The Canadian Journal of Statistics
Vol. 31, No. ?, 2003, Pages ???-???
La revue canadienne de statistique

Asymptotic normality of the posterior given a statistic

Ao YUAN and Bertrand CLARKE

Key words and phrases: partial information; local limit theorem; posterior normality.
MSC 2000: Primary 62F15; secondary 62F12.

Abstract: The authors consider data, X^n, from a member of a parametric family indexed by a d dimensional real parameter θ. The parameter is equipped with a prior density $w(\theta)$. For a consistent and asymptotically normal estimator $T_n = T(X^n)$, satisfying a uniform local central limit theorem, the authors show that the posterior density for θ given T_n is asymptotically normal and they identify the asymptotic variance. The proof is given for continuous random variables, but generalizes to lattice valued random variables. The result requires a uniform Edgeworth expansion to control the behavior of T_n. The authors show this holds in typical cases. They provide examples and show that their result helps identify reference priors.

Title in French: we can supply this

Résumé : The authors consider IID random variables $X_1, \ldots, X_n$ from a member of a parametric family indexed by a d dimensional real parameter θ. The parameter is equipped with a prior density $w(\theta)$. For a consistent and asymptotically normal estimator $T_n = T(X^n)$, the authors show that the posterior density for θ given T_n is asymptotically normal and they identify the asymptotic variance. The proof is given for continuous random variables, but generalizes to lattice valued random variables. The result requires a uniform Edgeworth expansion to control the behavior of T_n. The authors show this holds in typical cases. They provide examples and show that their result helps identify reference priors.

1. INTRODUCTION

Suppose a Bayesian wants to estimate a parameter but has only partial information in the form of a summary statistic. This may arise from modeling considerations or because the full data set is unwieldy. Such settings are not unusual, and our main result here establishes an asymptotic normality result for them. We show that the posterior distribution given a summary statistic such as a mean is asymptotically normal.

One setting where one wants to make inferences from a summary statistic arises from modeling in item response theory, the statistical theory of standardized tests. One wants to estimate a construct such as 'mathematics achievement' from a test score which is usually a weighted mean, rather than from the full data set. If the test score is a sufficient statistic then one would use it. However, in some commonly occurring models such as the 3-parameter logistic model there is no sufficient statistic. To see this, observe that the 3-parameter logistic model for the i^{th} dichotomous test item is

$$P_\theta(X_i = 1) = c_i + (1 - c_i) \cdot \frac{1}{e^{-a_i(\theta - b_i)} + 1}$$

in which the subject dependent unidimensional parameter θ encapsulates the notion of achievement and the item dependent parameters are a_i for the discrimination of the i^{th} item, b_i for its difficulty, and c_i for the fraction of times an examinee guesses. When $c_i = 0$ this family is of exponential form, however when $c_i \neq 0$ there is no sufficient statistic. Nevertheless, practitioners regard the test score as the physically relevant quantity from which to make inferences.

Another setting in which partial information arises naturally is when the data set is unwieldy. Consider longitudinal data. Suppose there are several measurements on each experimental unit and we take the maximum of them; this is done in pulmonary function tests and blood pressure for instance. It may be reasonable to treat the maximum for each subject as arising from independent extreme value distributions even though the underlying distribution is not known. If it is reasonable to take the average of these maxima as an estimator for the population maximum then one may want to condition on the mean of the maxima. Reducing to the mean rather than using the full data set may give results that are more robust, an issue with highly variable data. Indeed, Reiss (1989, Chap. 4) and Hall (1992, Chap. 2) provide the expansions to generalize our main results here to include percentiles, such as the median, which have more robustness properties. That is, one can extend our results below to permit Bayesian inference on a parameter given the median.

Whatever its origin, consider a statistic $T_n = T(n, X^n) = (T_1(n, X^n), \ldots, T_k(n, X^n))$, in which data X^n is assumed follow a distribution $P_{\theta,n}$. The distributions and their densities are indexed by a parameter $\theta \in \mathbb{R}^d$ in which $\theta = (\theta_1, \ldots, \theta_d)$. We assume that $f(x|\theta)$ is continuous in θ for each x. When θ is regarded as a random variable Θ it has a prior density $w(\theta)$ with respect to Lebesgue measure. We assume the statistic T_n has a density $f(T_n|\theta)$ with respect to some dominating measure – either Lesbesgue or counting measure. We assume that T_n is consistent for a function $\eta(\theta)$ and is asymptotically normal.

We are interested in the behavior of the posterior density for Θ given T_n as n increases. Thus we examine $w(\theta|T_n) = w(\theta)f(T_n|\theta)/m(T_n)$, in which $m(T_n)$ is the integral of the numerator, i.e., $m(T_n)$ is the mixture of the densities for T_n with respect to w. Here, we have one main result: For continuous data from a smooth parametric family equipped with a continuous prior, the posterior density for the parameter θ, $w(\theta|T_n)$ is asymptotically normal and we identify the form of the variance giving the limiting behavior.

In the special case that T_n is sufficient, there is nothing to prove because $w(\theta|T_n) = w(\theta|X^n)$, the posterior density for θ given all the data. This last quantity is what is usually studied in asymptotic normality results for the posterior, see for instance Bickel & Yahav (1969) and Walker (1969). A partial list of other contributors to this area includes Le Cam (1958), Hartigan (1983), Fraser & McDunnough (1984) and more recently Sweeting (1992), Ghosal, Ghosh & Samanta (1995), Clarke (1999) amongst many others.

When T_n is not sufficient there are relatively few results on the asymptotic normality of the posterior. Le Cam (1953) proves posterior normality given the maximum likelihood estimator which is asymptotically sufficient, and Doksum & Lo (1990) study equivariant estimators in location families. In the special case that T_n is a mean of lattice valued random variables Clarke & Ghosh (1995) establish asymptotic normality of the posterior for use in educational testing problems from Junker (1993).

Our approach can also be regarded as a variant on that taken by Severini (1993) who chose interval estimates by ensuring that their (frequentist) coverage probability would match their posterior probability to order $O(1/n^{3/2})$ and $O_P(1/n^{3/2})$ respectively. The coverage probability is introduced as a way to choose which 100(1-α)% credible set to announce. Alternative criteria such as using a highest posterior density region or requiring the tails outside the interval to have the same probability are also possible. Severini (1991) permits the other direction: It uses the likelihood ratio statistic to generate credibility sets.

One of the implications of the main result parallels Strasser (1981). Strasser showed that any set of conditions that gives consistency of the MLE also gives concentration of the posterior. Here, we convert the usual asymptotic normality of a statistic in a frequentist sense into asymptotic normality of the posterior given the statistic. In effect, we give hypotheses under which any set of conditions giving frequentist normality also gives posterior normality.

A second implication is for reference priors. Bernardo (1979) defined reference priors to be those that achieved the maximal Shannon mutual information between a parameter Θ and a sample X^n. He, and others, verified that reference priors have many desirable properties. Our main result here, Theorem 1, partially formalizes a heuristic argument that permits identification of novel reference priors. We give this argument in Section 4.

Unlike many posterior normality results, our technique is not based on Laplace's method. A proof using Laplace's method may be possible, but would be longer, and probably no more general than the result here. Nevertheless, one could extract a distribution for a statistic T_n, define and use the Fisher information for T_n and an MLE based on T_n, and then attempt to use Laplace's method on $w(\theta|T_n)$ using the asymptotic normality of T_n. It is unclear whether the hypotheses would be as easy to verify as the ones we give in Proposition 1.

The key hypothesis of the main result is that the density of the statistic on which the parameter is conditioned satisfies a uniform local limit theorem. That is, the density $f_{T_n}(t|\theta)$ can be approximated in the limit, in supremum norm, by a sum of normal densities weighted by polynomials with factors decreasing in powers of $1/\sqrt{n}$. After using this as a hypothesis in our main result stated in Section 2, we give hypotheses to ensure that the uniform local limit theorem is satisfied for some choices of T_n. In Section 3, we use the full strength of the local limit theorem and state a posterior normality result given T_n in an L^1 sense. In Section 4, we give computations to suggest a suitable size of n for our approximation to hold as well as giving several theoretical examples and using our result to identify reference priors in some new cases. All significant proofs are relegated to Section 5 at the end.

2. ASYMPTOTIC NORMALITY GIVEN A STATISTIC

We assume T_n converges almost surely (a.s.), or in P_θ probability, to a function of θ that we write as $\eta(\theta)$, and that it is asymptotically normal with a rate $\sqrt{n}$. That is, we assume T_n satisfies

$$\sqrt{n}\{T_n - \eta(\theta)\} \xrightarrow{L} N_k\{\mathbf{0}, \Omega(\theta)\} \tag{1}$$

in which $\Omega(\theta)$ is the $k \times k$ asymptotic variance matrix.

