{\phi}_n \in A$ following by symmetry. For any neighborhood V of ϕ_0 , if A is disjoint from V then (b) in the Theorem holds for n large enough. We will find a small enough V so that if A intersects V we can make a linear change of coordinates from ϕ into y so that $\hat{y} = 0$, the boundary ∂A is the hyperplane where the first coordinate y_1 is some constant, a unique MLE \tilde{y} in ∂A exists and is the point of A closest to 0, and the empirical Fisher information matrix $E_n(y)$ with respect to y is close to the identity matrix when y is close to 0.

We then obtain bounds on integrals $I_n(B) = \int_B e^{-nh(y)}\pi(y) dy$, where $\pi(\cdot)$ is the prior density in y coordinates, $nh(y)$ is the difference in log likelihoods at the MLE \hat{y} and y , so that $\pi_{x,n}(y \in C) = I_n(C)/I_n(\Theta)$. Outside of a neighborhood $U(\rho)$ of the MLE, $\pi_{x,n}$ is exponentially small, so $\pi_{x,n}(A)$ equals $I_n(A \cap U(\rho))/I_n(U(\rho))$ up to an exponentially small quantity. Propositions 2 and 3 give asymptotic evaluations of $I_n(U(\rho))$ and $I_n(A \cap U(\rho))$, which rely on Lemma 5, relating to Mills’ ratios and their extension to ratios of gamma tail probabilities to densities.

3. Sharper approximations. Extending multidimensional Laplace approximation formulas of Hsu (1948) and Fulks and Sather (1961) [see Wong (1989), Section IX.5], approximation formulas for integrals of likelihood functions over hyperplanes or manifolds have been proved under different conditions by Haughton (1984, 1988) and Poskitt (1987). Shun and McCullagh (1995) give further asymptotic expansion terms and consider high dimensions relative to n ; see also Barndorff-Nielsen and Wood [(1998), Section 4]. Leonard (1982), Tierney and Kadane (1986), Tierney, Kass and Kadane (1989), DiCiccio and Martin (1991) and DiCiccio and Stern (1993) applied the Laplace method to approximate a marginal posterior density in case of a parameter $\phi = (\psi, \eta)$ where $\psi = (\psi_1, \dots, \psi_p)$ is a parameter of interest and $\eta = (\eta_1, \dots, \eta_q)$ is a nuisance parameter. In our case, where we are interested in posterior probabilities of a half-space A which can be written $\psi_1 \geq \psi'_1$, we have $p = 1$, $\psi \equiv \psi_1$, $q = d - 1$, $\eta_j = \phi_{j+1}$ for $j = 1, \dots, d - 1$. Let $\pi(\phi)$ be the prior density, $Y = (Y_1, \dots, Y_n)$ the observations, $\tilde{\phi}(\psi) = (\psi, \tilde{\eta}(\psi))$ the MLE of ϕ for fixed ψ , if it exists, $\hat{\phi} = (\hat{\psi}, \tilde{\eta}(\hat{\psi}))$ the overall MLE, ℓ the log likelihood, $m(\psi) := \ell(\tilde{\phi}(\psi))$ the profile log likelihood, and $\ell_{\eta\eta} := \{\partial^2 \ell / \partial \eta_i \partial \eta_j\}_{i,j=1}^q$ the Hessian of ℓ with respect to η . Let $\pi_{\psi|Y}(\psi)$ be the marginal posterior density of ψ . Then the approximation in the form stated by DiCiccio and Stern (1993) is

$$(1) \quad \pi_{\psi|Y}^*(\psi) := c^* \left(\frac{\det[-\ell_{\eta\eta}(\tilde{\phi}(\psi))]}{\det[-\ell_{\eta\eta}(\hat{\phi})]} \right)^{-1/2} \frac{\pi(\tilde{\phi}(\psi))}{\pi(\hat{\phi})} e^{m(\psi) - m(\hat{\psi})},$$

where c^* is a normalizing constant. (Note that functions of $\hat{\phi}$ and $\hat{\psi}$ do not depend on the argument ψ and in that sense are also constants.) Let z be the likelihood root

$$(2) \quad z := \text{sgn}(\psi - \hat{\psi}) \sqrt{2(m(\hat{\psi}) - m(\psi))}.$$

If there is a pseudo-true ψ_0 and a 1-1, mutually C^1 relationship between z and ψ for ψ in some neighborhood of ψ_0 , as will be shown in the Appendix, we get an approximate posterior density for z ,

$$(3) \quad \pi_{z|Y}^{**}(z) := c^* \left(\frac{\det[-\ell_{\eta\eta}(\tilde{\phi}(\psi(z)))]}{\det[-\ell_{\eta\eta}(\hat{\phi})]} \right)^{-1/2} \frac{\pi(\tilde{\phi}(\psi(z)))}{\pi(\hat{\phi})} e^{-z^2/2} \frac{d\psi}{dz}.$$

We then define an approximation for the posterior probability of a half-space $A : \psi \geq \psi' > \hat{\psi}$ or equivalently $z \geq z_1 > 0$, say, by

$$(4) \quad \begin{aligned} & \pi^{**}(z \geq z_1) \\ & := \Phi(-z_1) \left(\frac{\det[-\ell_{\eta\eta}(\tilde{\phi}(\psi(z_1)))]}{\det[-\ell_{\eta\eta}(\hat{\phi})]} \right)^{-1/2} \frac{\pi(\tilde{\phi}(\psi(z_1)))}{\pi(\hat{\phi})} \frac{d\psi/dz|_{z=z_1}}{d\psi/dz|_{z=0}}. \end{aligned}$$

The approximation given by Theorem 1 is just $\Phi(-z_1)$. We can now state the following corollary.

COROLLARY 1. Under the hypotheses of Theorem 1, if A_n are half-spaces with $\hat{\phi}_n \notin A_n$ and $d(A_n, \phi_0) \rightarrow 0$ as $n \rightarrow \infty$ where for each n , A_n is written as $\{\psi \geq \eta_n\}$ (where the parameter ψ , a linear function of ϕ , can also depend on n), then the posterior probability $\pi_{x,n}(A_n)$ is asymptotic as $n \rightarrow \infty$ to its sharpened Laplace approximation $\pi^{**}(z \geq z_n)$, which is well defined almost surely for n large enough.

A proof will be given in the Appendix.

To improve the approximation of small posterior probabilities, one can consider π^{**} and the following approximations, but the rest of this section is nonrigorous.

The approximation $\pi^{**}(z \geq z_1)$ can be sharpened further. We will define approximations in the one-dimensional case $p = d = 1, q = 0$, assuming that the prior density π is differentiable at least once, or has higher derivatives for higher-order approximations. The ratio of determinants is replaced by 1 and $\tilde{\phi}(\psi) \equiv \psi$. Let τ be the prior density as a function of z , so $\tau(z) = \pi(\psi(z)) d\psi/dz$. Let $z_1 \geq 0$. The posterior probability $\pi_Y(z \geq z_1)$ equals

$$(5) \quad \int_{z_1}^{\infty} \tau(z)e^{-z^2/2} dz / \int_{-\infty}^{\infty} \tau(z)e^{-z^2/2} dz,$$

exactly if the 1-1, increasing, C^1 relationship between z and ψ holds for all $|z| < \infty$ (as in the beta case in Section 10), or up to an error exponentially small in n under broader conditions, with ∞ replaced by $K\sqrt{n}$ for some $K < \infty$.

Let $\pi_Y^{k,*}(z \geq z_1)$ denote the approximation to $\pi_Y(z \geq z_1)$ where τ is replaced by partial Taylor expansions through order k , around 0 in the denominator and around z_1 in the numerator. The integrals of polynomials in z times $\exp(-z^2/2)$ can be found explicitly. In the denominator, odd powers of z yield 0 integrals. We get $\pi^{**}(z \geq z_1) = \pi_Y^{0,*}(z \geq z_1)$ and for $\zeta := z_1$,

$$\begin{aligned} \pi_Y^{1,*}(z \geq z_1) &= \tau(0)^{-1} [\Phi(-z_1)\tau(z_1) + \tau'(z_1)(\phi(z_1) - z_1\Phi(-z_1))], \\ \pi_Y^{2,*}(z \geq \zeta) &= \frac{\Phi(-\zeta)[\tau(\zeta) - \zeta\tau'(\zeta) + \tau''(\zeta)(1 + \zeta^2)/2] + \phi(\zeta)[\tau'(\zeta) - \zeta\tau''(\zeta)/2]}{\tau(0) + \tau''(0)/2}. \end{aligned}$$

DiCiccio and Martin (1991) gave a related approximation. In one dimension, letting $l(\psi)$ denote the log likelihood function and $l^{(k)}$ its k th derivative, for a given ψ_1 , the posterior tail probability $\pi_n(\psi \leq \psi_1)$ is approximated by

$$(6) \quad \pi_Y^{DM,*} := \Phi(z_1) + \phi(z_1) \left[\frac{1}{z_1} + \frac{\sqrt{-l^{(2)}(\hat{\psi})\pi(\psi)}}{l^{(1)}(\psi)\pi(\hat{\psi})} \right],$$

where $\psi(z_1) = \psi_1$ and we have in mind $\psi_1 \leq 0$ in this case. These approximations will appear in the examples in Section 10.

Following a method going back at least to Bleistein (1966), Temme (1982) in effect expands an integral $\int_{-\infty}^w \phi(z) f(\delta z) dz$ for a smooth function f as

$$f(0)\Phi(w) + \int_{-\infty}^w \frac{f(0) - f(\delta z)}{z} d\phi(z),$$

integrating by parts and iterating the process to obtain an asymptotic expansion in powers of $\delta \downarrow 0$. Here Temme (1982), going beyond earlier work by analysts as far as we know, allows f to depend separately on δ if it satisfies suitable conditions. In our case $\delta = 1/\sqrt{n}$. Although the resulting expansions appear at first sight different from ours, we have verified that: if $\delta w \rightarrow 0$, while possibly $w \rightarrow -\infty$, the terms of orders δ^j , $j = 0, 1, 2$, are the same in our expansions of the numerator and denominator of $\pi_Y^{*,2}$ as in those of Temme (1982). Or, if $\delta \downarrow 0$, $w \rightarrow -\infty$ and $\delta w \rightarrow u < 0$ (a large deviation case), then our expansion and Temme's have different coefficients of $\Phi(w)$ and $\phi(w)$, but via

$$\Phi(w) = \phi(w) \left[-\frac{1}{w} + \frac{1}{w^3} + O(w^{-5}) \right]$$

as $w \rightarrow -\infty$, Temme's series, and ours except for the error bound, give

$$(7) \quad \phi(w) \left[-\frac{\delta f(u)}{u} + \delta^3 \left\{ \frac{f(u) - uf'(u)}{u^3} \right\} + O(\delta^5) \right],$$

where the $O(\delta^5)$ term also depends on f and u .

When applying Temme's expansion method to a beta probability

$$I_x(a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt / B(a, b), \quad 0 < x < 1, \quad a > 0, \quad b > 0,$$

there are different possible choices of δ such as $1/\sqrt{a+b}$ [Temme (1982)] or $1/\sqrt{a+b-2}$ for $a+b > 2$ or $1/\sqrt{a+b-1}$ for $a+b > 1$ (Section 10 below). Temme (1982, 1987) evaluates the complete beta function $B(a, b)$ directly, whereas we apply approximations to it as examples of how the approximations might work more generally.

There is an evident similarity between the sharpened Laplace approximations to posteriors just given and the "saddlepoint" approximation to tail probabilities for sample means due to Lugannani and Rice (1980); see also Daniels (1987) and Jensen [(1995), Chapter 3]. One contrast is that the Lugannani-Rice and related approximation formulas (e.g., for distributions of maximum likelihood estimators) are based on moment generating functions, sometimes for approximating distributions [e.g., Fraser, Reid and Wu (1999)], and do not require explicit knowledge of the densities of the sample means. Conversely, $\pi_Y^{i,*}$ and $\pi_Y^{\text{DM},*}$ do not involve moment generating functions, but they do use the likelihood functions and prior

densities (which are usually given explicitly in the Bayes case) and their derivatives.

In applications of the Laplace method to posterior probabilities, DiCiccio and Martin (1991) and Fraser, Reid and Wu (1999) gave error rates of $O(n^{-3/2})$ but only for Laplace deviations. We will see in Table 3 of Section 10 that $\pi_Y^{i,*}$ and $\pi_Y^{DM,*}$ also work well in a large deviation example.

4. Preliminary transformations. Here we begin the proof of Theorem 1. Let $L_n(\phi) := \exp(LL_n(\phi))$ be the likelihood function. We first note that the posterior law is eventually well defined under our assumptions: it follows from (A₅) that for a compact neighborhood \mathcal{N} of ϕ_0 , a.s. for each n large enough, $L_n(\phi)$ is bounded outside of \mathcal{N} . On \mathcal{N} it is bounded by continuity (A₆). Then $\int L_n(\phi)\pi_0(\phi) d\phi < \infty$. Also by (A₅), for the same n , $L_n(\phi) > 0$ for some $\phi \in \mathcal{N}$ and so by continuity $L_n(\phi) > 0$ on some nonempty open set. Then by (A₄) $\int L_n(\phi)\pi_0(\phi) d\phi > 0$. Thus a.s. for n large enough, $L_n(\phi)\pi_0(\phi) / \int L_n(\phi')\pi_0(\phi') d\phi'$ is a well-defined posterior probability density.

We will assume first and for most of the proof that $\hat{\phi}_n \notin A$. The case $\hat{\phi}_n \in A$ will follow by symmetry. We will make a linear change of coordinates in Θ so that we can assume $\hat{\phi}_n = 0$, $E(\phi_0)$ is the identity matrix, A is a half-space with boundary where the first coordinate ϕ_1 is some constant and $\tilde{\phi}$ is the point of A closest to 0. Moreover, we need to make these changes “without loss of generality,” which requires proof since A varies over all half-spaces and κ' must not depend on A . Also, our informal statements so far have assumed that unique maximum likelihood estimators $\hat{\phi}_n$ in Θ and $\tilde{\phi}_n$ in ∂A exist, which is true for $\hat{\phi}_n$ for n large enough, but not necessarily true for $\tilde{\phi}_n$ simultaneously for all half-spaces A . It is true if ∂A is close enough to $\hat{\phi}_n$ in a sense to be made precise as Case I (28) after some coordinate transformations. The full transformation involving $\tilde{\phi}_n$ will be made in Case I.