It is seen that the dimension of T_n and $\eta(\theta)$ is k, and the dimension of θ is d. When $d > k$, there are too few statistics for a method of moments approach to give unique solutions for parameter values. Consequently, in such a case the limiting normal would be degenerate if it existed. Thus, a limiting normal only makes sense if $d \leq k$. In this case, the asymptotic variance of the limiting normal would be $\Omega(\theta)$, the $k \times k$ covariance matrix of T_n postmultiplied by $(D\eta)(\theta)$, the Jacobian of η, and premultiplied by its transpose. Here, we do not use this generality. We only establish the result for $k = d$. Indeed, the regularity conditions (2) and (3) below are violated when $d < k$. (One can get around this by imposing assumptions 2 and 3 on $(D\eta)(\theta)^t\Omega(\theta)(D\eta)(\theta)$ as in Clarke & Ghosh 1995.)

We assume here that η is globally invertible. For Theorem 1 below, in which the mode of convergence is P_{θ_o}-probability, we believe it is sufficient that η be locally invertible at a fixed value θ_o taken to be true. However, Theorem 2 gives the result in L^1 convergence, a stronger mode. When η is differentiable at a parameter value θ as in (3) below, one can use the delta-method, see Serfling (1980, p. 122), to verify

$$\sqrt{n}\{\eta^{-1}(T_n) - \theta\} \xrightarrow{L} N_d\left[\mathbf{0}, \{(D\eta)^{-1}\}^t(\theta)\Omega(\theta)(D\eta)^{-1})(\theta)\right]$$

Let Δ denote the closure of the support of $w(\cdot)$. We require that Ω be continuous and have determinant bounded away from zero and infinity on Δ. That is,

$$0 < \inf_{\theta \in \Delta} |\Omega(\theta)| \leq \sup_{\theta \in \Delta} |\Omega(\theta)| < \infty, \tag{2}$$

and $(D\eta)(\theta)$, with entries $\partial\eta_i(\theta)/\partial\theta_j$ exists, is continuously differentiable and satisfies

$$0 < \inf_{\theta\in\Delta} |D\eta(\theta)| \le \sup_{\theta\in\Delta} |D\eta(\theta)| < \infty. \tag{3}$$

Here, $|A|$ means the absolute value of the determinant of the matrix A.

Our proof relies on Edgeworth expansions to control error terms. Indeed, we use an Edgeworth expansion for the density $f_{V_n}(v|\theta)$ of

$$V_n = \sqrt{n}\Omega(\theta)^{-1/2}\{T_n - \eta(\theta)\} \xrightarrow{L} N_d(\mathbf{0}, \mathbf{I_d}).$$

The Edgeworth condition we impose on V_n is the following.

CONDITION E: There is a bounded, non-negative, integrable, real-valued function $c(\cdot)$ so that as $n \to \infty$

$$\forall v \qquad \sup_{\theta\in\Delta} |f_{V_n}(v|\theta) - \phi_d(v)| \le o(1)c(v),$$

where the o(1) does not depend on v or θ, and $\phi_d(v)$ is the density function of an d-dimensional standard normal distribution, $N_d(0, I_d)$ and $d = \dim(v)$.

After stating the main result of this section, we will give sufficient conditions for Condition E to be satisfied when T_n is a function of a mean. When T_n is a percentile, work by Reiss (1989, Chap. 4) suggests that Condition E continues to hold.

To state our result formally, observe that if Ω is a continuous function of θ then the square root of $\Omega(\theta)$, which we denote by $\Omega(\theta)^{1/2}$, can be chosen to be a continuous function of θ: let $A(\theta)$ be the unique upper triangular matrix with non-negative diagonal elements such that $\Omega(\theta) = A(\theta)A^t(\theta)$. Then A is continuous in θ. Now write $\Sigma(\theta) = (D\eta)^{-1}(\theta)\Omega(\theta)(D\eta)^{-1}(\theta)^t$ and define the formal square root of Σ to be $\Sigma(\theta)^{1/2} = (D\eta)^{-1}(\theta)\Omega(\theta)^{1/2}$.

Our main result is the following.

THEOREM 1. *Suppose that T_n has an absolutely continuous distribution and satisfies (1) for θ almost everywhere with respect to $w(\cdot)$, and that Condition* **E** *is satisfied. Assume that η is invertible and has a derivative satisfying (3) and that Ω is continuous and satisfies (2). Let w be positive at $\theta_0 \in Int(\Delta)$, and be continuous and bounded on the parameter space. Then, the conditional law of Θ given T_n converges in distribution to a Gaussian distribution with mean $\eta^{-1}(T_n)$ and covariance matrix $(1/n)\Sigma\{\eta^{-1}(T_n)\}$. That is, for any bounded and continuous function h from $I\!R^d$ to $I\!R$ we have*

$$\lim_{n\to\infty} \int h\{\sqrt{n}A_n^{-1}(\theta - \theta_n)\}w(\theta|T_n)d\theta = \int h(z)\frac{1}{(2\pi)^{d/2}}e^{-||z||^2/2}dz, \quad P_{\theta_0} \tag{4}$$

where $\theta_n = \eta^{-1}(T_n)$ and $A_n = (D\eta)^{-1}(\theta_n)\Omega^{1/2}(\theta_n)$. If, in addition, $T_n \to \eta(\theta_o)$ almost surely then (4) holds almost surely as well.

COMMENT 1: We could obtain posterior normality in terms of the parameter $\eta = \eta(\theta)$ and then transform to get a posterior for θ by a smooth one-to-one function. What we do here is equivalent.

COMMENT 2: Note that convergence of the posterior probabilities holds in the same mode as the convergence of T_n to $\eta(\theta_0)$.

Proof: Deferred to Section 5.

Now we turn to Condition E. Theorem 19.2 from Bhattacharya & Ranga Rao (1986) gives moment conditions and conditions on the characteristic function of a random variable sufficient to ensure that Condition E holds pointwise in θ. For our purposes, we use Bhattacharya & Ranga Rao (1986, p. 194, eqn. 19.26) to assert

$$\lim_{n\to\infty} \sup_{x\in \mathbb{R}^m} (1+||x||^2)|q_n(x|\theta) - \phi_{0,\Sigma}(x)| = 0, \tag{5}$$

where $q_n(x)$ is the density of $\sum_{i=1}^n X_i/\sqrt{n}$ and $\phi_{0,\Sigma}$ is the density of a normal with mean zero and variance matrix $\Sigma(\theta)$. Thus, the Bhattacharya & Ranga Rao (1986) conditions imply Condition E when $T_n = \bar{X}$ and Δ is any singleton set. The function $c(\cdot)$ in Condition E corresponds to $1/(1+||x||^2)$ in (5) which is integrable only when $\dim(X) = 1$. From Bhattacharya & Ranga Rao (1986), it is clear that (5) holds pointwise in θ whenever T_n is an average of functions if IID random variables, i.e., $T_n = (1/n)\sum_{i=1}^n h(X_i)$.

Here, we want Condition E to hold uniformly for θ in Δ (with an integrable error term c) when Δ is a region in $\mathbb{R}^d$, and $T_n = g\{(1/n)\sum_{i=1}^n h(X_i)\}$ for some known functions $h(\cdot)$ and $g(\cdot)$ with a nonsingular matrix $Dg(\cdot)$ of partial derivatives, if not more generally. Set $\zeta(\theta) = E_\theta h(X_i)$, $\eta(\theta) = g\{\zeta(\theta)\}$ and $\Gamma(\theta) = \mathrm{Var}_\theta\{h(X_1)\}$. We also set

$$S_n = \sqrt{n}\Gamma(\theta)^{-1/2}\left\{\frac{1}{n}\sum_{i=1}^n h(X_i) - \zeta(\theta)\right\} \xrightarrow{L} N(\mathbf{0}, I_d), \tag{6}$$

and, more generally, by the delta method we have

$$V_n = \sqrt{n}\left\{(Dg(\zeta(\theta)))\Gamma(\theta)(Dg(\zeta(\theta)))^t\right\}^{-1/2}\{T_n - \eta(\theta)\} \xrightarrow{L} N(\mathbf{0}, I_d). \tag{7}$$

In (7), $Dg(\zeta(\theta))$ is assumed to be nonsingular. Let $\hat{h}(t|\theta)$ be the characteristic function of $Y_i = h(X_i)$, assumed to be continuous in both arguments. We have the following variant of Theorem 19.2, page 192, in Bhattacharya & Ranga Rao (1986). It is similar to Proposition 2.1 in Clarke & Ghosh (1995).

PROPOSITION 1: *Let Y_i be a sequence of IID random vectors with values in $\mathbb{R}^d$ having mean zero and positive definite covariance matrix $\Gamma = \Gamma(\theta)$. Let K be a compact set in Δ and assume that $\hat{h}(t|\theta)$ is in $L^p(\mathbb{R}^d)$ for some $p \geq 1$ when $\theta \in K$. Suppose also that $||Y_1||^s$ is integrable for $\theta \in K$ with $E_\theta||Y_1||^s$ continuous on K for some $s \geq \max(3, d+1)$.*
(i) If $g(\cdot)$ is the identity, then, for each θ, a bounded, continuous density f_{V_n} for the distribution of V_n exists and is continuous in θ for n large enough and we have

$$\sup_{\theta\in K}\sup_v (1+||v||^s)\left| f_{V_n}(v|\theta) - \phi_d(v) - \sum_{j=1}^{s-2} n^{-j/2}P_j(v)\right| = o(n^{-(s-2)}),$$

where the $P_j(v)$'s are "exponential polynomials" defined below.

(ii) More generally, if $g: \mathbb{R}^d \mapsto \mathbb{R}^k$, for $k \leq d$, is not the identity, suppose that $Dg(\cdot)$ has continuous and bounded components and has $rank(Dg(\cdot)) = k$ at all $\zeta(\theta)$ for $\theta \in K$. Then, for each θ, a bounded, continuous density f_{V_n} for the distribution of V_n exists and is continuous in θ for n large enough and we have

$$\sup_{\theta\in K}\sup_v (1+||v||^s)\left| f_{V_n}(v|\theta) - \phi_k(v) - \sum_{j=1}^{s-2} n^{-j/2}P_j(v)\right| = o(1).$$

COMMENT 3: The appropriate form for the $P_j(v)$'s is from Bhattacharya & Ranga Rao (1986), hereafter BR. For a single fixed θ, Proposition 1 follows from BR Section 19. To get Proposition

1 more generally, one must look at the form of the P_j's as a function of θ. From BR Section 7 we have that

$$P_j(v) = \sum_{i=1}^{j} \frac{1}{i!} \sum_{l_1+\cdots+l_i=j,\ |r_s|=l_s+2,(1\le s\le i)} \sum \frac{\chi_{r_1}\cdots\chi_{r_i}}{r_1!\cdots r_i!}(-1)^{r_1+\cdots r_i} D^{r_1+\cdots r_i}\phi_k(v),$$

where, for non-negative integer vectors $r_s = (r_{s1}, \ldots, r_{sk})$, $|r_s| = r_{s1} + \cdots + r_{sk}$,

$$D^u \phi_k(v) = \frac{\partial u_1 + \cdots u_k}{\partial v_1^{u_1} \cdots \partial v_k^{u_k}} \phi_k(v), \qquad \text{for a vector} \quad u = (u_1, \ldots, u_k).$$

For a vector r, χ_r indicates the r-th cumulants, see BR, and $\phi_k(x)$ is the density of the standard Normal$(0, I_k)$ random variable. The cumulants involved are continuous functions of θ by the continuity of $\hat{h}$ and of $\mathrm{E}_\theta||Y_1||^s$. Thus, $\sup_K \chi_r < \infty$. Moreover $\int_v |D^{r_1+\cdots+r_i}\phi_k(v)|dv < \infty$. Now, Part (ii) of Prop. 1 implies Condition **E** with

$$c(v) = 1/(1 + ||v||^s) + \sum_{j=1}^{s-2} P_j(v).$$

COMMENT 4: In Part (ii), the case of $k > d$ is of little practical interest, since it leads to a degenerate normal. The case $k \le d$ occurs regularly. It is seen that in Part (ii), the error term is much larger than in Part (i). Although one can identify an expansion achieving the same order in the error term, the approximation required to reduce the $o(1)$ is complicated. The terms in the corresponding exponential polynomials will include the derivatives of $g(\cdot)$ and expressions in powers of S_n will emerge.

Proof: Deferred to Section 5.

We comment that our technique of proof extends to the case that T_n is lattice valued rather than continuous, subject to relatively minor adjustments. This is so because it is the continuity of the parameter θ rather than that of the data which is necessary so the support of the limiting distribution will be d-dimensional real space. Indeed, it is clear from Theorem 19.2 in Bhattacharya & Ranga Rao (1986) that all of the results in this section can be generalized to the case of independent but not identically distributed random variables, discrete and continuous. The INID nonlattice, or continuous, case is similar in terms of the quantities introduced but the expansions are generalizations of what we have done here. We have not done this partially for ease of exposition and partially for length considerations. We comment that our theorem may prove more useful in settings in which the random variables are not identical because Markov Chain Monte Carlo (MCMC) computations are essentially only done in the IID setting and the definition of sufficiency is mostly tailored to the IID case. Clarke & Ghosh (1995) shows how the INID case can be handled in the lattice case.

It is unclear how to generalize the results to settings with dependent random variables because conditions for the validity of the expansions become difficult to interpret and verify. On the other hand, results due to Reiss (1989) and Hall (1992) give local limit theorems for sample percentiles that can be used to establish Condition E when T_n is a central order statistic. Thus, the posterior given a percentile will also be asymptotically normal but we have not identified the scale.

Recall that when T_n is sufficient, we have $w(\theta|T_n) = w(\theta|X_1, \ldots, X_n)$ which is well known to be asymptotically normal with limiting distribution of the form $N\{\theta_0, I^{-1}(\theta_0)/n\{$, where $I(\theta)$ is the Fisher information matrix and θ_0 is the true value of the parameter. That is, we get the usual $\sqrt{n}$ rate of concentration. Our results show that if the MLE $\hat{\theta}$ is a function g of the mean of $h(X_i)$'s then $w(\theta|\hat{\theta})$ is asymptotically normal with rate $\sqrt{n}$, located at $\xi\{g^{-1}(\hat{\theta})\}$ where $\xi(\theta) = E_\theta h(X_1)$, and asymptotic variance given by the Fisher information, since the MLE is efficient. That is, when T_n

is sufficient or is an MLE that satisfies Proposition 1 we get the best possible asymptotic variance. Conditioning on a statistic that is not sufficient or is not optimal while satisfying Condition E leads to posteriors which will typically have the same rate, $\sqrt{n}$, but an asymptotic variance larger than the inverse Fisher information. That is, using partial information may sacrifice the constant multiple of the rate of convergence but does not generally alter the rate itself. A ratio of the asymptotic variance of $w(\theta|T_n)$ to the inverse Fisher information measures the loss of efficiency from using T_n.

3. ASYMPTOTIC NORMALITY IN L^1

The point of this section is to assert that the main results of the last section extend to the L^1 mode of convergence. The L^1 mode is stronger than convergence in probability and incompatible with convergence almost sure.

Let $\phi_{\eta^{-1}(T_n),\Sigma\{\eta^{-1}(T_n)\}/n}(\theta)$ be the normal density function with mean $\eta^{-1}(T_n)$ and variance matrix $\Sigma\{\eta^{-1}(T_n)\}/n$. In view of Theorem 1, Proposition 1 suggests the following.

THEOREM 2. *Assume the hypotheses of Theorem 1, those of Proposition 1 and that w has bounded support on which it it is bounded away from zero. Then, we have*

$$\int\int \left|w(\theta|T_n = t) - \phi_{\eta^{-1}(T_n=t),\Sigma\{\eta^{-1}(T_n=t)\}/n}(\theta)\right| d\theta m_{T_n}(t)dt \to 0. \qquad (8)$$

Proof: Deferred to Section 5.

Note that local conditions on $w(\theta)$ are enough for Theorem 1. Here, for L^1, we seem to need stronger hypotheses because the influence of the tail properties of the prior, or of the boundary points of its support, would be otherwise difficult to control. Nevertheless, we conjecture that a suitable truncation argument would give a more general result.

4. EXAMPLES

In this section we develop three practical aspects of our main result. First, we computationally examine the behavior of our normal approximation as n increases, suggesting that in some cases it might be valid for sample sizes as small as 20. Then we give several closed form examples to see that they hypotheses are satisfied in some typical cases. Finally, we note that the asymptotic normality permits derivation of an estimator dependent reference prior – which was the original motivation for our result.

4.1 Computational Example

Here we present a computational example to suggest a reasonable size of n for our approximation to be valid. We argue that this has a better chance to generalize than a data driven example which would necessarily be closely tied to a specific setting. Moreover, as is seen in the next subsection using our theorem is straightforward: One is merely justifying the replacement of an unknown density with a normal density.

One of the strengths of our result is that it applies even when standard techniques in the Markov Chain Monte Carlo family cannot be used. In general MCMC-based techniques permit evaluation of a posterior from a joint density for a parameter and statistic that is not normalized rather than a joint density that cannot be written down in closed form. This means that our theorem can be used with statistics that do not admit tractable expressions for their conditional density given the parameter. For this reason any computational example has to be carefully chosen.