We can take $0 < \varepsilon < 1$. Take $\delta > 0$ small enough so that

$$(8) \quad (1 + \delta)^{3d+9} < 1 + \varepsilon.$$

Then $\delta < 1/10$. Let $\delta_1 := \delta^2/100$. By a fixed linear transformation of ϕ (not depending on A) we can obtain coordinates θ in which $E(\theta_0) = I$, the identity matrix, where θ_0 is ϕ_0 expressed in θ coordinates. By assumption (A₄) there exists a neighborhood V_0 of θ_0 such that, for the prior density $\pi_0(\phi) d\phi$ now expressed as $\pi_1(\theta) d\theta$, and $\eta := \inf_{\theta \in V_0} \pi_1(\theta)$,

$$(9) \quad \eta \leq \pi_1(\theta) \leq \eta(1 + \delta), \quad \theta \in V_0.$$

To recall some facts about matrix norms, let K and J be $d \times d$ real matrices. For the Euclidean norm $\|t\| := (t_1^2 + \dots + t_d^2)^{1/2}$, we use the matrix norm $\|K\| := \max_{\|t\|=1} \|Kt\|$. Let $K \leq J$ mean that $J - K$ is nonnegative definite and symmetric. For any symmetric real matrix K , $\|K\| = \max_i |\lambda_i|$ for the

eigenvalues λ_i of K . Thus if $-cI \leq K \leq cI$ and $c \geq 0$, then $\|K\| \leq c$. Conversely if J is symmetric and $\|J - I\| < 2c/3$ where $0 < c < 1/2$ then

$$(10) \quad I/(1+c) \leq J \leq (1+c)I.$$

Then, there is a $\gamma_1 > 0$ such that for W as in (A₆) and (A₇) and

$$(11) \quad V := \{\theta : \|\theta - \theta_0\| < 4\gamma_1\},$$

$V \subset W \cap V_0$ and $\|E(\theta) - I\| < \delta_1/3$ for all $\theta \in V$.

By a rotation which does not change distances, and keeps $E(\theta_0) = I$, we can assume that $v_A = (1, 0, \dots, 0)$, so A is a set

$$(12) \quad A = \{\theta : \theta_1 \geq M\}.$$

Let $E_n(\theta)$ be the empirical Fisher information matrix,

$$(13) \quad E_n(\theta) := \left\{ - \int (\partial^2 \log f(x, \theta) / \partial \theta_i \partial \theta_j) dP_n(x) \right\}_{i,j=1}^n.$$

Almost surely for n large enough, by (A₇),

$$(14) \quad \|E_n(\theta) - I\| < 2\delta_1/3$$

for all $\theta \in V$, and by (A₅), a.s. for n large enough the supremum of the likelihood will be attained at some point θ where $\|\theta - \theta_0\| \leq \gamma_1$. Specifically, this will occur for $n \geq n_0(\xi, \gamma_1)$ where ξ is a point of the probability space on which X_j are defined. At any such point, the gradient of $LL_n(\theta)$ exists and is 0. By (14) and (10) we have

$$(15) \quad I/(1+\delta_1) \leq E_n(\theta) \leq (1+\delta_1)I$$

for all $\theta \in V$. By (15), E_n is strictly positive definite on V , so almost surely,

$$(16) \quad \text{for } n \geq n_0(\xi, \gamma_1) \text{ a unique MLE } \hat{\theta}_n \text{ exists with } \|\hat{\theta}_n - \theta_0\| < \gamma_1.$$

By a translation we can assume

$$(17) \quad \hat{\theta}_n = 0.$$

Then from (12),

$$(18) \quad A = \{\theta : \theta_1 \geq \zeta\}$$

for some ζ depending on $\hat{\theta}_n$, A and the choice of coordinates. The translation preserves distances and the fact that $E(\theta_0) = I$.

By (18), $\zeta > 0$ since $\hat{\theta}_n = 0 \notin A$. Also by (15), LL_n is strictly concave on V . Thus if $A \cap V$ is nonempty, and since we assume $\hat{\theta}_n \notin A$, we have

$$(19) \quad \sup_{A \cap V} LL_n = \sup_{\partial A \cap V} LL_n.$$

For $\theta \in \Theta$, let

$$(20) \quad H(\theta) = H_n(\theta) = \frac{1}{n} \sum_{i=1}^n [\log f(X_i, 0) - \log f(X_i, \theta)].$$

Along every line segment $\theta = tw$ in V , say, $a < t < b$, where $\|w\| = 1$, we have by (15) that $1/(1 + \delta_1) \leq d^2 H(tw)/dt^2 \leq 1 + \delta_1$. A Taylor expansion around $0 = \hat{\theta}_n$, where $\nabla H = 0$, then gives $dH(tw)/dt|_{t=0} = H(0) = 0$,

$$(21) \quad 1/(1 + \delta_1) \leq [dH(tw)/dt]/t \leq 1 + \delta_1 \quad \text{for } t \neq 0,$$

$$(22) \quad t^2/[2(1 + \delta_1)] \leq H(tw) \leq t^2(1 + \delta_1)/2.$$

In what follows we let $H(\zeta, X) \equiv H(\zeta)$ if $d = 1$. For $\zeta \leq \gamma_1$ we have

$$(23) \quad H(\zeta, 0) \leq \zeta^2(1 + \delta_1)/2$$

[here 0 is the $(d - 1)$ -dimensional zero vector]. Also, for $(\zeta, X) \in V$,

$$(24) \quad H(\zeta, X) \geq (\zeta^2 + \|X\|^2)/[2(1 + \delta_1)].$$

Let

$$(25) \quad \rho := \gamma_1/(1 + 6^{-1}\delta) \leq \gamma_1/(1 + (\delta^2/36))^{1/2}.$$

Let $r := \rho/2$. Let Ω_{d-1} be the total $(d - 1)$ -dimensional area of the sphere $S^{d-1} := \{\omega \in \mathbf{R}^d : \|\omega\| = 1\}$. Then $\Omega_0 := 2$, $\Omega_1 = 2\pi$, $\Omega_2 = 4\pi, \dots$, $\Omega_k = 2\pi^{(k+1)/2}/\Gamma((k + 1)/2)$ for $k = 0, 1, \dots$. Let $C_1 := 1/2$, $C_2 := \pi/9$ and $C_d := \Omega_{d-1}/(2\Omega_{d-2})$ for $d \geq 3$. Let, for κ in (A_5) with $\mathcal{N} = V$,

$$(26) \quad \zeta_0 := \zeta_0(d, r, \delta) := \min(C_d r \delta, \sqrt{\kappa}).$$

Since $\Omega_d = 2 \int_0^1 \Omega_{d-1}(1 - x^2)^{(d-2)/2} dx \leq 2\Omega_{d-1}$ for $d \geq 2$, we have for all $d = 1, 2, \dots$,

$$(27) \quad C_d \leq 1.$$

5. Cases of the proof. Recall ζ_0 as defined in (26). The proof of Theorem 1 will be divided into two cases:

$$(28) \quad \text{Case I:} \quad 0 < \zeta < \zeta_0,$$

where the boundary $\partial A = \{\theta_1 = \zeta\}$ is not too far from the MLE $\hat{\theta}_n = 0$, and

$$(29) \quad \text{Case II:} \quad \zeta \geq \zeta_0,$$

where ∂A is bounded away from $\hat{\theta}_n$ by a fixed amount.

Here, we will begin the main proof in Case I, where either (a) or (b) in Theorem 1 may hold. We will find a unique MLE $\tilde{\theta}_n$ in A or equivalently by (19), (23) and (26) in ∂A . The following arguments through (38) are immediate if $d = 1$ since (ζ, X) is replaced by ζ and X by $0 \in \mathbf{R}^{d-1}$.

For fixed ζ with $0 < \zeta < \zeta_0$, $\inf_{(\zeta, X) \in V} H(\zeta, X)$ can by (23) and (24) be restricted to (ζ, X) such that $(\zeta^2 + \|X\|^2)/(1 + \delta_1) \leq \zeta^2(1 + \delta_1)$, or equivalently $1 + \|X\|^2/\zeta^2 \leq (1 + \delta_1)^2$, and thus since $\delta < 1/10$ and $\delta_1 = \delta^2/100$,

$$(30) \quad \|X\|^2/\zeta^2 \leq 2\delta_1 + \delta_1^2 \leq \delta^2/36.$$

This then implies $\zeta^2 + \|X\|^2 \leq \gamma_1^2$ by (25) since $\zeta < \rho$. Thus $(\rho, X) \in V$ by definition (11) of V and (16) for $n \geq n_0(\xi, \gamma_1)$. On the compact, convex set of (ζ, X) satisfying (30) for fixed $\zeta > 0$, H is a C^2 function, strictly convex by (15), so H attains its minimum on $\{\theta_1 = \zeta\} \cap V$ at a unique point $\tilde{\theta} := (\zeta, \tilde{X})$, where $\tilde{X} = \tilde{X}(\zeta)$.

Define new coordinates y by the linear transformation $y = T(\theta)$ such that for $u^{(1)} := (u_2, \dots, u_d)$, where u may be y or θ ,

$$(31) \quad y_1 = \theta_1, \quad y^{(1)} = \theta^{(1)} - \theta_1 \tilde{X}/\zeta.$$

We have $\zeta < \rho$, so that $(\zeta, 0) \in U(\rho) := \{y : \|y\| < \rho\}$. In the y coordinates, A still is $\{y_1 \geq \zeta\}$ and $\hat{y} = 0$. Now $\tilde{y} = (\zeta, 0)$. Define $h = h_n$ so that

$$(32) \quad h(y) := h_n(y) = H_n(\theta) = H_n(y_1, y^{(1)} + y_1 \tilde{X}/\zeta).$$

Then since $\zeta < \rho$, $\|T - I\| \leq \delta/6$ by (30). Since T^{-1} and its transpose $(T^{-1})^t$ have the same form as T and T^t , respectively, in this case, replacing \tilde{X} by $-\tilde{X}$, also

$$(33) \quad \|T^{-1} - I\| = \|(T^{-1})^t - I\| \leq \delta/6.$$

Thus since $\zeta < \zeta_0$, if $\|y\| < \rho$ then $\|\theta\| = \|\theta - \hat{\theta}_n\| \leq (1 + (\delta/6))\|y\| < \gamma_1$ by (25), so $\theta \in V \subset V_0$ by choice of γ_1 and (16), for $n \geq n_0(\xi, \gamma_1)$. We have $D^{(2)}H_n = E_n$ by (13) and (20), and $D^{(2)}h_n = (T^{-1})^t(D^{(2)}H_n)T^{-1}$. By (14) and (33) and a short calculation we then have $\|D^{(2)}h_n - I\| < 2\delta/3$. Thus by (10), if $\|y\| < \rho$,

$$(34) \quad I/(1 + \delta) \leq D^{(2)}h_n \leq (1 + \delta)I.$$

Note that the determinants (Jacobians) of an orthogonal transformation, translation and T (31) are all ± 1 , so that the volume element $dy = d\theta$ is unchanged. Then by (9), for $\|y\| < \rho$,

$$(35) \quad \eta \leq \pi(y) \leq \eta(1 + \delta),$$

where $\pi(y)$ is the density of the prior π_1 for θ expressed in the y coordinates.

Recall that $U(\rho) := \{y : \|y\| < \rho\}$. Note that $U(\rho)$, by (31), depends on $\hat{\theta}_n$, $\zeta > 0$, \tilde{X} , and its radius $\rho > 0$ for y , $U(\rho) = U(\rho, \zeta, \tilde{X}, \hat{\theta}_n)$.

Reviewing then, in Case I, since $0 < \zeta < \rho < \gamma_1$, in the y coordinates, H_n is expressed as $h_n \equiv h_n(y)$, the maximum likelihood estimate \hat{y} is equal to 0, and the maximum likelihood estimate on the set $\partial A = \{y_1 = \zeta\}$ equals $(\zeta, 0)$.

6. Bounds for the likelihood function and integrals. It follows from (20), (32), (34) and a Taylor expansion around $y = 0$ as in (22) that

$$(36) \quad \|y\|^2/[2(1 + \delta)] \leq h(y) \leq (1 + \delta)\|y\|^2/2$$

for all y in $U(\rho)$. We have $A \cap U(\rho) = \{y_1 \geq \zeta\} \cap U(\rho)$. Since h is minimized at $(\zeta, 0)$ on $\{y_1 = \zeta\} \cap U(\rho)$, it follows that $\partial h/\partial y_j = 0$ at $(\zeta, 0)$ for $j = 2, \dots, d$. By a Taylor expansion around $(\zeta, 0)$, we have by (34) for $(\zeta, X) \in U(\rho)$ and

$$(37) \quad \tau := h(\zeta, 0),$$

$$(38) \quad \tau + \frac{1}{2(1 + \delta)}\|X\|^2 \leq h(\zeta, X) \leq \tau + \frac{(1 + \delta)}{2}\|X\|^2.$$

Note that

$$(39) \quad n\tau = \Delta$$

by (17), (20), (32) and (37). Assume that $d \geq 2$ (the case $d = 1$ is easier and will be omitted). Define

$$(40) \quad I_n(B) := \int_B e^{-nh(y)}\pi(y) dy$$

for any measurable set B . Note that by (20) and (32), the likelihood function equals $ML e^{-nh}$, where ML is the maximum likelihood on Θ .

Recall that $\pi_{x,n}$ is the posterior distribution on Θ , so a.s. for n large enough, for any measurable set $C \subset \Theta$ in y coordinates, $\pi_{x,n}(y \in C)$ equals

$$(41) \quad ML \int_C e^{-nh(y)}\pi(y) dy / \left(ML \int_{\Theta} e^{-nh(y)}\pi(y) dy \right) = I_n(C)/I_n(\Theta).$$

We will prove Theorem 1 by way of three propositions. Proposition 1 will say that $\pi_{x,n}(U(\rho)^c)$ is exponentially small in n ; Proposition 2 gives an asymptotic evaluation for $I_n(U(\rho))$ and Proposition 3 does so for $I_n(A \cap U(\rho))$. For any $\rho' > 0$ not depending on n let $U' := U'(\rho') := B(\hat{\theta}_n, \rho') = B(0, \rho')$ in the θ coordinates as ultimately chosen, ending with (17).

PROPOSITION 1. *For some $\nu > 0$ not depending on the observations, a.s. for n large enough, $\pi_{x,n}(U'(\rho/2)^c) < e^{-n\nu}$ and whenever $0 < \zeta \leq \rho$, $\pi_{x,n}(U(\rho)^c) < e^{-n\nu}$.*

PROPOSITION 2. *We have almost surely for n large enough,*

$$(42) \quad (1 + \delta)^{-(3d+10)/2} \leq \frac{1}{\eta} I_n(U(\rho))(n/2\pi)^{d/2} \leq (1 + \delta)^{(3d+6)/2}.$$

PROPOSITION 3. *In Case I, for $d \geq 2$, we have almost surely for n large enough, not depending on A ,*

$$(43) \quad \eta(2\pi/n)^{d/2}(1 + \delta)^{-(3d+10)/2}\Phi(-\sqrt{2n\tau}) \leq I_n(A \cap U(\rho))$$

$$(44) \quad \leq \eta(2\pi/n)^{d/2}(1 + \delta)^{(3d+6)/2}\Phi(-\sqrt{2n\tau}).$$

The three propositions will be proved in Section 8 after Lemma 5. Recall that the orthogonal transformation just before (12) and the translation just after (16) did not change distances in the θ coordinates.

Let $\omega = (\omega_1, \dots, \omega_d)$, $S(d^+) := \{\omega \in \mathbf{R}^d : \|\omega\| = 1, \omega_1 > 0\}$. Let $d\omega$ be the surface element on $S(d^+)$. For any Borel set $S \subset S(d^+)$, the surface area measure is $\Omega_{d-1}(S) := \int_S d\omega$. Let $D_\omega := \zeta/\omega_1$, and $h_\omega := D_\omega\|\omega^{(1)}\|$, where $\omega^{(1)} = (\omega_2, \dots, \omega_d) \in \mathbf{R}^{d-1}$. For $0 \leq a \leq b \leq \infty$ let $[[a, b)) := \{\omega : a \leq D_\omega < b\}$. Then

$$(45) \quad D_\omega^2 = \zeta^2 + h_\omega^2.$$

Let the spherical coordinates of y be (t, ω) where $\omega := y/\|y\|$ and $t = \|y\|$. Then $dy = t^{d-1} d\omega dt$. Let $S(\zeta, \rho) := \{\omega \in \mathbf{R}^d : \|\omega\| = 1, \omega_1 \geq \zeta/\rho\}$. We have

$$(46) \quad I_n(A \cap U(\rho)) = \int_{S(\zeta, \rho)} \int_{D_\omega}^\rho e^{-nh(t\omega)} \pi(t\omega) t^{d-1} dt d\omega.$$

If $\pi(\cdot)$ were constant and h linear then the integrals $\int_{D_\omega}^\rho = \int_{D_\omega}^\infty - \int_\rho^\infty$ would be proportional to gamma probabilities. If h is quadratic without a linear term they reduce to gamma probabilities or, if $d = 1$, to normal probabilities. We will bound the corresponding integrals above and below by gamma probabilities. Thus we consider gamma and normal probabilities in the following section.

For $x > 0$, $[[x, \infty)) = \{\omega : \zeta/\omega_1 \geq x\} = \{\omega : 0 \leq \omega_1 \leq \zeta/x\}$. Recalling ζ_0 (26) and Ω_d, C_d as defined just before (26), we have the lemma.

LEMMA 1. *For any ζ with $0 < \zeta < \zeta_0$, $\Omega_{d-1}([[r, \infty))) < \delta\Omega_{d-1}/2$, where $\Omega_0(\cdot)$ is a sum of unit point masses at 1 and -1 .*

PROOF. For $d = 1$, $S(1^+) = \{1\}$, so $[[r, \infty))$ is empty by (26) and (27) since $\zeta < \zeta_0 < r$. For $d = 2$, $\Omega_1([[r, \infty))) = 2 \arcsin(\zeta/r)$. Since \arcsin is a convex function on $[0, 1]$, $\arcsin x \leq 2x$ for $0 \leq x \leq 1$. So since $\zeta/r < 1$, and by (26), $\Omega_1([[r, \infty))) \leq 4\zeta/r < 4C_2\delta \leq \delta\pi/2 = \delta\Omega_1/4$. This proves the lemma for $d = 2$.

Let now $d \geq 3$. We have, again by (26), and since $0 < \zeta < \zeta_0$,

$$\begin{aligned} \Omega_{d-1}([[r, \infty))) &= \Omega_{d-1}\{\omega : 0 \leq \omega_1 \leq \zeta/r\} = \int_0^{\zeta/r} \Omega_{d-2}(1-x^2)^{(d-3)/2} dx \\ &< \Omega_{d-2}\zeta_0/r \leq \delta\Omega_{d-1}/2 \end{aligned}$$

by definition of C_d above (26). Lemma 1 is proved. \square

7. Gamma and normal tail/density ratios. We begin with normal probabilities. Let $M(x) = \Phi(-x)/\varphi(x)$ for $x \geq 0$ (M is sometimes called Mills' ratio).

LEMMA 2. For all $x \geq 0$ and $0 \leq \delta \leq 1$, $e^{-\delta} \leq M(x(1 + \delta))/M(x) \leq 1$.

PROOF. For $x = 0$ the result is clear. It is easily shown that

$$(47) \quad M'(x) = -1 + xM(x)$$

since $\varphi'(x) = -x\varphi(x)$. By, for example, Lemma 12.1.6(a) in Dudley (1993), we have for all $x > 0$ that $\Phi(-x) \leq \varphi(x)/x$ and so $M(x) \leq 1/x$. It follows that $M'(x) \leq 0$ for all $x > 0$. So the second inequality is proved. For the first, and $x > 0$, $M(x)$ can be written as Laplace's continued fraction [Wall (1948), 92.15],

$$M(x) = 1/(x + 1/(x + 2/(x + 3/(x + \dots)))$$

So again $M(x) < 1/x$, and now we get

$$(48) \quad M(x) > 1/\left(x + \frac{1}{x}\right) = x/(x^2 + 1).$$

It follows that $[\log M(x)]' = M'(x)/M(x) = [-1 + xM(x)]/M(x) \geq -1/x$, so

$$[\log M(\cdot)]_x^{x+\delta} \geq -\int_x^{x+\delta} \frac{1}{t} dt = -\log[1 + \delta] \geq -\delta.$$

This completes the proof of the first inequality, and the proof of Lemma 2. \square

For $\alpha > 0$ and $s > 0$, let

$$(49) \quad M_\alpha(s) := \int_s^\infty x^{\alpha-1} e^{-x} dx / (s^{\alpha-1} e^{-s}).$$

LEMMA 3. (a) For any $\alpha \geq 1$, $s > 0$ and $\delta > 0$, we have $M'_\alpha(s) \leq 0$, $M_\alpha(s) \geq 1$ and $e^{-\delta(\alpha-1)} \leq M_\alpha(s(1 + \delta))/M_\alpha(s) \leq 1$.

(b) For $\alpha = 1/2$, $s > 0$ and $0 < \delta \leq 1$, we have $M'_{1/2}(s) \geq 0$ and $1 \leq M_{1/2}(s(1 + \delta))/M_{1/2}(s) \leq (1 + \delta)^{1/2}$.

PROOF. For (a), note first that $M_1 \equiv 1$. So assume $\alpha > 1$. Using $x^{\alpha-1} \geq s^{\alpha-1}$ for $x \geq s$ it follows that $M_\alpha(s) \geq 1$ for all $s > 0$. We have $M'_\alpha(s) = -1 + M_\alpha(s)[1 - (\frac{\alpha-1}{s})]$. An integration by parts shows that $M_\alpha(s) \leq 1 + (\frac{\alpha-1}{s})M_\alpha(s)$. So $M'_\alpha(s) \leq 0$. This proves the first part of (a). Now $d \log(M_\alpha(s))/ds = -\frac{1}{M_\alpha(s)} + 1 - (\frac{\alpha-1}{s}) \leq 0$. So, $\frac{d}{ds}[\log(M_\alpha(s))] \geq -(\frac{\alpha-1}{s})$ since $M_\alpha(s) \geq 1$. Thus,

$$\int_s^{s+\delta s} \frac{d}{dt}[\log(M_\alpha(t))] dt \geq -(\alpha - 1) \log(1 + \delta) \geq -(\alpha - 1)\delta.$$

So $M_\alpha(s + \delta s)/M_\alpha(s) \geq e^{-\delta(\alpha-1)}$. This completes the proof of (a).

For (b), we have $M_{1/2}(s) \equiv \sqrt{2s}M(\sqrt{2s})$. So by (47), with $u := (2x)^{1/2}$, $M'_{1/2}(x) = [M(u)(1 + u^2) - u]/u$. Since $M(u) > u/(1 + u^2)$ by (48), $M'_{1/2}(x) \geq 0$. This proves the first part of (b) and the first inequality of the second part. For the last inequality, Lemma 2 gives $M_{1/2}((1 + \delta)s) \leq [2s(1 + \delta)]^{1/2}M([2s]^{1/2})$, so $M_{1/2}((1 + \delta)s) \leq (1 + \delta)^{1/2}M_{1/2}(s)$. \square

Next, let $a > 0$, and let X_a be a random variable distributed according to a Γ_a distribution with density $x^{a-1}e^{-x}/\Gamma(a)$ for $x > 0$. We have:

LEMMA 4. For any $a > 0$, (a) $M_a(s) \rightarrow 1$ as $s \rightarrow \infty$.

(b) For α fixed in $(0, 1)$ and $0 \leq x \leq \alpha y$, as $y \rightarrow \infty$, $P(X_a \geq y)/P(X_a \geq x) \rightarrow 0$ uniformly in such x .

PROOF. (a) The definition of M_a (49) and integration by parts give

$$\int_s^\infty x^{\alpha-1}e^{-x} dx = s^{\alpha-1}e^{-s} + (\alpha - 1) \int_s^\infty x^{\alpha-2}e^{-x} dx,$$

and (a) follows since $\int_s^\infty x^{\alpha-2}e^{-x} dx \leq s^{-1} \int_s^\infty x^{\alpha-1}e^{-x} dx$.

(b) By (a) it suffices to note that as $y \rightarrow \infty$, $y^{a-1}e^{-y}/[(\alpha y)^{a-1}e^{-\alpha y}] = \alpha^{1-a} \times e^{(\alpha-1)y} \rightarrow 0$. \square

8. The rest of the proof. In this section we will finish the proof of Theorem 1. The following bounds the inner integral in (46) via gamma probabilities:

LEMMA 5. Let $U_v := (0, v)$ with $v > 0$ and $g : U_v \mapsto \mathbf{R}$. Suppose that for some $0 < \delta \leq 1$, for all t in U_v , $1/(1 + \delta) \leq g''(t) \leq 1 + \delta$ and assume that g'' and thus g and g' can be extended to be continuous on $[0, v]$, with $g(0) = g'(0) = 0$. Let $\pi(\cdot)$ be measurable on U_v and such that for some $\eta > 0$, we have $\eta \leq \pi(t) \leq \eta(1 + \delta)$ for all $t \in U_v$. For $0 \leq a < v$, $n = 1, 2, \dots$ and $k = 0, 1, 2, \dots$, let

$$(50) \quad \begin{aligned} I_{n,a,k} &:= I_{n,a,k,v} := \int_a^v e^{-ng(t)}\pi(t)t^k dt, \\ \Gamma_{n,a,k} &:= \eta e^{-ng(a)}a^{k-1}M_{(k+1)/2}(na^2/2)n^{-1}. \end{aligned}$$

We then have $I_{n,a,k,v} \leq \Gamma_{n,a,k}(1 + \delta)^{k+2}$, and $I_{n,a,k,v}$ is bounded below by

$$\frac{\Gamma_{n,a,k}}{(1 + \delta)^{k+1}} - \frac{\eta e^{-ng(a)}}{n\sqrt{1 + \delta}}v^{k-1} \exp\left[\frac{n}{2}(1 + \delta)(a^2 - v^2)\right]M_{(k+1)/2}\left(\frac{nv^2}{2}\right).$$

PROOF. We have that $t/(1 + \delta) \leq g'(t) \leq t(1 + \delta)$, thus $t^2/[2(1 + \delta)] \leq g(t) \leq t^2(1 + \delta)/2$ for all t in U_v . Now for $0 \leq a \leq t \leq v$,

$$g(t) = g(a) + \int_a^t g'(s) ds \geq g(a) + \int_a^t \frac{s}{1 + \delta} ds = g(a) + \frac{(t^2 - a^2)}{2(1 + \delta)},$$

so

$$I_{n,\alpha,k,v} \leq \eta(1 + \delta) \int_a^\infty e^{-ng(a)} \exp\left[-\frac{n(t^2 - a^2)}{2(1 + \delta)}\right] t^k dt.$$

For any $\lambda > 0$ let

$$(51) \quad J_{a,\lambda,k} := \int_a^\infty \exp(-\lambda t^2) t^k dt.$$

Then

$$(52) \quad J_{a,\lambda,k} = (2\lambda)^{-1} a^{k-1} \exp(-\lambda a^2) M_{(k+1)/2}(\lambda a^2)$$

by the change of variables $u = \lambda t^2$. Set $\lambda_1 := n/[2(1 + \delta)]$. It follows that

$$(53) \quad I_{n,\alpha,k,v} \leq \eta(1 + \delta) \exp\left[-ng(a) + \frac{na^2}{2(1 + \delta)}\right] J_{a,\lambda_1,k}$$

and

$$(54) \quad J_{a,\lambda_1,k} = \frac{(1 + \delta)}{n} a^{k-1} \exp\left[-\frac{na^2}{2(1 + \delta)}\right] M_{(k+1)/2}\left(\frac{n}{2(1 + \delta)} a^2\right).$$

Next,

$$(55) \quad e^\delta \leq (1 + \delta)^2 \quad \text{for } 0 \leq \delta \leq 1,$$

as is easily checked. Set

$$(56) \quad s := \lambda_1 a^2, \quad \alpha := (k + 1)/2$$

and recall that $\lambda_1 = n/[2(1 + \delta)]$. Then $s(1 + \delta) = na^2/2$.