For ease of computation suppose we have $X_1, \ldots, X_n$ IID $N(\mu, \sigma^2)$. Let $T_n = \overline{X}/S_n$, where $\overline{X}$ is the sample mean and S_n^2 is the sample variance. Note $\sqrt{n}T_n$ has a noncentral $t_{n-1}(\theta)$ distribution where θ is the mean-to-standard-deviation ratio given by $\theta = \mu/\sigma$.

We first observe the key hypothesis of our theorem, the convergence in distribution of $\sqrt{n}(T_n - \theta)$ to a normal, is satisfied. So, let $Y_i = (X_i, X_i^2)$, then $\sqrt{n}(\overline{Y} - \zeta) \xrightarrow{L} N(\mathbf{0}, \Sigma)$, where $\zeta = (\mu, \sigma^2 + \mu^2)$ and

$$\Sigma = \begin{pmatrix} \sigma^2 & 2\mu\sigma^2 \\ 2\mu\sigma^2 & 2\sigma^4 + 4\mu^2\sigma^2 \end{pmatrix}.$$

Now, for $y = (y_1, y_2)^t$, let $g(y) = y_1/\sqrt{y_2 - y_1^2}$. Then $T_n = g(\overline{Y})$, and $\theta = g(\zeta)$. Since the partial derivative is $Dg = (\frac{\partial g}{\partial y_1}, \frac{\partial g}{\partial y_2}) = (\frac{y_2}{(y_2 - y_1^2)^{3/2}}, -\frac{y_1}{2(y_2 - y_1^2)^{3/2}})$, we have

$$V_n := \sqrt{n}\Omega(\theta)^{-1/2}(T_n - \theta) \xrightarrow{L} N(0, 1),$$

where $\Omega(\theta) = Dg(\zeta)\Sigma Dg(\zeta)^t = (2\sigma^2 + \mu^2)/(2\sigma^2) = 1 + \theta^2/2$. We assume the characteristic function of X_i is integrable (see the comment after Example 1). Since Dg is continuous and bounded componentwise on a compact neighborhood of the "true" parameter θ_0^* (or on a compact neighborhood of the its empirical version), Part (ii) of Proposition 1 gives that Condition **E** is satisfied for the density $f_{V_n}(v|\theta)$ of V_n, so the assumptions of Theorem 1 are satisfied. Now, we can conclude that $w(\theta|T_n)$ is asymptotically $N\left[T_n, (1/n)\{1 + (T_n)^2/2\}\right]$.

To exemplify this result, we use Markov Chain Monte Carlo to sample the posterior distribution $w(\theta|T_n)$ to observe its convergence to its theoretical limit. We take the prior $w(\theta)$ to be $N(.5, 1)$ and set $\theta_0 = 0.5$, with $(\mu_0, \sigma_0^2) = (0.5, 1)$ and generate samples of T_n for $n = 10, 20, 30$ and 50. We take the transition density to be the prior density, and $f_n(t|\theta)$ be the density of T_n. At each MCMC iteration k, we draw a candidate sample θ^* from the transition density $w(\theta^*|\theta^{(k-1)}) = w(\theta^*)$, and compute the Metropolis-Hastings ratio

$$r = \frac{w(\theta^*|T_n)w(\theta^{(k-1)}|\theta^*)}{w(\theta^{(k-1)}|T_n)w(\theta^*|\theta^{(k-1)})} = \frac{f_n(T_n|\theta^*)}{f_n(T_n|\theta^{(k-1)})}.$$

With probability $\min\{r, 1\}$, we accept θ^* as $\theta^{(k)}$, otherwise set $\theta^{(k)} = \theta^{(k-1)}$. The density $f_n(t|\theta)$ of T_n is

$$f_n(t|\theta) = \frac{\sqrt{n}(n-1)^{(n-1)/2}e^{-n\theta^2/2}}{\Gamma\left(\frac{n-1}{2}\right)(n-1+nt^2)^{n/2}} \sum_{i=0}^{\infty} \Gamma\left(\frac{n+i}{2}\right) \frac{(\sqrt{n}\theta)^i}{i!} \left(\frac{2nt^2}{n-1+nt^2}\right)^{i/2}.$$

However, when the noncentrality parameter θ is zero, it reduces to the t_{n-1} distribution, in which case we know how close to normal our quantities are. However, as the noncentrality parameter increases, we anticipate needing ever larger n's for the approximation of T_n by a normal to be valid.

It is difficult to compute $f_n(t|\theta)$ because of the infinite summation and because the values of $\exp(n\theta^2/2)$ and $\Gamma\{(n + i)/2\}$ get large fast. Taking finitely many terms in the first and using Stirling's approximation in the second does not appear to solve the problem. Indeed, it is unclear how to choose which terms to include and to evaluate how good the approximation is. Instead, we use the asymptotic distribution of T_n as an approximation, i.e. we use $T_n \sim N\{\theta, v^2(\theta)/n\}$ approximately. Thus the Metropolis-Hasting ratio is approximated by

$$r \approx \frac{v(\theta_{(k-1)})}{v(\theta^*)} e^{\frac{n}{2}\left\{\frac{(T_n - \theta_{(k-1)})^2}{v^2(\theta_{(k-1)})} - \frac{(T_n - \theta^*)^2}{v^2(\theta^*)}\right\}}$$

$$= \sqrt{\frac{2 + \theta_{(k-1)}^2}{2 + \theta^{*2}}} e^{n\left\{\frac{(T_n - \theta_{(k-1)})^2}{2 + \theta_{(k-1)}^2} - \frac{(T_n - \theta^*)^2}{2 + \theta^{*2}}\right\}}$$

Four plots from our simulations appear in Figure 1. To generate them, we drew 15,000 MCMC samples and used only the last 5,000 to estimate the posterior density. Over this computed density we have plotted the normal approximation given by our theorem. It is seen that when the true value of θ is 0.5 the values of T_n cluster around 0.5, because it is consistent and asymptotically normal. The approximating normal from our theorem is located at T_n and the computed posterior also appears to have its mode near 0.5. Moreover, the spread of the limiting normal appears to be close to the spread of the computed posterior. It is seen in this case that when $n = 50$ the approximation is quite good in the sense that the locations and scales of the two densities are nearly indistinguishable. For n near 30, the heights of the modes match but the locations and scales are distinguishable despite being close. Even for n as low as 20, the approximation is pretty good – although for n around 10 it is probably not adequate.

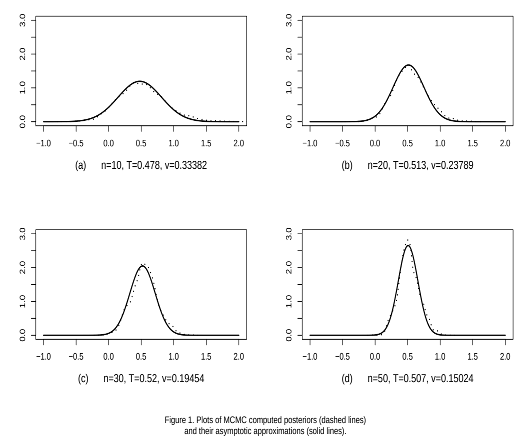

Figure 1: Caption for figure 1.

As a generality, we conjecture that our approximation will be reasonably good in well-behaved cases by the time n is around 30. By well-behaved we mean a statistic with a unimodal conditional density, a unimodal prior that is not far wrong, and an asymptotic variance that is not too large.

4.2 Closed Form Examples

Here we argue that the range of standard examples covered by our results is large. Indeed, given a prior and model our result automates the generation of asymptotically valid credible sets given a wide range of summary statistics for use in initial data analysis.

For instance, suppose T_n is the MLE $\hat{\theta}$ for θ and that $\eta(\theta) = \theta$ and $\Sigma(\theta) = I^{-1}(\theta)$, the inverse of the Fisher information matrix of X_1, when it exists. Then, when $\hat{\theta}$ is a function of the form defined in Proposition 1, we have that $w(\theta|\hat{\theta})$ is asymptotically $N\{\hat{\theta}, (nI)^{-1}(\hat{\theta})\}$. Also, when $T_n = \eta(\hat{\theta})$, the MLE for $\eta(\theta)$, we get that $w(\theta|\eta(\hat{\theta}))$ is asymptotically

$$N\left\{\hat{\theta}, n^{-1}(D\eta)^{-1}(\hat{\theta})\Omega(\hat{\theta})(D\eta)^{-1}(\hat{\theta})\right\},$$

where $\Omega = \Omega(\theta)$ is the asymptotic variance of T_n.