By Lemma 3(a), for $k \geq 1$ [so that $\alpha \geq 1$ by (56)], then by (55), we have

$$M_\alpha(s) \leq M_\alpha(na^2/2) e^{\delta(\alpha-1)} = M_\alpha(na^2/2) e^{\delta(k-1)/2} \leq M_\alpha(na^2/2) (1 + \delta)^{k-1}.$$

If $k = 0$, $M_{1/2}(s) \leq M_{1/2}(na^2/2)$ by Lemma 3(b). So, in all cases, for $k = 0, 1, 2, \dots$,

$$M_{(k+1)/2}(\lambda_1 a^2) \leq M_{(k+1)/2}(na^2/2) (1 + \delta)^k,$$

and by (54),

$$J_{a,\lambda_1,k} \leq (1 + \delta)^{k+1} n^{-1} a^{k-1} \exp\left[-\frac{na^2}{2(1 + \delta)}\right] M_{(k+1)/2}(na^2/2).$$

Thus by (53) and (50),

$$I_{n,a,k,v} \leq \eta(1 + \delta)^{k+2} \exp[-ng(a)] n^{-1} a^{k-1} M_{(k+1)/2}(na^2/2) = (1 + \delta)^{k+2} \Gamma_{n,a,k},$$

proving the first statement in Lemma 5.

For the lower bound, we have by the first line of the proof,

$$\begin{aligned} g(t) &= g(a) + \int_a^t g'(x) dx \leq g(a) + \left(\int_a^t x dx \right) (1 + \delta) \\ &= g(a) + (t^2 - a^2)(1 + \delta)/2. \end{aligned}$$

So

$$I_{n,a,k,v} \geq \eta \exp[-ng(a) + na^2(1 + \delta)/2] \int_a^v t^k e^{-(1+\delta)nt^2/2} dt$$

or equivalently, recalling (51) and setting $b := (1 + \delta)n/2$,

$$(57) \quad I_{n,a,k,v} \geq \eta \exp[-ng(a) + na^2(1 + \delta)/2][J_{a,b,k} - J_{v,b,k}].$$

By (52) with $\lambda = b$, we have

$$(58) \quad J_{a,b,k} = \frac{1}{n(1 + \delta)} M_{(k+1)/2}((1 + \delta)na^2/2) a^{k-1} \exp[-na^2(1 + \delta)/2].$$

For $k \geq 1$, $(k + 1)/2 \geq 1$, so by Lemma 3(a),

$$M_{(k+1)/2}((1 + \delta)na^2/2) \geq M_{(k+1)/2}(na^2/2) e^{-\delta(k-1)/2}.$$

Since $0 < \delta < 1$ by (8), we have $e^\delta \leq (1 + \delta)^2$ by (55), and so

$$M_{(k+1)/2}((1 + \delta)na^2/2) \geq M_{(k+1)/2}(na^2/2) \frac{1}{(1 + \delta)^{k-1}}$$

for $k \geq 1$. For $k = 0$, by Lemma 3(b), $M_{1/2}((1 + \delta)na^2/2) \geq M_{1/2}(na^2/2)$. So,

$$(59) \quad J_{a,b,k} \geq (\exp[-(1 + \delta)na^2/2]) a^{k-1} \frac{1}{n} M_{(k+1)/2}(na^2/2) (1 + \delta)^{-k-1}$$

for all $k = 0, 1, \dots$. Now, by (58) with v in place of a , and by both parts of Lemma 3, for $k = 0, 1, \dots$,

$$J_{v,b,k} \leq \frac{v^{k-1}}{n(1 + \delta)} (1 + \delta)^{1/2} M_{(k+1)/2}(nv^2/2) \exp[-(1 + \delta)nv^2/2].$$

By (57), (50) and (59), it follows that $I_{n,a,k,v}$ is bounded below by

$$\frac{\Gamma_{n,a,k}}{(1 + \delta)^{k+1}} - \frac{\eta}{n} e^{-ng(a)} v^{k-1} \exp[n(1 + \delta)(a^2 - v^2)/2] M_{(k+1)/2}(nv^2/2) / (1 + \delta)^{1/2}.$$

This completes the proof of Lemma 5. \square

PROOF OF PROPOSITION 1, ASSUMING PROPOSITION 2. We have for $0 < \zeta \leq \rho$ that $\|y\| \leq (1 + (\delta/6))\|\theta\| \leq 2\|\theta\|$ as noted just after (32). So $U(\rho)^c \subset U'(\rho/2)^c$ and it will suffice to prove the statement about $U'(\rho/2)^c$. Almost surely $\|\hat{\theta}_n - \theta_0\| < \rho/8$ for $n \geq n_0(\xi, \rho/8)$ by (16), so that in the θ coordinates as finally chosen, $\|\theta_0\| < \rho/8$. Let $\mathcal{B} := B(\theta_0, \rho/4)$ in the θ coordinates. Then

$U'(\rho/2)^c \subset \mathcal{B}^c$. By assumption (A₅), there is a $\kappa > 0$ such that a.s. for n large enough, specifically $n \geq n_2(\xi, \rho)$, where we can take $n_2(\xi, \rho) \geq n_0(\xi, \rho/8)$, $\sup_{\theta \notin \mathcal{B}} LL_n(\theta) < \sup_{\theta \in \mathcal{B}} LL_n(\theta) - n\kappa$. Thus by (20), $\inf_{\theta \notin \mathcal{B}} H(\theta) \geq \kappa$, and by (32) and (40), $I_n(\mathcal{B}^c) \leq e^{-n\kappa}$. Also, $\|\theta\| \leq 2\|y\|$ for $0 < \zeta \leq \rho$ by (33) so $\mathcal{B} \supset U'(\rho/8) \supset U(\rho/16)$. Thus we have a.s. for $n \geq n_4$ for some $n_4(\xi, \rho, \gamma_1, \delta, d)$,

$$I_n(\mathcal{B}) \geq I_n(U'(\rho/8)) \geq \eta(2\pi/n)^{d/2}(1 + \delta)^{-(3d+10)/2}$$

by (42) for $U(\rho/16)$. Thus by (41) for $C = \mathcal{B}$,

$$\begin{aligned} \pi_{x,n}(U(\rho)^c) &\leq \pi_{x,n}(\mathcal{B}^c) = \frac{I_n(\mathcal{B}^c)}{I_n(\Theta)} = \frac{I_n(\mathcal{B}^c)}{I_n(\mathcal{B}) + I_n(\mathcal{B}^c)} \\ &\leq \frac{e^{-n\kappa}}{I_n(\mathcal{B})} \leq e^{-n\kappa}(n/2\pi)^{d/2}(1 + \delta)^{(3d+10)/2}\eta^{-1}. \end{aligned}$$

We have $(1 + \delta)^{(3d+10)/2} < 1 + \varepsilon < 2$ by (8), and $2(n/(2\pi))^{d/2} \leq \eta e^{n\kappa/2}$ for $n \geq n_5$ for some $n_5 = n_5(d, \eta, \kappa)$. Proposition 1 then holds with $\nu = \kappa/2$ and $n_3 := \max(n_2, n_4, n_5)$. \square

PROOF OF PROPOSITION 2, ASSUMING PROPOSITION 3. For $n \geq \max(n_0(\xi, \gamma_1), n_1(\delta, \rho, d))$, the bounds (43) and (44) hold uniformly in $0 < \zeta \leq \zeta_0$ and then we can let $\zeta \rightarrow 0$ to obtain the two inequalities when $\zeta = 0$ and $\hat{\theta}_n \in \partial A$. Consider the half-space $A_1 := \{\theta_1 \geq 0\} = \{y_1 \geq 0\}$ by (31). Adding inequalities for A_1 and its complement then gives (42) (which could be proved directly without half-spaces), proving Proposition 2. \square

PROOF OF PROPOSITION 3. Recall ρ as defined in (25) and D_ω and h_ω as defined before (45). By Lemma 5, applied to $k = d - 1$, with $g(t) = h_n(t\omega)$ in light of (34), since $\eta \leq \pi(t\omega) \leq \eta(1 + \delta)$ for $t\omega \in U(\rho)$ by (35), we have, for ω such that $D_\omega \leq \rho$,

$$\begin{aligned} (60) \quad &\int_{D_\omega}^\rho \exp[-nh_n(t\omega)]\pi(t\omega)t^{d-1} dt \\ &\leq \eta \exp[-nh_n(D_\omega\omega)](D_\omega)^{d-2}M_{d/2}(nD_\omega^2/2)n^{-1}(1 + \delta)^{d+1}. \end{aligned}$$

Since $h_n(D_\omega\omega) \geq \tau + h_\omega^2/[2(1 + \delta)]$ by (38), we have by (46) and (60) that $I_n(A \cap U(\rho))$ is less than or equal to

$$(61) \quad \frac{\eta}{n}(1 + \delta)^{d+1}e^{-n\tau} \int_{S(\zeta, \rho)} \exp\left[-\frac{nh_\omega^2}{2(1 + \delta)}\right]D_\omega^{d-2}M_{d/2}\left(\frac{nD_\omega^2}{2}\right) d\omega.$$

Let $Y(\omega) := Y(\omega, n, d, \delta) := D_\omega^{d-2}M_{d/2}(nD_\omega^2/2)/(1 + \delta)^d$ and

$$\begin{aligned} Z(\omega) &:= Z(\omega, n, d, \delta, \rho) \\ &:= \rho^{d-2} \exp[n(1 + \delta)(D_\omega^2 - \rho^2)/2]M_{d/2}(n\rho^2/2)/(1 + \delta)^{1/2}. \end{aligned}$$

Then by Lemma 5 and (32),

$$(62) \quad \int_{D_\omega}^\rho e^{-nh(t\omega)} \pi(t\omega) t^{d-1} dt \geq \frac{\eta}{n} \exp[-nh(D_\omega\omega)] \{Y(\omega) - Z(\omega)\}.$$

Recalling $r := \rho/2$ after (25) and $U(\rho) := \{y : \|y\| < \rho\}$ after (35), we have $I_n(A \cap U(\rho)) \geq I_n(A \cap U(\rho) \cap \{D_\omega \leq r\})$, and for $S(\zeta, \rho)$ defined before (46) and D_ω before (45), $S(\zeta, \rho) \cap \{D_\omega \leq r\} = S(\zeta, r)$, so by (46) and (62),

$$(63) \quad I_n(A \cap U(\rho)) \geq \eta n^{-1} \int_{S(\zeta, r)} \exp[-nh(D_\omega\omega)] [Y(\omega) - Z(\omega)] d\omega.$$

We have $D_\omega \geq \zeta > 0$ for all $\omega \in S(d+)$. Next,

$$(64) \quad \frac{Z(\omega)}{Y(\omega)} = \frac{\rho^{d-2} \exp[-\frac{\eta}{2}(1+\delta)\rho^2] M_{d/2}(n\rho^2/2)}{D_\omega^{d-2} \exp[-\frac{\eta}{2}(1+\delta)D_\omega^2] M_{d/2}(nD_\omega^2/2)} (1+\delta)^{d-1/2},$$

which from the definition (49) of M_α is easily seen to equal a product $T_1 T_2 T_3$ where $0 < T_1 := \exp[n\delta(D_\omega^2 - \rho^2)/2]$, $T_2 := (1+\delta)^{d-1/2}$ and since $D_\omega < \rho$ in (63),

$$(65) \quad T_3 := \frac{P(X_{d/2} \geq n\rho^2/2)}{P(X_{d/2} \geq nD_\omega^2/2)} < 1,$$

where $X_{d/2}$ is a gamma random variable with density $x^{(d-2)/2} e^{-x} / \Gamma(d/2)$ for $x > 0$. Then for $0 < D_\omega \leq r = \rho/2$ we have $T_1 T_2 \leq 1$ for n large enough, $n \geq N_0(\delta, \rho, d)$, and then $Y(\omega) \geq Z(\omega)$.

Also, $h(\zeta, X) \leq \tau + (1+\delta)\|X\|^2/2$ by (38) for $(\zeta, X) = D_\omega\omega$. So for $n \geq N_0(\delta, \rho, d)$, by (45) and (63), $I_n(A \cap U(\rho))$ is greater than or equal to

$$(66) \quad \eta n^{-1} e^{-n\tau} \int_{S(\zeta, r)} \exp[-n(1+\delta)h_\omega^2/2] [Y(\omega) - Z(\omega)] d\omega.$$

By Lemma 4(b), we have, for any fixed $d \geq 1$, uniformly for $D_\omega \leq r = \rho/2$,

$$(1+\delta)^{d-1/2} \frac{P(X_{d/2} \geq n\rho^2/2)}{P(X_{d/2} \geq nD_\omega^2/2)} \rightarrow 0$$

as $n \rightarrow \infty$, recalling $0 < \delta < 1$. So for $n \geq n_1(\delta, \rho, d) \geq N_0(\delta, \rho, d)$ large enough,

$$(1+\delta)^{d-1/2} \frac{P(X_{d/2} \geq n\rho^2/2)}{P(X_{d/2} \geq nD_\omega^2/2)} < \frac{\delta}{2}$$

for $0 < D_\omega \leq r$. It follows by (66), (64), (65) and the definition of $Y(\omega)$ after (60) that $I_n(A \cap U(\rho))$ is bounded below by

$$\eta n^{-1} e^{-n\tau} \int_{S(\zeta, r)} \exp[-n(1+\delta)h_\omega^2/2] \frac{D_\omega^{d-2} M_{d/2}(nD_\omega^2/2)}{(1+\delta)^d} \left(1 - \frac{\delta}{2}\right) d\omega.$$

As is easily shown, for $0 \leq \delta < 1$, we have $1 - (\delta/2) \geq 1/(1 + \delta)$, so for $n \geq n_1(\delta, \rho, d)$, $I_n(A \cap U(\rho))$ is greater than or equal to

$$(67) \quad \eta n^{-1} e^{-n\tau} \int_{S(\zeta, r)} \exp[-n(1 + \delta)h_\omega^2/2] \frac{D_\omega^{d-2} M_{d/2}(nD_\omega^2/2)}{(1 + \delta)^{d+1}} d\omega.$$

Recalling again (45) that $D_\omega^2 = \zeta^2 + h_\omega^2$, the integrand in (67) equals

$$(68) \quad I(\zeta, \omega) := I_{n,d}(\zeta, \omega) := (1 + \delta)^{-d-1} \exp[n\zeta^2(1 + \delta)/2] f(D_\omega)$$

with $f(x) := \exp[-nx^2(1 + \delta)/2] x^{d-2} M_{d/2}(nx^2/2)$. Then

$$f(x) = \exp[-nx^2\delta/2] \Gamma(d/2) (2/n)^{(d-2)/2} P(X_{d/2} \geq nx^2/2),$$

so $f(x)$ is a decreasing function of x . Let $\omega(r)$ be a direction ω such that $D_{\omega(r)} = r$ (choose one such direction). Recall that $[[r, \infty)) = \{D_\omega \geq r\}$, $[[0, r)) = \{D_\omega < r\}$. The function $I(\zeta, \omega)$ depends on ω only through D_ω and is decreasing in D_ω since $f(x)$ is decreasing, so

$$(69) \quad \int_{[[r, \infty))} I(\zeta, \omega) d\omega \leq I(\zeta, \omega(r)) \Omega_{d-1}([[r, \infty))$$

and

$$(70) \quad \int_{[[0, r))} I(\zeta, \omega) d\omega \geq I(\zeta, \omega(r)) \Omega_{d-1}([[0, r)).$$

Now by Lemma 1 and (69),

$$\int_{S(d+)} I(\zeta, \omega) d\omega \leq \int_{[[0, r))} I(\zeta, \omega) d\omega + I(\zeta, \omega(r)) \delta \Omega_{d-1}/2.$$

So by (70),

$$\begin{aligned} & \int_{S(d+)} I(\zeta, \omega) d\omega \\ & \leq \int_{[[0, r))} I(\zeta, \omega) d\omega + \frac{\delta}{2} \Omega_{d-1} \int_{[[0, r))} I(\zeta, \omega) d\omega / \Omega_{d-1}([[0, r)) \\ & < \int_{[[0, r))} I(\zeta, \omega) d\omega \left(1 + \frac{\delta}{1 - \delta}\right), \end{aligned}$$

since $\Omega_{d-1}([[0, r)) > (1 - \delta)\Omega_{d-1}/2$ by Lemma 1. Since $\delta < 1/10$, it is clear that $1 + \delta/(1 - \delta) = 1/(1 - \delta) < (1 + \delta)^2$. So

$$\int_{S(d+)} I(\zeta, \omega) d\omega \leq \left[\int_{[[0, r))} I(\zeta, \omega) d\omega \right] (1 + \delta)^2.$$

Thus, returning to (67), then since $I(\zeta, \omega)$ is the integrand in it,

$$\begin{aligned}
 I_n(A \cap U(\rho)) &\geq \eta n^{-1} e^{-n\tau} \int_{[[0,r))} I(\zeta, \omega) d\omega \\
 (71) \quad &\geq \eta n^{-1} e^{-n\tau} \int_{S(d+)} I(\zeta, \omega) d\omega / (1 + \delta)^2 \\
 &= \frac{\eta n^{-1} e^{-n\tau}}{(1 + \delta)^{d+3}} \int_{S(d+)} \exp[-n(1 + \delta)h_\omega^2/2] D_\omega^{d-2} M_{d/2}(nD_\omega^2/2) d\omega.
 \end{aligned}$$

Then by (49), (45) and a few calculations, the integral in (71) equals

$$(72) \quad n e^{[n\zeta^2(1+\delta)/2]} \int_{S(d+)} \exp[-n\delta D_\omega^2/2] \int_{D_\omega}^\infty u^{d-1} e^{(-nu^2/2)} du d\omega.$$

Returning now to Cartesian coordinates, we have $u^{d-1} du d\omega = dy$ with $u = \|y\|$, so the integral in (71) equals, where $dy := dy_d \cdots dy_1$,

$$\begin{aligned}
 n \exp[n\zeta^2(1 + \delta)/2] &\int_\zeta^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp[-n\|y\|^2/2] \exp[-n\delta\|y\|^2\zeta^2/(2y_1^2)] dy \\
 &\geq n \exp[n\zeta^2(1 + \delta)/2] \int_\zeta^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp[-n(1 + \delta)\|y\|^2/2] dy_d \cdots dy_1 \\
 &= (1 + \delta)^{-d/2} n^{(2-d)/2} (2\pi)^{(d-1)/2} M(\zeta\sqrt{n(1 + \delta)})
 \end{aligned}$$

(after a few calculations). So from (71),

$$(73) \quad I_n(A \cap U(\rho)) \geq \eta n^{-d/2} e^{-n\tau} (1 + \delta)^{-(3d+6)/2} (2\pi)^{(d-1)/2} M(\zeta\sqrt{n(1 + \delta)}).$$

It follows from (36) and the definition (37) of τ that $\zeta^2/[2(1 + \delta)] \leq \tau = h(\zeta, 0) \leq (1 + \delta)\zeta^2/2$ (since $(\zeta, 0) \in U(\rho)$), and therefore

$$(74) \quad \sqrt{n}\zeta/\sqrt{1 + \delta} \leq \sqrt{2n\tau} \leq \sqrt{n(1 + \delta)}\zeta.$$

So by Lemma 2,

$$M(\zeta[n(1 + \delta)]^{1/2}) \geq e^{-\delta} M(\sqrt{n}\zeta/\sqrt{1 + \delta}) \geq e^{-\delta} M(\sqrt{2n\tau}),$$

since M is a nonincreasing function, so

$$I_n(A \cap U(\rho)) \geq \eta n^{-d/2} e^{-n\tau} (1 + \delta)^{-(3d+6)/2} (2\pi)^{(d-1)/2} e^{-\delta} M(\sqrt{2n\tau}),$$

thus by definition of $M(\cdot)$ before Lemma 2,

$$I_n(A \cap U(\rho)) \geq \eta(2\pi/n)^{d/2} (1 + \delta)^{-(3d+6)/2} \Phi(-\sqrt{2n\tau}) e^{-\delta}.$$

By (55), for $0 \leq \delta \leq 1$, $e^{-\delta} \geq 1/(1 + \delta)^2$; thus (43) follows.

Turning to the upper bound (61), consider the integral

$$I := \int_{S(\zeta,\rho)} \exp[-nh_\omega^2/[2(1 + \delta)]] D_\omega^{d-2} M_{d/2}(nD_\omega^2/2) d\omega.$$

A few calculations and (49) show that as in (72),

$$I = n \exp\left[\frac{n\zeta^2}{2(1+\delta)}\right] \int_{S(\zeta,\rho)} \exp\left[\frac{\delta}{1+\delta} D_\omega^2 \frac{n}{2}\right] \int_{D_\omega}^\infty u^{d-1} \exp[-nu^2/2] du d\omega.$$

Returning to Cartesian coordinates, we have $dy = u^{d-1} du d\omega$ and $u = \|y\|$. Then for $dy := dy_d \cdots dy_1$,

$$\begin{aligned} I &\leq n \exp\left[\frac{n\zeta^2}{2(1+\delta)}\right] \int_\zeta^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp\left[\frac{\delta}{1+\delta} \frac{n}{2} \frac{\|y\|^2 \zeta^2}{y_1^2}\right] \exp\left[-\frac{n}{2} \|y\|^2\right] dy \\ &\leq n^{(2-d)/2} (1+\delta)^{(d-1)/2} (2\pi)^{(d-1)/2} M(\zeta\sqrt{n}) \end{aligned}$$

after a few calculations, using $y_j^2 \zeta^2 / y_1^2 \leq y_j^2$ for $j = 2, \dots, d$. Altogether, using (61),

$$I_n(A \cap U(\rho)) \leq n^{-1} \eta (1+\delta)^{d+1} e^{-n\tau} n^{(2-d)/2} (1+\delta)^{d/2} (2\pi)^{(d-1)/2} M(\zeta\sqrt{n}).$$

Now, since $\sqrt{2n\tau} \leq \sqrt{n}(1+\delta)\zeta$ by (74) and by Lemma 2, we have

$$M(\zeta\sqrt{n}) \leq M(\zeta\sqrt{n}(1+\delta))e^\delta \leq M(\sqrt{2n\tau})e^\delta.$$

So again by (55), (44) follows and Proposition 3 is proved. \square

PROOF OF THEOREM 1. Continuing with Case I, we have $\hat{\theta}_n \notin A$ and $\zeta < \zeta_0 < \rho$. By Proposition 1, there exists $\nu > 0$ such that by (41), for $n \geq n_3$,

$$\begin{aligned} \pi_{x,n}(A) &= \pi_{x,n}(A \cap U(\rho)) + \pi_{x,n}(A \cap U(\rho)^c) \leq \pi_{x,n}(A \cap U(\rho)) + e^{-n\nu} \\ &= I_n(A \cap U(\rho)) / I_n(\Theta) + e^{-n\nu} \leq \frac{I_n(A \cap U(\rho))}{I_n(U(\rho))} + e^{-n\nu}. \end{aligned}$$

Since $\zeta < \rho$ then by Proposition 3 (44) and Proposition 2 (42), for $n \geq n_3$,

$$(75) \quad \pi_{x,n}(A) \leq (1+\delta)^{3d+8} \Phi(-\sqrt{2n\tau}) + e^{-n\nu}.$$

We now distinguish two subcases, subcase (i) where $e^{-n\nu} < \delta \Phi(-\sqrt{2n\tau})$ and subcase (ii) otherwise. In subcase (ii), $\Phi(-\sqrt{2n\tau}) \leq \delta^{-1} e^{-n\nu}$ and

$$\pi_{x,n}(A) \leq [(1+\delta)^{3d+8} \delta^{-1} + 1] e^{-n\nu} \leq e^{-n\nu/2}$$

for n large enough so that $e^{n\nu/2} \geq (1+\delta)^{3d+8} \delta^{-1} + 1$, in addition to $n \geq n_3$. Thus in subcase (ii), we have (b) in Theorem 1 for any $\kappa' \leq \nu/2$.

In subcase (i), by (75), (8) and (39),

$$(76) \quad \pi_{x,n}(A) \leq (1+\delta)^{3d+9} \Phi(-\sqrt{2n\tau}) \leq (1+\varepsilon) \Phi(-\sqrt{2\Delta}),$$

so that the upper bound in (a) of Theorem 1 holds.

Now to treat lower bounds for $\pi_{x,n}(A)$, by (41),

$$\pi_{x,n}(A) \geq \pi_{x,n}(A \cap U(\rho)) = I_n(A \cap U(\rho)) / I_n(\Theta).$$

Note that a.s. for $n \geq n_3$ by Proposition 1 and (42),

$$\begin{aligned} I_n(\Theta) &= I_n(U(\rho)) + I_n(U(\rho)^c) \leq I_n(U(\rho)) + e^{-nv} \\ &\leq \eta(2\pi/n)^{d/2}(1+\delta)^{(3d+6)/2} + e^{-nv} \leq \eta(2\pi/n)^{d/2}(1+\delta)^{(3d+7)/2}, \end{aligned}$$

for n large enough, say, $n \geq \max(n_3, n_6)$ for some $n_6(v, \eta, d)$. Then by (43), (39) and (8),

$$(77) \quad \pi_{x,n}(A) \geq (1+\delta)^{-(3d+9)} \Phi(-\sqrt{2n\tau}) \geq \Phi(-\sqrt{2\Delta})/(1+\varepsilon).$$

Thus by (76) conclusion (a) of Theorem 1 is proved in Case I(i).

To prove Theorem 1 when $\hat{\theta}_n \in \partial A$ we can, as in the proof of Proposition 2, let $\zeta \downarrow 0$ in Proposition 3, then apply the same argument as for Case I(i).

In Theorem 1, the statement about U and $A \subset U^c$ follows from Proposition 1. For the statement about A_n , since $d(A_n, \phi_0) \rightarrow 0$ and $d(\hat{\phi}_n, \phi_0) \rightarrow 0$ we have by (37), (20), (23), noting that $h(\zeta, 0) \leq H(\zeta, 0)$, and (32) that $\tau = \tau_n \rightarrow 0$ as $n \rightarrow \infty$. Thus we have subcase (i) [after (75)]. We apply Propositions 1–3 and (8) to get that $\pi_{x,n}(A_n)/\Phi(-\sqrt{2\Delta_n}) \doteq 1$ within a fixed power of $1 + \varepsilon$ for n large enough. Letting $\rho \downarrow 0$ and thus $\varepsilon \downarrow 0$, we get $\pi_{x,n}(A_n)/\Phi(-\sqrt{2n\tau_n}) \rightarrow 1$ as $n \rightarrow \infty$. This finishes the proof of Theorem 1 in Case I for $d \geq 2$.

In Case II (29), $\zeta \geq \zeta_0$. Let $B := \{\theta : \theta_1 \geq \zeta_0\}$ in the coordinates as in (18). Recalling that, for B , by (37) and (39) for $\zeta = \zeta_0$, $\Delta_{B,n} = nh(\zeta_0, 0)$, we have by (A5), $\Phi(-\sqrt{2\Delta_{B,n}}) \leq \exp(-n\kappa)$ for some $\kappa > 0$ and n large enough. Then also for A , $\Phi(-\sqrt{2\Delta}) \leq e^{-n\kappa}$, uniformly for $\zeta \geq \zeta_0$. B is disjoint from $U'(\zeta_0)$, so by Proposition 1, for some $v' > 0$, a.s. for n large enough, $\pi_{x,n}(A) \leq \pi_{x,n}(U'(\zeta_0)^c) \leq e^{-v'n}$ uniformly over all A with $\zeta \geq \zeta_0$. So in Case II we have conclusion (b) for $\kappa' = \min(v', \kappa)$. The proof of Theorem 1 for $d \geq 2$ is complete.

For $d = 1$ the proof is somewhat simpler and is omitted because there is no transformation to spherical coordinates and we do not need Lemma 1. This completes the proof of Theorem 1. \square

9. A multivariate normal location family. For this family, explicit calculations can be done. We will see how (a) for a normal prior, slowness of approach of the half-space to ϕ_0 can cause slow convergence to 0 of the relative error of our simplest approximation (to some small probabilities), and (b) for a double-exponential prior, the absolute error for some half-spaces is no smaller than $O(1/\sqrt{n})$, even for the approximation sharpened via (1).