Note, this differs from the usual posterior normality result informally stated as $w(\theta|X^n)$ is asymptotically $N[\hat{\theta}, 1/\{nI(\theta_0)\}]$, where $\hat{\theta}$ is the MLE for a unidimensional θ based on X^n, see, for instance Walker (1969). The difference is that $\dim(T_n) = d$ whereas $\dim(X^n) = n$. Setting $T_n = X^n$ gives an increasing dimension, violating our use of local limit theorems. In essence we are asking 1) In what sense must a finite dimensional statistic be informative? And, 2) How informative must it be so as to permit meaningful posterior inference? Apart from regularity conditions such as the invertibility of η and the positive definiteness of Fisher information matrices, our answer comes down to asymptotic normality of the statistic and Condition **E** , for 1) and 2) respectively. That is, if a statistic is asymptotically normal and this leads to a suitable property of densities (Condition E) then posterior inference given the statistic is feasible. Note that in principle, asymptotic normality of T and Condition E can be verified in many settings.

Next we give four standard statistics which our result covers.

Example 1: Sample Mean. Let $T_n = \overline{X}$, set $\eta(\theta) = E_\theta X_1$ and $\sigma^2 = Var_\theta(X_1)$, which we assume to be a bounded function of θ. Now, $T_n \overset{a.s.}{\to} \eta(\theta)$, and $\sqrt{n}\{T_n - \eta(\theta)\} \overset{L}{\to} N(0, \sigma^2)$. We set $g(x) \equiv h(x) \equiv x$, so that $g(\cdot)$ is continuously differentiable. If we choose the parameter space to be a compact set K and require X_1 to have finite third absolute moment, then $||Y_1||^3 = |x|^3$ is integrable uniformly on K. Now, when the characteristic function $\hat{h}(t|\theta)$ of X_1 is continuous in (t, θ) and is p^{th} power integrable Proposition 1 gives that Condition E is satisfied. Lastly, we see that T_n satisfies (1) for all $\theta \in K$; $\eta(\theta) = \theta$ is invertible and differentiable and satisfies (3); and $\Omega(\theta) = \sigma^2(\theta)$ satisfies (2). Provided we choose a bounded and continuous prior $w(\cdot)$ for θ, all the conditions in Theorem 1 are satisfied. Consequently, we have that $w(\theta|T_n)$ is asymptotically $N\{\eta^{-1}(T_n), \sigma^2/n\}$, as was shown in Clarke & Ghosh (1995) for the regular lattice case. Note that Theorem 2 also applies for any prior strictly positive with compact support, in this and the following examples.

We comment that many commonly occurring parametric families have integrable characteristic functions. These include the normal and Gamma (for a wide variety of parameters) as well as the double exponential, the Cauchy, the triangular distribution, and the cosine distribution $(1 - \cos x)/x^2$. Consequently, in the following examples, we can make similar appropriate assumptions as in Example 1, so that all the conditions of Condition E and of Theorem 1 will be satisfied. We have omitted these details below to avoid redundancy.

Example 2: Percentiles. Let $T_n = Q_\alpha$, the $100\alpha^{th}$ sample percentile, where and $0 < \alpha < 1$. It is well known that $T_n \overset{a.s.}{\to} \eta(\theta)$, and that $\sqrt{n}\{T_n - \eta(\theta)\} \overset{L}{\to} N[0, \alpha(1-\alpha)/f^2\{\eta(\theta)|\theta\}]$, where $\eta(\theta) = q_\alpha(\theta)$ is the α^{th} percentile of the distribution, and $f(\cdot|\theta)$ is the density function of X_1. Since Condition E is satisfied in this case (Reiss, 1989) our results give that $w(\theta|T_n)$ is asymptotically $N\left(\eta^{-1}(T_n), \alpha(1-\alpha)/[nf^2\{T_n|\eta^{-1}(T_n)\}]\right)$, the same as was shown in Yuan & Clarke (1993) in L^1. See Reiss (1989, Theorem 4.7.1) and Hall (1992, Theorem 2.5) for the local limit theorems and asymptotic expansions that can be used to justify the extension of the main results here to the case of conditioning on sample percentiles or more general statistics.

Example 3: The Computed Case of 4.1 Revisited. In a two dimensional setting, let $\theta = (\mu, \sigma^2)$, where $\mu = E(X_1)$ and $\sigma^2 = Var(X_1)$. Let $T_n = (\overline{X}, \sqrt{n}\overline{X}/S_n)$ with the second component being the usual t_{n-1}-statistic for testing $\mu = 0$, where S_n^2 is the sample variance. Write $\eta(\theta) = (\mu, \mu/\sigma^2)$. Set $Y_i = (X_i, X_i^2)$ with mean $\zeta(\theta) = (\mu, \sigma^2 + \mu^2)$, and variance $\mathrm{V}ar(Y_1) = \Sigma(\theta)$. Now, $\overline{Y} \overset{a.s.}{\to} \zeta$ and $\sqrt{n}\{\overline{Y} - \zeta(\theta)\} \overset{L}{\to} N\{0, \Sigma(\theta)\}$.

For $g(y) = (y_1, y_1/\sqrt{y_2 - y_1^2})$, we have $\eta(\theta) = g(\zeta)$ and $T_n = g(\overline{Y})$. So, we get $T_n \overset{a.s.}{\to} \eta(\theta)$, and $\sqrt{n}\{T_n - \eta(\theta)\} \overset{L}{\to} N\{0, \Omega(\theta)\}$, where $\Omega(\theta) = Dg\{\zeta(\theta)\}\Sigma(\theta)Dg\{\zeta(\theta)\}^t$. Thus, $w(\theta|T_n)$ is asymptotically $N[\eta^{-1}(T_n), n^{-1}\Sigma\{\eta^{-1}(T_n)\}]$.

Example 4: Covariance. Let $X_1, \ldots, X_n$ be IID $f(\cdot|\theta)$ with $X_i = (X_{i,1}, X_{i,2})$. Consider the five dimensional parameter $\zeta(\theta) = (E_\theta(X_{1,1}), E_\theta(X_{1,2}), E_\theta(X_{1,1}X_{1,2}), E_\theta(X_{1,1}^2), E_\theta(X_{1,2}^2))$, and set

$\eta(\theta) = \mathrm{C}orr_\theta(X_{1,1}, X_{1,2})$. The law of large numbers gives

$$Y_n = (1/n)\sum_{i=1}^{n}(X_{i,1}, X_{i,2}, X_{i,1}X_{i,2}, X_{i,1}^2, X_{i,2}^2) \stackrel{a.s.}{\to} \zeta(\theta),$$

and the central limit theorem gives $\sqrt{n}\{Y_n - \zeta(\theta)\} \stackrel{L}{\to} N\{\mathbf{0}, \Sigma(\theta)\}$, where $\Sigma(\theta) = \mathrm{V}ar_\theta(Y_1)$. Now if we let

$$g(y_1, y_2, y_3, y_4, y_5) = (y_3 - y_1 y_2)/\sqrt{(y_4 - y_1^2)(y_5 - y_2^2)},$$

it is easy to check that $\eta(\theta) = g\{\zeta(\theta)\}$. Setting $T_n = g(Y_n)$ we have that $T_n \stackrel{a.s.}{\to} \eta(\theta)$, and $\sqrt{n}\{T_n - \eta(\theta)\} \stackrel{L}{\to} N\{0, Dg(\theta)\Sigma(\theta)Dg(\theta)^t\}$, see Serfling (1980, p. 122). So $w(\theta|T_n)$ is asymptotically $N\left[\eta^{-1}(T_n), n^{-1}\Sigma\left\{\eta^{-1}(T_n)\right\}\right]$.

4.3 An Estimator Dependent Reference Prior

Finally, our result permits a heuristic derivation of a new class of reference priors. We, have that, asymptotically, $w(\theta|T_n)$ is $N\left[\eta^{-1}(T_n), \Sigma\{\eta^{-1}(T_n)\}/n\right]$. If we use this and (7) then we get an asymptotic expression for the mutual information $I(\Theta, T_n)$. In fact,

$$I(\Theta, T_n) = \int w(\theta) R_{(w,T_n)}(\theta) d\theta,$$

where

$$R_{(w,T_n)}(\theta) = \int f_{T_n}(t|\theta)\log\frac{f_{T_n}(t|\theta)}{m_{T_n}(t)}dt$$

$$= \int f_{T_n}(t|\theta)\log f_{T_n}(t|\theta)dt - \int f_{T_n}(t|\theta)\log m_{T_n}(t)dt.$$

The first term is minus the entropy, $-H(T_n)$, of T_n. One can verify that typically

$$H(T_n) \sim \frac{d}{2}\log\frac{n}{2\pi} - \log|\Omega(\theta)|^{1/2}.$$

The second term is asymptotically $\log w(\theta)|D^{-1}\eta(\theta)|$, when θ is true. Thus the asymptotic approximation of $I(\Theta, T_n)$ is

$$\frac{d}{2}\log\frac{n}{2\pi e} + H(\Theta) + \int w(\theta)\log\frac{|D\eta(\theta)|}{|\Omega(\theta)|^{1/2}}d\theta,$$

where the entropy H of Θ is $H(\Theta) = -\int \log w(\theta)d\theta$. Now, a calculus of variations argument to maximize $I(\Theta, T_n)$ over the marginal density, $w(\cdot)$, for Θ gives

$$w^*(\theta) = |D\eta(\theta)|/|\Omega(\theta)|^{1/2}.$$

Clarke and Yuan (2003) develops this approach for use in prior selection and information theory.