Consider the location family $N(\mu, I)$, $\mu \in \mathbf{R}^d$, on \mathbf{R}^d with, first, a prior $N(0, \sigma^2 I)$ for μ . (In this case the Jeffreys prior, Lebesgue measure, is improper.) Let $x = (X_1, \dots, X_n)$ be observed i.i.d. $N(\mu, I)$. Let $S_n := X_1 + \dots + X_n$ and $Q_n := \sum_{j=1}^n |X_j|^2$. Let A be a half-space $A := \{\mu \cdot t \geq c\}$ with $|t| = 1$. A brief calculation shows that the posterior probability is $\pi_{x,n}(A) = \Phi(\tau_n S_n \cdot t - c\tau_n^{-1})$ where $\tau_n := (n + \sigma^{-2})^{-1/2}$. Our simple likelihood root approximation to the posterior, provided that $S_n/n \notin A$, is $\Phi(-\sqrt{2\Delta}) = \Phi(n^{-1/2} S_n \cdot t - c\sqrt{n})$. As

$\sigma \rightarrow \infty$, so that the prior becomes more and more diffuse, and converges in a sense to the Jeffreys prior, the exact posterior distribution converges to the likelihood root approximation. If we instead fix, for example, $\sigma = 1$, $\mu = \mu_0 = 0$ and $t = (1, 0, \dots, 0)$ and let $c = c_n = n^{-\alpha}$ for some $\alpha \in (0, 1/2)$, we obtain $\pi_{x,n}(A_n) = \Phi(S_{n,1}/\sqrt{n+1} - n^{-\alpha}\sqrt{n+1})$ and $\Phi(-\sqrt{2\Delta_n}) = \Phi(S_{n,1}/\sqrt{n} - n^{-\alpha}\sqrt{n})$. For $a > 0$ and $0 < \delta < 1$ a short calculation gives

$$(78) \quad \Phi(a + \delta) - \Phi(a) = \phi(a)(1 - e^{-a\delta})(1 + O(\delta^2))/a.$$

We will use the fact that if Z_n are any standard normal variables then $|Z_n| = O(\sqrt{\log n}) = O(n^\varepsilon)$ a.s. as $n \rightarrow \infty$ for any $\varepsilon > 0$. Let $Z_n := S_{n,1}/\sqrt{n}$, $a := a_n := n^{0.5-\alpha} - Z_n$ and $\delta := \delta_n := n^{-\alpha}\sqrt{n+1} - Z_n\sqrt{n}/\sqrt{n+1} - a_n$. Then we have a.s. $\delta_n = \frac{1}{2}n^{-0.5-\alpha} + \frac{Z_n}{2n} + O(n^{-\alpha-3/2})$. Since by l'Hospital's rule and facts stated after (47), $\Phi(-x) \sim \phi(x)/x$ as $x \rightarrow +\infty$, it follows from (78) that the relative error in the likelihood root approximation of the posterior is almost surely asymptotic to $a_n\delta_n \sim n^{-2\alpha}/2$ as $n \rightarrow \infty$, which converges to 0 slowly for small $\alpha > 0$. The probabilities being approximated converge to 0 rather rapidly in this case.

For the approximation π^{**} (4) the factors in terms of $\ell_{\eta\eta}$ and $d\mu_1/dz \equiv n^{-1/2}$ are constants and divide out, so the interesting factor is $\rho_n := \pi_1(\tilde{\theta})/\pi_1(\hat{\theta})$, which a brief calculation shows to be $\exp(-a_n\delta_n + O_p(\delta_n^2))$. The relative error in this case is $r_{SL} := [\Phi(-a_n)\rho_n - \Phi(-a_n - \delta_n)]/\Phi(-a_n - \delta_n)$. By (78) we have

$$\Phi(-a_n - \delta_n) = \Phi(-a_n) - \phi(a_n)(1 - \exp(-a_n\delta_n))(1 + O(\delta_n^2))/a_n,$$

and so the numerator of r_{SL} is asymptotic to

$$[\phi(a_n)a_n^{-1} - \Phi(-a_n)](1 - \exp(-a_n\delta_n)) + O(\delta_n^2)\Phi(-a_n).$$

Since $\Phi(-x) - \phi(x)/x \sim -\phi(x)/x^3$ as $x \rightarrow +\infty$, also by l'Hospital's rule, the numerator is asymptotic to $\phi(a_n)\delta_n/a_n^2$. Since the denominator is asymptotic to $\phi(a_n)/a_n$, we have $r_{SL} \sim \delta_n/a_n \sim 1/(2n)$, not depending on α . But in the examples in the next section, the relative error of π^{**} goes to 0 at a slower rate.

Often, rates of convergence for approximations of posteriors have been stated in the literature without precise assumptions. Under certain conditions on the likelihood functions, it is sufficient for convergence without rates that the prior $\pi(\cdot)$ should be continuous and strictly positive [e.g., Walker (1969), Theorem 1 and Corollary 1 above]. For faster rates such as $O(1/n)$ or $O(n^{-3/2})$, beside stronger assumptions on the likelihood functions, it has been assumed that $\pi(\cdot)$ has continuous partial derivatives through order 4 [e.g., Bickel and Ghosh (1990), Erkanli (1994)]. We will see that for such rates, even for the very smooth normal location likelihood, some smoothness of the prior is needed.

A strictly positive, continuous, in fact Lipschitz, but not differentiable (at 0) prior density on \mathbf{R}^1 is the double-exponential $\pi(\theta) = e^{-|\theta|}/2$, $-\infty < \theta < \infty$. One can do closed-form calculations with this prior and the normal location family as

noted by Pericchi and Smith (1992) and Choy and Smith (1997). Let $d = 2$ and take $N(\phi, I)$, $\phi = (\psi, \eta)$, $S_n = (S_{n1}, S_{n2})$. The profile log likelihood is

$$m(\psi) = -n \log(2\pi) - \frac{1}{2} Q_n + \psi S_{n1} - \frac{n}{2} (\psi^2 - \bar{y}^2),$$

where $\bar{y} := S_{n2}/n$. For the log likelihood ℓ , $\partial^2 \ell / \partial \eta^2 \equiv -n$.

For any product prior density $\pi(\psi, \eta) = \pi_1(\psi)\pi_2(\eta)$, the approximation (1) to the marginal posterior density is exact here. Instead we take the 45 degree rotated double-exponential product prior $\pi(\psi, \eta) = \frac{1}{2} \exp(-|\psi + \eta| - |\psi - \eta|)$. Then π is not differentiable where $\psi = \pm\eta$. The approximation (1) is given by

$$\pi_{\psi|Y}^*(\psi) = c^{**} \exp(-|\psi + \bar{y}| - |\psi - \bar{y}| + \psi S_{n1} - n\psi^2/2),$$

where the normalizing constant c^{**} may depend on \bar{y} , S_{n1} and n but not on ψ . The exact marginal posterior can be evaluated straightforwardly, say, for $\psi \geq 0$, setting

$$\int_{-\infty}^{\infty} d\eta = \left(\int_{-\infty}^{-\psi} + \int_{-\psi}^{\psi} + \int_{\psi}^{\infty} \right) d\eta,$$

as $c \exp(\psi S_{n1} - n\psi^2/2) \tau_n(\psi, \bar{y})$ where c is a constant with respect to ψ and

$$\begin{aligned} \tau_n(\psi, \bar{y}) := & e^{-2\psi} [\Phi(\sqrt{n}(\psi - \bar{y})) - \Phi(-\sqrt{n}(\psi + \bar{y}))] \\ & + e^{2\bar{y}+2/n} \Phi\left(-\sqrt{n}\left(\psi + \bar{y} + \frac{2}{n}\right)\right) \\ & + e^{-2\bar{y}+2/n} \Phi\left(-\sqrt{n}\left(\psi - \bar{y} + \frac{2}{n}\right)\right). \end{aligned}$$

For $u > 0$ we have $1 - \Phi(\sqrt{nu}) = \Phi(-\sqrt{nu}) \rightarrow 0$ exponentially as $n \rightarrow \infty$. Thus the relative error in the approximation of $\tau_n(\psi, \bar{y})$ by $\exp(-|\psi - \bar{y}| - |\psi + \bar{y}|)$ for $\psi = a|\bar{y}|$, $a \geq 0$, $\bar{y} \neq 0$, as $n \rightarrow \infty$, is asymptotic to $-2/n$ for $0 \leq a < 1$ and to $\sqrt{2/(\pi n)}$ for $a = 1$ (and is exponentially small for $a > 1$). Taking a in a $1/\sqrt{n}$ -neighborhood of 1, if for the true parameters $\psi_0 = \eta_0 \neq 0$, then with probability not approaching 0 we have an absolute error in the approximation of posterior probabilities of some half-spaces via (1) of order $1/\sqrt{n}$.

10. Beta numerical results. As an example, we considered the sample space of two points $\{0, 1\}$ and the family of binomial probabilities with $p \in \Theta = (0, 1)$, and the fixed half-space (segment) $p \leq x = 0.7$. We consider in Tables 1 and 3 the Jeffreys prior, which is the $\beta(1/2, 1/2)$ distribution with density $p^{-1/2}(1-p)^{-1/2}/\pi$, $0 < p < 1$, and in Table 2 the uniform prior. If in n independent trials, there are k successes (1's) and $n - k$ failures (0's) then the posterior is the $\beta(k + 1/2, n - k + 1/2)$ distribution in Tables 1 and 3, $\beta(k + 1, n - k + 1)$ in Table 2. Thus the posterior probability of the interval $[0, x]$ in Tables 1 and 3 is

$$J_{x,k,n} := I_{x,k+1/2,n-k+1/2} := \int_0^x t^{k-1/2}(1-t)^{n-k-1/2} dt / B(k+1/2, n-k+1/2),$$

TABLE 1
Jeffreys prior beta examples

n	k	$J_{0.7,k,n}$	r_{cl}	r_{Δ}	r_0	r_1	r_{DM}	r_2
50	44	$1.3735 \cdot 10^{-3}$	-0.940	-0.091	-0.0103	$-8.6 \cdot 10^{-3}$	$-2.8 \cdot 10^{-3}$	$3.05 \cdot 10^{-5}$
100	85	$2.5483 \cdot 10^{-4}$	-0.932	-0.068	-0.0054	$-3.7 \cdot 10^{-3}$	$-8.5 \cdot 10^{-4}$	$3.30 \cdot 10^{-6}$
200	165	$2.5816 \cdot 10^{-5}$	-0.927	-0.053	-0.0030	$-1.7 \cdot 10^{-3}$	$-2.8 \cdot 10^{-4}$	$5.3 \cdot 10^{-7}$
500	394	$4.6722 \cdot 10^{-6}$	-0.834	-0.034	-0.0016	$-6.2 \cdot 10^{-4}$	$-5.8 \cdot 10^{-5}$	$3 \cdot 10^{-8}$
1000	767	$1.1125 \cdot 10^{-6}$	-0.752	-0.025	-0.0010	$-2.9 \cdot 10^{-4}$	$-1.9 \cdot 10^{-5}$	$ \cdot \leq 10^{-9}$
2000	1501	$2.6838 \cdot 10^{-7}$	-0.661	-0.018	-0.0007	$-1.4 \cdot 10^{-4}$	$-6.3 \cdot 10^{-6}$	$ \cdot \leq 10^{-9}$
5000	3671	$4.7368 \cdot 10^{-8}$	-0.534	-0.012	-0.0004	$-5.6 \cdot 10^{-5}$	$-1.5 \cdot 10^{-6}$	$ \cdot \leq 10^{-9}$

where $B(\cdot, \cdot)$ is the beta function. The MLE for p is $\hat{p} = k/n$. Thus, for the MLE to be outside the half-space (in this case, interval) $[0, x]$ we consider $k \geq nx$. For such n, x and k we have the exact beta probability, which we computed by the algorithm of Holt (1986), and our approximation $\Phi_{\Delta} := \Phi_{x,k,n} := \Phi(-\sqrt{2\Delta})$. We evaluated the relative error $r_{\Delta} := \Phi_{x,k,n}/J_{x,k,n} - 1$. For comparison, we took a classical approximation by a normal distribution with mean \hat{p} and variance such that the second derivative of the log of its density (a constant) equals the second derivative of the log of the posterior density at \hat{p} . This gave an approximation we call Φ_{cl} with a relative error r_{cl} . [Taking the posterior mode $(k - 1/2)/(n - 1)$ in place of \hat{p} or the second derivative of the log likelihood gave worse approximations in Table 1.] Also, in all three tables, we consider the sharpened approximations $\pi_Y^{i,*}$ and $\pi_Y^{DM,*}$ defined in Section 3 to the posterior probabilities and give their relative errors r_i for $i = 0, 1, 2$, and r_{DM} . The relative error r_2 is the smallest in each row and r_{DM} is usually next smallest.