5. PROOFS OF MAJOR RESULTS

5.1 Proof of Theorem 1.

There are three steps in the proof. The first identifies the two statements we must establish; the last two prove them.

Step 1: Denote by $f_{T_n}(\cdot|\theta)$ and $f_{V_n}(\cdot|\theta)$ the conditional densities of T_n and V_n given $\Theta = \theta$. To prove the result, it suffices to show that for any bounded, continuous function h from $\mathbb{R}^d$ to $\mathbb{R}$, we have

$$\lim_{n\to\infty} R_n(h) = w(\theta_0)|(D\eta)^{-1}(\theta_0)|E(h(Z)) \qquad P_{\theta_0} \quad a.s., \tag{9}$$

where Z has a multivariate central Gaussian distribution with covariance matrix I_d, and

$$R_n(h) = \int h\{\sqrt{n}A_n^{-1}(\theta - \theta_n)\} f_{T_n}(T_n|\theta)w(\theta)d\theta.$$

Expression (9) is enough to establish Theorem 1 because it implies $R_n(h)/R_n(1)$ converges P_{θ_0} a.s. to $E\{h(Z)\}$. This is exactly what we need to show, because the conditional density $w(\theta|T_n)$ of Θ, given T_n, is $w(\theta|T_n) = f_{T_n}(T_n|\theta)w(\theta)/R_n(1)$.

The proof of (9) can be done in two steps. First, we prove (9) for functions h with compact support. Second, we show that for any $\epsilon > 0$, there is $M > 0$ so that

$$\limsup_{n\to\infty} \int I_{\{\sqrt{n}||A_n^{-1}(\theta-\theta_n)||>M\}} f_{T_n}(T_n|\theta)w(\theta)d\theta \le \epsilon \qquad P_{\theta_0}, \quad a.s. \tag{10}$$

These two facts combined prove (9) for bounded, continuous functions h. Observe that establishing (10) is equivalent to showing that the law of $\sqrt{n}A_n^{-1}\{\Theta - \eta^{-1}(T_n)\}$, given T_n, is almost surely tight with respect to P_{θ_0}.

Step 2 : We show (9), for functions h with compact support. First note that for any β, (2) and (3) imply that both $|\Omega\{\eta^{-1}(T_n + \beta/\sqrt{n}))\}$ and $|(D\eta)\{\eta^{-1}(T_n + \beta/\sqrt{n})\}|$ converge P_{θ_0} almost surely to $|\Omega(\theta_0)|$ and $|D\eta(\theta_0)|$, and these two sequences are uniformly bounded away from zero and infinity.

Then, setting $v = v(t) = \sqrt{n}\Omega^{-1/2}(\theta)\{t - \eta(\theta)\}$, it is easy to verify that

$$f_{T_n}(t|\theta) = n^{d/2}|\Omega^{-1/2}(\theta)|f_{V_n}\{v(t)|\theta\}. \tag{11}$$

Next, since h is bounded and has compact support, Condition E and (11) imply that

$$\begin{aligned}
R_n(h) =& \int h\left\{\sqrt{n}A_n^{-1}(\theta - \theta_n)\right\} \frac{n^{d/2}}{|\Omega(\theta)|^{1/2}} f_{V_n}\left[\sqrt{n}\Omega^{-1/2}(\theta)\{T_n - \eta(\theta)\}\Big|\,\theta\right] w(\theta)d\theta \\
=& n^{d/2}\int h\left\{\sqrt{n}A_n^{-1}(\theta - \theta_n)\right\} \frac{e^{-\frac{n}{2}\{T_n-\eta(\theta)\}^t\Omega^{-1}(\theta)\{T_n-\eta(\theta)\}}}{(2\pi)^{d/2}|\Omega(\theta)|^{1/2}} w(\theta)d\theta + o(1) \\
=& \int \frac{h\left[\sqrt{n}A_n^{-1}\{\eta^{-1}(T_n + \beta/\sqrt{n}) - \theta_n\}\right]}{|D\eta^{-1}\{\eta^{-1}(T_n + \beta/\sqrt{n})\}|} \\
&\times \frac{e^{-\frac{1}{2}\beta^t\Omega^{-1}(T_n+\beta/\sqrt{n})\beta}}{(2\pi)^{d/2}|\Omega\{\eta^{-1}(T_n + \beta/\sqrt{n})\}|^{1/2}} w\{\eta^{-1}(T_n + \beta/\sqrt{n})\}d\beta + o(1) && (12) \\
\to& \frac{w(\theta_0)}{|D\eta(\theta_0)||\Omega(\theta_0)|^{1/2}} \int h\{\Omega^{-1/2}(\theta_0)\}\frac{e^{-\frac{1}{2}\beta^t\Omega^{-1}(\theta_0)\beta}}{(2\pi)^{d/2}}d\beta && (13) \\
=& \frac{w(\theta_0)}{|D\eta(\theta_0)|}E\{h(Z)\},
\end{aligned}$$

where we used the change of variables $\theta \longmapsto \beta = \sqrt{n}\{\eta(\theta) - T_n\}$ explicitly to get (12). This change of variables is also used implicitly in the $o(1)$ error term, which is, in fact, $o(1)$ by the integrability of the function c in Condition E. Expression (13) follows from the Dominated Convergence Theorem.

Step 3: Under (1), we have that there is a positive number C_1 so that

$$\{\sqrt{n}||A_n^{-1}(\theta - \theta_n)|| > M\} \subset \{||\beta|| > M/C_1\}$$

holds with P_{θ_o} probability as close to one as desired, for n large enough. Now, Condition E and the change of variables used in Step 2 gives that the main quantity in (10),

$$\int I_{\{\sqrt{n}||A_n^{-1}(\theta-\theta_n)||>M\}} f_{T_n}(T_n|\theta)w(\theta)d\theta,$$

is bounded by

$$C_2 \int_{||\beta||>M/C_1} G(C_3||\beta||)d\beta,$$

for positive constants C_2 and C_3, and some integrable function G, for n large enough. Now, from the integrability of G one can conclude that (10) can be made smaller than any pre-assigned positive number. □

We comment that the integrations in this proof are over the parameter space for θ, not over the sample space.

5.2 Proof of Proposition 1.

This result follows from examining the proof of Theorem 19.2 in Bhattacharya & Ranga Rao (1986) and noting the modifications carefully. Essentially, one must retain extra terms in the expansion of the characteristic function in Theorems 9.12 and 9.9 of Bhattacharya & Ranga Rao (1986). We describe the details below.

The integrability of the characteristic function, $\hat{h}$ is equivalent for each θ to the existence of a density for $\bar{Y}$, see BR Theorem 19.1 p. 189. The continuity in θ of the distribution for Y_1 carries over to the continuity of the density for $\bar{Y}$ in θ, when the sample size is large enough, under the conditions stated. The proof in BR p. 190-191 extends to cover this case; we have not recorded the details for this here. The modifications are similar in spirit to the modifications we describe next for getting the main statement of the proposition from the work of BR.

(i) Next, we consider the case that the function g is the identity. When K is not a singleton we must examine the proof of Theorem 19.2 in BR. It is seen that the proof of Theorem 19.2 in BR relies on Theorem 9.12 in BR. Therefore, the proof of Theorem 19.2 carries over to the present setting if Theorem 9.12 in BR which is pointwise in θ extends to compact sets K. Theorem 9.12 uses the hypotheses of Theorem 9.10 and can be proved by extending the technique of proof used in Theorem 9.9 by taking one more term in the Taylor expansion of the characteristic function, CF, (see p. 83, BR). The key assumptions in Theorem 9.9 involve making sure that two small open sets of t's (the argument of the CF $\hat{h}$ of P_θ) must contain zero and be uniformly well behaved over θ's. The first open set is to ensure the Taylor expansion of the CF has uniformly small remainder over a range of θ's. The second open set is to ensure that the CF, $h(\cdot|\theta)$ is uniformly bounded away from one. See BR, p. 77, equation 9.37.

Both of these sets exist by using the continuity of $p(\cdot|\theta)$. That is, the first set exists because the continuity of $p(\cdot|\theta)$ means that as a function of t its CF can be approximated by its Taylor expansion with uniformly small remainder, locally around any θ for t not too far from zero. Similarly, the continuity of $p(\cdot|\theta)$ gives that the t set for the CF to be bounded away from one is good for all θ's in a range.