Table 3 treats a “large deviation” case in which $x = 0.7$ and $k/n = 0.74$ both remaining constant as n becomes large. Here the relative error r_{Δ} of the simple likelihood root approximation may approach a nonzero constant, which would not contradict our theorem. For all the sharpened approximations, the relative errors

TABLE 2
Uniform prior beta examples

n	k	$I_{0.7,k+1,n-k+1}$	r_{cl}	r_{Δ}	r_0	r_1	r_{DM}	r_2
50	44	$1.9625 \cdot 10^{-3}$	-0.977	-0.364	-0.0232	$-3.1 \cdot 10^{-3}$	$8.5 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$
100	85	$3.3070 \cdot 10^{-4}$	-0.960	-0.282	-0.0164	$-3.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$7.8 \cdot 10^{-5}$
200	165	$3.1374 \cdot 10^{-5}$	-0.948	-0.221	-0.0106	$-2.1 \cdot 10^{-3}$	$9.4 \cdot 10^{-4}$	$2.1 \cdot 10^{-5}$
500	394	$5.2625 \cdot 10^{-6}$	-0.860	-0.143	-0.0061	$-1.0 \cdot 10^{-3}$	$2.1 \cdot 10^{-4}$	$4.0 \cdot 10^{-6}$
1000	767	$1.2096 \cdot 10^{-6}$	-0.777	-0.103	-0.0040	$-5.6 \cdot 10^{-4}$	$7.0 \cdot 10^{-5}$	$1.1 \cdot 10^{-6}$
2000	1501	$2.8479 \cdot 10^{-7}$	-0.684	-0.075	-0.0026	$-2.9 \cdot 10^{-4}$	$2.4 \cdot 10^{-5}$	$3.1 \cdot 10^{-7}$
5000	3671	$4.9194 \cdot 10^{-8}$	-0.553	-0.049	-0.0015	$-1.2 \cdot 10^{-4}$	$5.9 \cdot 10^{-6}$	$6 \cdot 10^{-8}$

TABLE 3
Jeffreys prior large deviation beta examples

n	k	$J_{0.7,k,n}$	r_{cl}	r_{Δ}	r_0	r_1	r_{DM}	r_2
50	37	0.27308	-0.037	-0.0269	-0.0134	$-2.5 \cdot 10^{-3}$	$-3.4 \cdot 10^{-4}$	$-5.5 \cdot 10^{-5}$
100	74	0.19233	-0.049	-0.0220	-0.0085	$-1.5 \cdot 10^{-3}$	$-1.4 \cdot 10^{-4}$	$-1.5 \cdot 10^{-5}$
200	148	0.10740	-0.074	-0.0188	-0.0052	$-9.3 \cdot 10^{-4}$	$-6.1 \cdot 10^{-5}$	$-3.8 \cdot 10^{-6}$
500	370	0.02432	-0.141	-0.0162	-0.0026	$-4.5 \cdot 10^{-4}$	$-2.1 \cdot 10^{-5}$	$-4.8 \cdot 10^{-7}$
1000	740	0.00261	-0.241	-0.0151	-0.0015	$-2.5 \cdot 10^{-4}$	$-9.9 \cdot 10^{-6}$	$-8 \cdot 10^{-8}$
2000	1480	0.00004	-0.405	-0.0144	-0.0008	$-1.4 \cdot 10^{-4}$	$-4.7 \cdot 10^{-6}$	$-1 \cdot 10^{-8}$
5000	3700	$2 \cdot 10^{-10}$	-0.713	-0.0140	-0.0003	$-5.7 \cdot 10^{-5}$	$-1.8 \cdot 10^{-6}$	$ \cdot \leq 10^{-9}$

do appear to approach 0. For the sharpest approximation $\pi_Y^{2,*}$ the relative error is quite small and becomes smaller very rapidly. We caution, however, that this is a one-dimensional case and the approximation of a marginal of a multidimensional posterior as in (1) and (3) could result in larger relative errors of order $O(1/n)$ [Tierney, Kass and Kadane (1989)]. Alternatively, if the likelihoods or priors are not smooth enough, some of the approximations involving derivatives may not be defined, and others may be defined but less accurate, as seen in Section 9 for a double-exponential prior.

For Table 1, we wrote a computer program which, given n , x , and $\alpha > 0$, starts with the smallest integer $j \geq nx$ and considers $k = j, j + 1, j + 2, \dots$ as long as $J_{x,k,n} \geq \alpha$, and finds the k in that range with the largest relative error r_{Δ} in absolute value. We chose $\alpha = 1/n^2$. Then $k/n \rightarrow x$ as n becomes large. Some terms in approximations vanish when $n = 2k$ (so, near $x = 0.5$) and one vanishes for the uniform prior when $k/n \doteq 0.916$. Thus we chose $x = 0.7$. As is perhaps not surprising, we found the largest relative errors when the probabilities $J_{x,k,n}$ were as small as allowed, with $J_{x,k,n} > \alpha > J_{x,k+1,n}$. We used the same k for each n in Table 2 as in Table 1. Since $\alpha \rightarrow 0$ as $n \rightarrow \infty$ at a slower than exponential rate, by Theorem 1, the relative error r_{Δ} in the approximation Φ_{Δ} should approach 0, which fits with Tables 1 and 2. The notation $|\cdot| \leq 10^{-9}$ indicates that the relative error (of the approximation $\pi_Y^{2,*}$) is less in absolute value than $1.5 \cdot 10^{-9}$; we do not give more exact values since the Holt algorithm by which we computed beta probabilities was constructed to give a relative error less than 10^{-9} . For the same reason, we round off possibly unreliable digits of r_2 when $|r_2| < 10^{-6}$.

The relative errors r_{cl} for the classical approximation Φ_{cl} based on the second derivative at the MLE decrease slowly in magnitude in Tables 1 and 2, even increase in size with n in Table 3, and they are quite large, except in the upper rows of Table 3. The Φ_{cl} approximation can work rather well in the middle of the posterior distribution. The smallest relative error r_2 in Table 3 is of order $n^{-5/2}$, in agreement with (7).

The approximation $\Phi(-\sqrt{2\Delta})$ of Theorem 1, which is the same in corresponding rows of Tables 1 and 2, is a better approximation to posterior probabilities for

the Jeffreys prior (Table 1) than it is for the uniform prior (Table 2), by a factor about 4 in the relative errors r_Δ . To understand this, observe as noted, for example, by Woodroffe (1992) that if we take as a coordinate parameter $z = -\sqrt{2\Delta}$ for $p \leq k/n$ and $\sqrt{2\Delta}$ for $p > k/n$ as in (5) then the likelihood function becomes exactly normal. The uniform prior dp on $0 < p < 1$ and the Jeffreys prior have densities with respect to dz , say, $\tau_u(z)$ and $\tau_J(z)$, respectively. One can evaluate the logarithmic derivatives of both at $z = 0$, where we find

$$\frac{\tau'_u(0)}{\tau_u(0)} = 4 \frac{\tau'_J(0)}{\tau_J(0)} = -\frac{2(2k - n)}{3\sqrt{nk(n - k)}}$$

which is not 0 unless $k = n/2$. Thus with respect to dz , the Jeffreys prior is more uniform than the prior dp around the MLE.

For higher order approximations the situation is more complicated. The approximations again tend to be more accurate for the Jeffreys prior, but not necessarily to the extent shown in Tables 1 and 2 for r_2 . For example, when $n = 500$, $x = 0.7$ and $k = 394$, for the Jeffreys prior $r_2 \doteq 3 \cdot 10^{-8}$ and for the uniform prior $r_2 \doteq 4 \cdot 10^{-6}$. The Jeffreys r_2 changes sign nearby from $k = 391$ to 392, so it is unusually small in absolute value. We found a similar sign change for $x = 0.8$, $n = 500$, $k = 437, 438$.

For accurate calculation of the approximations when $|x - (k/n)|$ is very small, analytic subtraction is needed, as for the Lugannani–Rice approximations [e.g., Daniels (1987), page 43; Reid (1996), page 143].

APPENDIX

The relationship between a coordinate and a profile likelihood root.

PROPOSITION 4. *Let $\psi(\phi) = a + v \cdot \phi$, $v \neq 0$, so ψ is a nonconstant affine function of ϕ . Under the assumptions made in Theorem 1, let ϕ_0 be the pseudo-true parameter of (A_5) . Then there is an open interval (c, d) containing $\psi_0 := \psi(\phi_0)$ such that almost surely for all n large enough, in (2) there is a 1–1, mutually C^1 relationship, depending on n and the observations, between z and ψ for $c < \psi < d$.*

PROOF. As after (8), we can make a linear coordinate change from ϕ to θ with $E(\theta_0) = I$. In the neighborhood V of θ_0 defined by (11), almost surely for n large enough, $\|E_n(\theta) - I\| < 2\delta_1/3$ by (14). By (16), almost surely for n large enough, there is a unique MLE $\hat{\theta}_n$ with $\|\hat{\theta}_n - \theta_0\| < \gamma_1$. Then $\theta \in V$ whenever $\|\theta - \hat{\theta}_n\| < \rho_1 := 3\gamma_1$ by (11). As in (17) and similarly to (12), by a translation and rotation of coordinates which does not change any of the preceding (but depends on n and $\hat{\theta}_n$), we can assume that $\hat{\theta}_n = 0$ and that $\psi = a + b\theta_1$ for some constants a, b with $b > 0$ where θ_1 is the first coordinate of θ . Thus we can assume $\psi = \theta_1$.

Let $\eta := \theta^{(1)} = (\theta_2, \dots, \theta_d) \in \mathbf{R}^{d-1}$. Consider the vector-valued function $\nabla_\eta H_n$

for H_n defined by (20) from $\{\|\theta\| < \rho_1\}$ into \mathbf{R}^{d-1} , where $\nabla_\eta = (\partial/\partial\theta_2, \dots, \partial/\partial\theta_d)$. Then $\nabla_\eta H_n$ is C^1 and $\nabla_\eta H_n(0) = 0$. By (14), the Hessian $\partial^2 H_n/\partial\theta_i\partial\theta_j$ is within $2\delta_1/3$ of the identity on the open set $\{\|\theta\| < \rho_1\}$, which almost surely for all n large enough includes all θ within $2\gamma_1$ of θ_0 .

If I_r is the $r \times r$ identity matrix, a matrix B is $d \times d$ and $B_{ij}^{(1)} = B_{ij}$ for $i, j = 2, \dots, d$, then it is easily seen that $\|B^{(1)} - I_{d-1}\| \leq \|B - I_d\|$. The derivative $G_n := \nabla_\eta \nabla_\eta H_n$ is the Hessian of H_n with respect to η , so by (14) $\|G_n - I_{d-1}\| < 2\delta_1/3$ on $\{\|\theta\| < \rho_1\}$. In particular G_n is invertible at $\theta = 0$. Thus by the implicit function theorem [e.g., Rudin (1976), Theorem 9.28], there are an open set $U_1 \subset \mathbf{R}^d$ and an open interval $W_1 \subset \mathbf{R}$, say $(-w, w)$, both containing 0, such that for all $\theta_1 \in W_1$, there is a unique $\eta(\theta_1)$ such that $\tilde{\theta} := \tilde{\theta}(\theta_1) := (\theta_1, \eta(\theta_1)) \in U_1$ and $\nabla_\eta H_n(\tilde{\theta}(\theta_1)) = 0$. Moreover, the function $\eta(\cdot)$ is C^1 on W_1 and $\eta(0) = 0$. We have by (22) that if $\|\theta\| = \rho_1$ then $H(\theta) \geq \rho_1^2/(2(1 + \delta_1))$, while if in addition $|\theta_1| < \rho_1/2$ then $H(\theta_1, 0) < \rho_1^2(1 + \delta_1)/8 < H(\theta)$. Thus if $|h| < \rho_1/2$ then $\inf\{H(\theta) : \theta_1 = h, \|\theta\| \leq \rho_1\}$ is attained at a point $\theta = \tilde{\theta}(h)$ at which $\|\theta\| < \rho_1$ and $\nabla_\eta H(\theta) = 0$. By strict convexity of $H(\cdot)$ on $\|\theta\| < \rho_1$, $\tilde{\theta}(h)$ is unique, so $\tilde{\theta}(\cdot)$ is a well-defined function from $(-\rho_1/2, \rho_1/2)$ into $\{\theta : \|\theta\| < \rho_1\}$, we can take $w \leq \rho_1/2$ and then the definitions of $\tilde{\theta}$ agree for $|\theta_1| < w$. By the implicit function theorem, $\tilde{\theta}(\cdot)$ is a C^1 function on some neighborhood of each point in $(-\rho_1/2, \rho_1/2)$. So by compactness of any closed subinterval, we can take $w = \rho_1/2$.

By (A5), there is a $\kappa > 0$ such that almost surely for n large enough,

$$(79) \quad \inf_{\theta \notin V} H_n(\theta) > \kappa.$$

Inequality (22) gives $H_n(\theta) \leq \|\theta\|^2(1 + \delta_1)/2 \leq \|\theta\|^2$ for $\|\theta\| < \rho_1$. Thus $H_n(\theta) \leq \kappa$ for $\|\theta\| \leq \min(\rho_1, \sqrt{\kappa})$, so $H_n(\tilde{\theta}(\theta_1)) \leq \kappa$ for $|\theta_1| < \gamma_2 := \min(\rho_1, \sqrt{\kappa})/2$. By strict convexity, for $|\theta_1| < \gamma_2$, $H_n(\theta_1, \eta)$ has a strict minimum with respect to η at $\eta = \eta(\theta_1)$ for $(\theta_1, \eta) \in V$ and by (79) for all η . In other words, the profile log likelihood $m(\theta_1) = \ell(\tilde{\theta}(\theta_1)) = \ell(0) - nH_n(\tilde{\theta}(\theta_1))$, so by (2),

$$(80) \quad z = (\text{sgn } \theta_1) \sqrt{2nH_n(\tilde{\theta}(\theta_1))}.$$

Let $g(t) := H_n(\tilde{\theta}(t))$ for $|t| < \gamma_2$. For $-\gamma_2 < u < u + v < u/2 < 0$,

$$\begin{aligned} g(u + v) &= H_n(u + v, \eta(u + v)) \leq H_n\left(u + v, \frac{u + v}{u}\eta(u)\right) \\ &= H_n(u, \eta(u)) + v \frac{d}{dt} H_n\left(u + t, \frac{u + t}{u}\eta(u)\right) \Big|_{t=\tau} \end{aligned}$$

for some τ , $0 < \tau < v$, by the mean value theorem. By (21) and since $\|(1, \eta(u)/u)\| \geq 1$, this derivative is less than or equal to $-|u|/[2(1 + \delta_1)]$. Letting $v \downarrow 0$ we get $g'(u) < 0$. Likewise we get $g'(u) > 0$ for $0 < u < \gamma_2$. Thus

$$\frac{dz}{d\theta_1} = \frac{\sqrt{n}|dH_n(\tilde{\theta}(\theta_1))/d\theta_1|}{\sqrt{2H_n(\tilde{\theta}(\theta_1))}} > 0$$

for $|\theta_1| < \gamma_2$ and $\theta_1 \neq 0$ since $H_n(\tilde{\theta}(\theta_1)) > 0$ for $\theta_1 \neq 0$. Also, $dz/d\theta_1$ is continuous in θ_1 for $0 < |\theta_1| < \gamma_2$.