Thus, the statement of Theorem 19.2 follows uniformly for $\theta \in K$, when $s = 2$. (The case $s \geq 2$ also follows now if we replace $\phi_{0,V}$ with the approximation used in expression 19.17 p. 192 of BR which is $\phi_{0,V}$ plus the first $s-2$ terms of the Edgeworth expansion.)

(ii) For more general functions g, first consider the case $k = d$. We begin by Taylor expanding $g((1/n)\sum_{i=1}^n h(X_i))$ at $\zeta(\theta)$. This gives $V_n = Q^{-1}(\zeta_n, \theta)S_n$, where

$$Q^{-1}(\zeta_n, \theta) = (Dg(\zeta(\theta))\Gamma(\theta)Dg(\zeta(\theta))^t)^{-1/2}Dg(\zeta_n)\Gamma^{1/2}(\theta),$$

$$\zeta_n = \zeta(\theta) + (n^{-1}\sum_{i=1}^{n} h(X_i) - \zeta(\theta)) \otimes \beta_n,$$

and $\beta_n = (\beta_{n1}, ..., \beta_{nd})$ with $0 < \beta_{nk} < 1$ for $k = 1, ..., d$. We have written Q^{-1} here since its inverse Q will be more important below. Also, we have used the notation $\otimes$ to mean

$$a \otimes b = (a_1 b_1, ..., a_d b_d)^t$$

for vectors $a = (a_1, ..., a_d)$ and $b = (b_1, ..., b_d)$. Note that $\zeta_n \to \zeta(\theta)$ a.s., so $Q^{-1}(\zeta_n, \theta) \to Q^{-1}(\zeta(\theta), \theta)$ a.s., so by (6), (7) and Slutsky's theorem, we must have

$$Q^{-1}(\zeta(\theta), \theta)(Q^{-1}(\zeta(\theta), \theta))^t = I_d \quad \text{or} \quad (Q(\zeta(\theta), \theta))^t Q(\zeta(\theta), \theta) = I_d, \quad \forall \theta \in K$$

otherwise the asymptotic variance matrix in (7) cannot be I_d for some $\theta \in K$. Since $rank(Dg(\cdot)) = d$, the Inverse Function Theorem, see Apostol (1974, Theorem 13.6), gives that there is a differentiable inverse $g^{-1}(\cdot)$ for $g(\cdot)$. We have

$$S_n = \sqrt{n}\Gamma^{-1/2}(\theta)\left[g^{-1}\left(\eta(\theta) + n^{-1/2}(Dg\Gamma(\theta)(Dg)^t)^{1/2} V_n\right) - \zeta(\theta)\right],$$

and $D(g^{-1})(\cdot) = (Dg)^{-1}(g^{-1}(\cdot))$, the Jacobian of the transformation is

$$J_n(v) = \left|\frac{\partial S_n}{\partial V_n}\right|(V_n = v) = |\Gamma^{-1/2}(\theta)(Dg)^{-1}\left(\eta(\theta) + n^{-1/2}(Dg\Gamma(\theta)(Dg)^t)^{1/2} v\right)(Dg\Gamma(\theta)(Dg)^t)^{1/2}|$$

$$= |Q(g^{-1}(n^{-1/2}(\eta(\theta) + Dg\Gamma(\theta)(Dg)^t)^{1/2} v), \theta)|.$$

So, the density $f_{V_n}(v|\theta)$ of V_n is

$$f_{V_n}(v|\theta) = J_n(v) f_{S_n}\left(\sqrt{n}\Gamma^{-1/2}(\theta)\left[g^{-1}\left(\eta(\theta) + n^{-1/2}(Dg\Gamma(\theta)(Dg)^t)^{1/2} v\right) - \zeta(\theta)\right]|\theta\right). \quad (14)$$

Taylor expanding $g^{-1}\left(\eta(\theta) + n^{-1/2}(Dg\Gamma(\theta)(Dg)^t)^{1/2} v + \right)$ at $\zeta(\theta) = g^{-1}(\eta(\theta))$, the argument in $f_{S_n}(\cdot|\theta)$ is seen to be

$$Q(g^{-1}(\eta(\theta) + \xi_n \otimes v), \theta)v, \qquad \xi_n = n^{-1/2}(Dg\Gamma(\theta)(Dg)^t)^{1/2}\alpha_n,$$

where $\alpha_n = (\alpha_{n1}(\theta), ..., \alpha_{nd}(\theta))^t$ with $0 < \alpha_{ni}(\theta) < 1$, for $i = 1, ..., d$, $\forall \theta \in K$. Now, replacing f_{V_n} by (14) we have

$$\sup_{\theta \in K} \sup_v (1 + ||v||^s)\left|f_{V_n}(v|\theta) - \phi_d(v) - \sum_{j=1}^{s-2} n^{-j/2} P_j(v)\right|$$

$$= \sup_{\theta \in K} \sup_v (1 + ||v||^s)\left|J_n(v) f_{S_n}\left(Q(g^{-1}(\eta(\theta) + \xi_n \otimes v), \theta)v|\theta\right) - \phi_d(v) - \sum_{j=1}^{s-2} n^{-j/2} P_j(v)\right|. \quad (15)$$

Expression (15) can be bounded from above by adding and subtracting two terms and using the triangle inequality. This gives a three term upper bound:

$$\sup_{\theta \in K} \sup_v (1 + ||v||^s) J_n(v)\left|f_{S_n}\left(Q(g^{-1}(\eta(\theta) + \xi_n \otimes v), \theta)v|\theta\right) - \phi_d(Q(g^{-1}(\eta(\theta) + \xi_n \otimes v), \theta)v)\right.$$

$$\left. - \sum_{j=1}^{s-2} n^{-j/2} P_j(Q(g^{-1}(\eta(\theta) + \xi_n \otimes v), \theta)v)\right| \quad (16)$$

$$+\sup_{\theta\in K}\sup_{v}(1+||v||^s)\left|J_n(v)\phi_d\left(Q(g^{-1}(\eta(\theta)+\xi_n\otimes v),\theta)v\right)-\phi_d(v)\right| \quad (17)$$

$$+\sup_{\theta\in K}\sup_{v}(1+||v||^s)\left|J_n(v)\sum_{j=1}^{s-2}n^{-j/2}P_j(Q(g^{-1}(\eta(\theta)+\xi_n\otimes v),\theta)v)-\sum_{j=1}^{s-2}P_j(v)\right|. \quad (18)$$

Expression (16) is bounded by

$$\sup_u J_n(u)\sup_u\frac{1+||u||^s}{1+||Q(g^{-1}(\eta(\theta)+\xi_n\otimes u),\theta)u||^s}\sup_{\theta\in K}\sup_{v}(1+||v||^s)\left|f_{S_n}(v|\theta)-\phi_d(v)-\sum_{j=1}^{s-2}P_j(v)\right|$$

which is seen to be $o(n^{-(s-2)})$ using Part (i) of the Proposition, and the boundedness of both $J_n(\cdot)$ and $(1+||u||^s)/(1+||Q(g^{-1}(\eta(\theta)+\xi_n\otimes u),\theta)u||^s)$.

It remains to show that (17) and (18) are $o(1)$. It is enough to do this for (17) because the argument for (18) is similar. So, for each n, let (θ_n^*,v_n^*) be the value of (θ,v) achieving the supremum in (17). We begin by distinguishing two cases. First suppose

$$\overline{\lim}_n||v_n^*||<\infty.$$

Observe that $Q^t(\zeta(\theta),\theta)Q(\zeta(\theta),\theta)=I_d$, implies $|Q(\zeta(\theta),\theta)|=1$ because the determinant of a matrix equals the determinant of its transpose. Also, $Q(\cdot,\theta)$ is continuous, $(Dg\Gamma(\theta)(Dg)^t)^{1/2}$ is bounded, and $g^{-1}(\eta(\theta_n^*)+\xi_n\otimes v_n^*)\sim g^{-1}(\eta(\theta_n^*))=\zeta(\theta_n^*)$, where $a_n\sim b_n$ means the limit over n of their ratio equals one. This gives:

$$Q(g^{-1}(\eta(\theta_n^*)+\xi_n\otimes v_n^*),\theta_n^*)\sim Q(\zeta(\theta_n^*),\theta_n^*),$$

$$J_n(v_n^*)\sim|Q(\zeta(\theta_n^*),\theta_n^*)|=1,$$

and

$$\phi_d(Q(\zeta(\theta_n^*),\theta_n^*)v_n^*)=(2\pi)^{-d/2}e^{-\frac{1}{2}(v_n^*)^tQ^t(\zeta(\theta_n^*),\theta_n^*)Q(\zeta(\theta_n^*),\theta_n^*)v_n^*}=\phi_d(v_n^*).$$

Thus (17) is $o(1)$.