Let $a_{ij} := \frac{1}{2} \partial^2 H_n(0) / \partial \theta_i \partial \theta_j$ for $i, j = 1, \dots, d$, and $b_j := d\tilde{\theta}_j/d\theta_1|_{\theta_1=0}$ for $j = 1, \dots, d$. Then by Taylor's theorem with remainder,

$$\sqrt{H_n(\tilde{\theta})/|\theta_1|} = \sqrt{\sum_{i,j=1}^d a_{ij} \tilde{\theta}_i \tilde{\theta}_j + o(\theta_1^2)} / |\theta_1| \rightarrow \sqrt{\sum_{i,j=1}^d a_{ij} b_i b_j}$$

as $\theta_1 \rightarrow 0$. The limit is strictly positive because the matrix $\{a_{ij}\}$ is positive definite (being close to $\frac{1}{2}I_d$) and the vector b is nonzero since $b_1 = 1$. By the chain rule, we have for $\tilde{\theta} = \tilde{\theta}(\theta_1)$ that

$$\frac{dH_n(\tilde{\theta})}{d\theta_1} = \frac{\partial H_n}{\partial \theta_1}(\tilde{\theta}) + \sum_{j=2}^d \frac{\partial H_n}{\partial \theta_j}(\tilde{\theta}) \frac{d\tilde{\theta}_j}{d\theta_1} = \frac{\partial H_n}{\partial \theta_1}(\tilde{\theta})$$

(the latter equation will not be used here). We have as $\|\theta\| \rightarrow 0$ for $j = 1, \dots, d$, $\partial H_n(\theta) / \partial \theta_j = 2 \sum_{i=1}^d a_{ij} \theta_i + o(\|\theta\|)$, and so

$$\frac{dH_n(\tilde{\theta}(\theta_1))}{d\theta_1} = 2 \sum_{j=1}^d \sum_{i=1}^d a_{ij} b_i b_j \theta_1 + o(|\theta_1|)$$

as $\theta_1 \rightarrow 0$. It follows that as $\theta_1 \rightarrow 0$, $(\text{sgn } \theta_1) d\sqrt{H_n(\tilde{\theta}(\theta_1))}/d\theta_1 \rightarrow \sqrt{\sum_{i,j=1}^d a_{ij} b_i b_j}$. So by (80), since z is a continuous function of θ_1 , and has a derivative approaching a positive limit as $\theta_1 \rightarrow 0$, z is a C^1 function of θ_1 with $dz/d\theta_1 > 0$ for $|\theta_1| < \gamma_2$.

Almost surely for n large enough we will have $\|\hat{\theta}_n - \theta_0\| < \gamma_2/2$, or in our eventual θ coordinates $\|\theta_0\| < \gamma_2/2$, so $|\theta_{01}| < \gamma_2/2$. Then $|\theta_1 - \theta_{01}| < \gamma_2/2$ implies $|\theta_1| < \gamma_2$. A C^1 function with a strictly positive derivative has a C^1 inverse. Thus the open interval $(\theta_{01} - \gamma_2/2, \theta_{01} + \gamma_2/2)$ has the properties stated in the proposition for ψ replaced by θ_1 , and the proposition is proved. \square

PROOF OF COROLLARY 1. By the last statement in Theorem 1, $\pi_{x,n}(A_n) / \Phi(-z_n) \rightarrow 1$. Almost surely for n large enough, we have Case I of the proof (28), so the MLE $\tilde{\theta}_n$ in ∂A_n exists and is unique almost surely, as shown just before (80). Write A_n as $\{y_1 \geq \zeta_n\}$ for coordinates y , depending on A_n , as defined in Section 5, equation (31). We have $d(\partial A_n, \phi_0) \rightarrow 0$, $\|\hat{\phi}_n - \phi_0\| \rightarrow 0$ a.s., and the coordinate transformations $\phi \mapsto \theta \mapsto y$ in Sections 4 and 5 increase distances at most by a constant factor not depending on A_n , so $\|\tilde{\theta}_n - \hat{\theta}_n\| \rightarrow 0$ and $\zeta_n \rightarrow 0$. Thus by uniform continuity of $\pi_1(\cdot)$ in a neighborhood of θ_0 and $\pi(\cdot)$ in the corresponding open set in y coordinates we have $\pi_1(\tilde{\theta})/\pi_1(\hat{\theta}) \rightarrow 1$ and as in (35), $\pi(\zeta_n, 0)/\pi(0) \rightarrow 1$ as $n \rightarrow \infty$. Similarly, the Hessians $\ell_{\eta\eta}$ are uniformly continuous in a neighborhood of θ_0 by (A₆) and (A₇), and their determinants are bounded away from 0 almost surely as $n \rightarrow \infty$ by (15) for θ coordinates or (34) for

y coordinates. Thus $\det[-\ell_{\eta\eta}(\tilde{\theta})]/\det[-\ell_{\eta\eta}(\hat{\theta})] \rightarrow 1$ and likewise in y coordinates. Let $\tilde{y} := \tilde{y}(y_1)$ be the MLE of y for given y_1 , or $\tilde{\theta}$ in y coordinates, which is well-defined for $|y_1|$ small enough. Then let $(y_1, \tilde{y}_2, \dots, \tilde{y}_d) := (y_1, \tilde{y}^{(1)}) := \tilde{y}(y_1)$. By (31) and (32),

$$(81) \quad y_1 \equiv \theta_1 \quad \text{and} \quad h_n(\tilde{y}) \equiv H_n(\tilde{\theta}).$$

Recall also that $\tilde{y}(\zeta_n) = (\zeta_n, 0)$ from just before (32) and so

$$\frac{dh_n(\tilde{y})}{dy_1} \Big|_{y_1=\zeta_n} = \frac{\partial h_n(y)}{\partial y_1} \Big|_{y=(\zeta_n, 0)} =: \frac{\partial h_n(\zeta_n, 0)}{\partial y_1}.$$

Thus

$$(82) \quad \frac{d\sqrt{h_n(\tilde{y})}}{dy_1} \Big|_{y_1=\zeta_n} = \frac{\partial h_n(\zeta_n, 0)/\partial y_1}{2\sqrt{h_n(\zeta_n, 0)}}.$$

By (36), $\zeta_n^2/[2(1+\delta)] \leq h_n(\zeta_n, 0) \leq (1+\delta)\zeta_n^2/2$, so $\zeta_n/\sqrt{1+\delta} \leq \sqrt{2h_n(\zeta_n, 0)} \leq \zeta_n\sqrt{1+\delta}$. Similarly from (34), $y_1/(1+\delta) \leq \partial h_n/\partial y_1 \leq y_1(1+\delta)$ for $y_1 > 0$ small enough. Thus by (82),

$$2^{-1/2}(1+\delta)^{-3/2} \leq d\sqrt{h_n(\tilde{y})}/dy_1|_{y_1=\zeta_n} \leq 2^{-1/2}(1+\delta)^{3/2}$$

for n large enough. So $d\sqrt{h_n(\tilde{y})}/dy_1|_{y_1=\zeta_n} \rightarrow 2^{-1/2}$ as $n \rightarrow \infty$ since we can let $\delta \downarrow 0$. By (80) and (81), and the end of the proof of Proposition 4, as $n \rightarrow \infty$

$$(83) \quad dz/dy_1|_{y_1=\zeta_n} \sim \sqrt{n} \quad \text{and} \quad dz/dy_1|_{y_1=0} \sim \sqrt{n}.$$

It follows from this and Proposition 4 that

$$\frac{dz/dy_1|_{y_1=\zeta_n}}{dz/dy_1|_{y_1=0}} \rightarrow 1, \quad \text{so} \quad \frac{dy_1/dz|_{z=z_n}}{dy_1/dz|_{z=0}} \rightarrow 1,$$

and Corollary 1 is proved. \square

Acknowledgments. We thank the Editors and two referees for a number of helpful comments, Richard Melrose for advice on implicit function domains and Michael Woodroffe for pointing out the paper of Berk (1966) to us.

REFERENCES

- ALON, N., BEN-DAVID, S., CESA-BIANCHI, N. and HAUSSLER, D. (1997). Scale-sensitive dimensions, uniform convergence, and learnability. *J. ACM* **44** 615–631.
- BARNDORFF-NIELSEN, O. E. and WOOD, A. T. A. (1998). On large deviations and choice of ancillary for p^* and r^* . *Bernoulli* **4** 35–63.
- BERGER, J. O. and MORTERA, J. (1999). Default Bayes factors for nonnested hypothesis testing. *J. Amer. Statist. Assoc.* **94** 542–554.
- BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58. [Correction (1996) **37** 745–746.]

- 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.
- BLEISTEIN, N. (1966). Uniform asymptotic expansions of integrals with stationary point near algebraic singularity. *Comm. Pure Appl. Math.* **19** 353–370.
- CHEN, C.-F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *J. Roy. Statist. Soc. Ser. B* **47** 540–546.
- CHOY, S. T. B. and SMITH, A. F. M. (1997). On robust analysis of a normal location parameter. *J. Roy. Statist. Soc. Ser. B* **59** 463–474.
- DANIELS, H. E. (1987). Tail probability approximations. *Internat. Statist. Rev.* **55** 37–48.
- DICICCIO, T. J. and MARTIN, M. A. (1991). Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference. *Biometrika* **78** 891–902.
- DICICCIO, T. J. and STERN, S. E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika* **80** 731–740.
- DUDLEY, R. M. (1993). *Real Analysis and Probability*, 2nd ed., corrected. Chapman and Hall, New York.
- DUDLEY, R. M. (1998). Consistency of M -estimators and one-sided bracketing. In *High Dimensional Probability* (E. Eberlein, M. Hahn and M. Talagrand, eds.) 33–58. Birkhäuser, Basel.
- DUDLEY, R. M., GINÉ, E. and ZINN, J. (1991). Uniform and universal Glivenko–Cantelli classes. *J. Theoret. Probab.* **4** 485–510.
- DUDLEY, R. M. and HAUGHTON, D. (1997). Information criteria for multiple data sets and restricted parameters. *Statist. Sinica* **7** 265–284.
- DUDLEY, R. M. and HAUGHTON, D. (2001). One-sided hypotheses in a multinomial model. In *Goodness-of-Fit Tests and Model Validity* (C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah, eds.) 387–399. Birkhäuser, Boston.
- ERKANLI, A. (1994). Laplace approximations for posterior expectations when the mode occurs at the boundary of the parameter space. *J. Amer. Statist. Assoc.* **89** 250–258.
- FRASER, D. A. S., REID, N. and WU, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* **86** 249–264.
- FULKS, W. and SATHER, J. O. (1961). Asymptotics. II. Laplace’s method for multiple integrals. *Pacific J. Math.* **11** 185–192.
- HAUGHTON, D. (1984). On the choice of a model to fit data from an exponential family. Ph.D. dissertation, MIT.
- HAUGHTON, D. M. A. (1988). On the choice of a model to fit data from an exponential family. *Ann. Statist.* **16** 342–355.
- HIPP, C. and MICHEL, R. (1976). On the Bernstein–v. Mises approximation of posterior distributions. *Ann. Statist.* **4** 972–980.
- HOLT, R. J. (1986). Computation of gamma and beta tail probabilities. Technical report, Dept. Mathematics, MIT.
- HSU, L. C. (1948). A theorem on the asymptotic behavior of a multiple integral. *Duke Math. J.* **15** 623–632.
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.
- JENSEN, J. L. (1995). *Saddlepoint Approximations*. Oxford Univ. Press.
- JOHNSON, R. A. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–864.
- LAWLEY, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43** 295–303.

- LE CAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.* **1** 227–329.
- LEONARD, T. (1982). Comment. *J. Amer. Statist. Assoc.* **77** 657–658.
- LUGANNANI, R. and RICE, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Adv. in Appl. Probab.* **12** 475–490.
- PAULER, D. K., WAKEFIELD, J. C. and KASS, R. E. (1999). Bayes factors and approximations for variance component models. *J. Amer. Statist. Assoc.* **94** 1242–1253.
- PERICCHI, L. R. and SMITH, A. F. M. (1992). Exact and approximate posterior moments for a normal location parameter. *J. Roy. Statist. Soc. Ser. B* **54** 793–804.
- POSKITT, D. S. (1987). Precision, complexity and Bayesian model determination. *J. Roy. Statist. Soc. Ser. B* **49** 199–208.
- REID, N. (1996). Likelihood and higher-order approximations to tail areas: A review and annotated bibliography. *Canad. J. Statist.* **24** 141–166.
- RUDIN, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, New York.
- SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.
- SHUN, Z. and MCCULLAGH, P. (1995). Laplace approximation of high-dimensional integrals. *J. Roy. Statist. Soc. Ser. B* **57** 749–760.
- SKINNER, L. A. (1980). Note on the asymptotic behavior of multidimensional Laplace integrals. *SIAM J. Math. Anal.* **11** 911–917.
- TALAGRAND, M. (1987). The Glivenko–Cantelli problem. *Ann. Probab.* **15** 837–870.
- TEMME, N. M. (1982). The uniform asymptotic expansion of a class of integrals related to cumulative distribution functions. *SIAM J. Math. Anal.* **13** 239–253.
- TEMME, N. M. (1987). Incomplete Laplace integrals: Uniform asymptotic expansion with application to the incomplete beta function. *SIAM J. Math. Anal.* **18** 1638–1663.
- TIERNEY, L. and KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* **81** 82–86.
- TIERNEY, L., KASS, R. E. and KADANE, J. B. (1989). Approximate marginal densities of nonlinear functions. *Biometrika* **76** 425–433. [Correction (1991) **78** 233–234.]
- VAN DER VAART, A. W. and WELLNER, J. A. (2000). Preservation theorems for Glivenko–Cantelli and uniform Glivenko–Cantelli classes. In *High Dimensional Probability II* (E. Giné, D. M. Mason and J. A. Wellner, eds.) 115–133. Birkhäuser, Boston.
- WALKER, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31** 80–88.
- WALL, H. S. (1948). *Analytic Theory of Continued Fractions*. Van Nostrand, New York.
- WONG, R. (1973). On uniform asymptotic expansion of definite integrals. *J. Approx. Theory* **7** 76–86.
- WONG, R. (1989). *Asymptotic Approximations of Integrals*. Academic Press, New York.
- WOODROOFE, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2** 91–111.

DEPARTMENT OF MATHEMATICS
 MASSACHUSETTS INSTITUTE OF TECHNOLOGY
 ROOM 2-245
 CAMBRIDGE, MASSACHUSETTS 02139
 E-MAIL: rmd@math.mit.edu

DEPARTMENT OF MATHEMATICAL SCIENCES
 BENTLEY COLLEGE
 175 FOREST ST.
 WALTHAM, MASSACHUSETTS 02154
 E-MAIL: dhaughton@bentley.edu