Second, suppose

$$\underline{\lim}_n||v_n^*||=\infty.$$

Since $Q(\cdot,\theta_n^*)$ is bounded above and away from singularity by giving conditions, and $J_n(\cdot)$ is bounded, we have

$$\sup_{\theta\in K}\sup_{v\in M}(1+||v||^s)\left|J_n(v)\phi_d\left(Q(g^{-1}(\eta(\theta)+\xi_n\otimes v),\theta)v\right)-\phi_d(v)\right|$$

$$=(1+||v_n^*||^s)\left|J_n(v_n^*)\phi_d\left(Q(g^{-1}(\eta(\theta_n^*)+\xi_n\otimes v_n^*),\theta_n^*)v_n^*\right)-\phi_d(v_n^*)\right|$$

$$\leq(1+||v_n^*||^s)J_n(v_n^*)\phi_d\left(Q(g^{-1}(\eta(\theta_n^*)+\xi_n\otimes v_n^*),\theta_n^*)v_n^*\right)+(1+||v_n^*||^s)\phi_d(v_n^*). \quad (19)$$

Expression (19) tends to zero because $J_n(\cdot)$ is bounded and

$$\lim_{||v||\to\infty}(1+||v||^s)\phi_d(A_nv)=0,$$

when A_n bounded away singularity.

Third, in the intermediate case, we have $\underline{\lim}_n||v_n^*||<\infty$ but $\overline{\lim}_n||v_n^*||=\infty$. Let B be the set of all the limits of $\{v_n^*\}$. If the elements of B are bounded, we are in case 1. If the elements of B are all unbounded, we are in case 2. If some elements of B are finite and ∞ is in B, then for given

$\epsilon > 0$ and n_0, by the argument for case 2, we can find b_r such that for for any sequence $\{n_l\}$ with $\lim_l ||v^*_{n_l}|| \geq b_r$ and $n_l > n_1$,

$$(1+||v^*_{n_l}||^s)\left| J_{n_l}(v^*_{n_l})\phi_d\Big(Q(g^{-1}(\xi_{n_l}\otimes v^*_{n_l}+\eta(\theta^*_{n_l})),\theta^*_{n_l})v^*_{n_l}\Big) - \phi_d(v^*_{n_l})\right| < \epsilon.$$

Also, since the set of elements of B that are less than b_r, $\{b_1, ..., b_{r-1}\}$, is finite, we can find n_2 such that for all sequences $\{n_l\}$ with $\lim_l ||v^*_{n_l}|| \in \{b_1, ..., b_{r-1}\}$,

$$(1+||v^*_{n_l}||^s)\left| J_{n_l}(v^*_{n_l})\phi_d\Big(Q(g^{-1}(\xi_{n_l}\otimes v^*_{n_l}+\eta(\theta^*_{n_l})),\theta^*_{n_l})v^*_{n_l}\Big) - \phi_d(v^*_{n_l})\right| < \epsilon.$$

Now the limits of $\{v^*_n\}$ fall into the above two situations, and in either situation the conclusion of (ii) of the Proposition is true since $\epsilon > 0$ is arbitrary.

Now, consider the case $k < d$. Since $rank(Dg) = rank(Dg\Gamma) = k$, for each fixed θ, the equations

$$Dg(\zeta(\theta))\Gamma(\theta)\mathbf{x} = \mathbf{0}$$

have $d-k$ linearly independent solutions in $\mathbf{x}$. We denote a set of such solutions by $(Dg^{ind}(\zeta(\theta)))^t$, so

$$Dg(\zeta(\theta))\Gamma(\theta)(Dg^{ind}(\zeta(\theta)))^t = \mathbf{0}.$$

Since $Dg(\cdot)$ is continuous and bounded on a neighborhood of $\zeta(\theta))$, so is $Dg^{ind}(\cdot)$. Also, it is seen that $Dg^{ind}(\cdot)$ is linearly independent of $Dg(\cdot)$: Otherwise there would be a $(d-k)\times k$ matrix of the form $A = (a_1, ..., a_{d-k})^t$, in which the a_j's are d-vectors satisfying $Dg^{ind}(\zeta(\theta)) = ADg$. If this were true we would have

$$Dg(\zeta(\theta))\Gamma(\theta)(Dg^{ind}(\zeta(\theta)))^t = Dg(\zeta(\theta))\Gamma(\theta)(Dg(\zeta(\theta)))^t A^t = \mathbf{0},$$

or

$$Dg(\zeta(\theta))\Gamma(\theta)(Dg(\zeta(\theta)))^t a_j = \mathbf{0}, \quad j = 1, ..., d-k.$$

However, both are impossible since $Dg(\zeta(\theta))\Gamma(\theta)(Dg(\zeta(\theta)))^t$ is a non-degenerate variance matrix.

Now, let $g^{ind}(\cdot)$ be a function with $Dg^{ind}(\cdot)$, $\tilde{g}(\cdot) := (g(\cdot), g^{ind}(\cdot))^t$ and $D\tilde{g} = (Dg(\cdot), Dg^{ind}(\cdot))$. We see that $\tilde{g}(\cdot)$ is $\mathbb{R}^d \mapsto \mathbb{R}^d$, has rank$(D\tilde{g}) = d$, and satisfies all the conditions in (ii) of the Proposition for the case $k = d$.

Write $\eta^{ind}(\theta) = g^{ind}(\zeta(\theta))$, $\tilde{\eta}(\theta) = (\eta(\theta), \eta^{ind}(\theta))$,

$$\tilde{T}_n = (g\{(1/n)\sum_{i=1}^n h(X_i)\}, g^{ind}\{(1/n)\sum_{i=1}^n h(X_i)\}) := (T_n, T_n^{ind}),$$

and

$$\tilde{V}_n = \sqrt{n}\tilde{\Omega}(\theta)^{-1/2}(\tilde{T}_n - \tilde{\eta}(\theta)) = (V_n, V_n^{ind}) \xrightarrow{d} N(\mathbf{0}, I_d),$$

where $\tilde{\Omega}(\theta) = D\tilde{g}(\zeta(\theta))\Gamma(\theta)(D\tilde{g}(\zeta(\theta)))^t$ is

$$\begin{pmatrix} \Omega(\theta) & \mathbf{0} \\ \mathbf{0} & \Omega^{ind}(\theta) \end{pmatrix}$$

$$= \begin{pmatrix} Dg(\zeta(\theta))\Gamma(\theta)(Dg(\zeta(\theta)))^t & \mathbf{0} \\ \mathbf{0} & Dg^{ind}(\zeta(\theta))\Gamma(\theta)(Dg^{ind}(\zeta(\theta)))^t \end{pmatrix}.$$

The inverse root is given by

$$\tilde{\Omega}(\theta)^{-1/2} = \begin{pmatrix} \Omega(\theta)^{-1/2} & \mathbf{0} \\ \mathbf{0} & \Omega^{ind}(\theta)^{-1/2} \end{pmatrix},$$

so we set $V_n^{ind} = \sqrt{n}\Omega^{ind}(\theta)^{-1/2}(T_n^{ind} - \eta^{ind}(\theta))$.

Now the result for the case that $D\tilde{g}(\cdot)$ is of full rank gives that $\tilde{v} = (v, v^{ind})^t$ satisfies

$$f_{\tilde{V}_n}(\tilde{v}|\theta) - \phi_d(\tilde{v}) - \sum_{j=1}^{s-2} n^{-j/2} P_j(\tilde{v}) = \frac{o(1)}{1 + ||\tilde{v}||^s}, \quad \text{uniformly for } \theta \in K.$$

Since $\int f_{\tilde{V}_n}(\tilde{v}|\theta) dv^{ind} = f_{V_n}(v|\theta)$, integrating over v^{ind} on both sides of last expression gives

$$f_{V_n}(v|\theta) - \phi_k(v) - \sum_{j=1}^{s-2} n^{-j/2} P_j(v) = \frac{o(1)}{1 + ||v||^s}, \quad \text{uniformly for } \theta \in K.$$

5.3 Proof of Theorem 2.

The left hand side of (8) is

$$\int\int m_{T_n}(t) \left| \frac{f_{T_n}(t|\theta) w(\theta)}{m_{T_n}(t)} - \phi_{\eta^{-1}(t), \Sigma\{\eta^{-1}(t)\}/n}(\theta) \right| d\theta dt$$

$$= \int\int |f_{T_n}(t|\theta) w(\theta) - m_{T_n}(t) \phi_{\eta^{-1}(t), \Sigma\{\eta^{-1}(t)\}/n}(\theta)| d\theta dt$$

$$= \int\int \Big| n^{d/2} |\Omega(\theta)^{-1/2}| f_{V_n}\left[\sqrt{n}\Omega(\theta)^{-1/2}\{t - \eta(\theta)\}|\theta\right] w(\theta)$$

$$- \frac{n^{d/2}}{|\Sigma\{\eta^{-1}(t)\}|^{1/2}} \phi_d\left[\sqrt{n}\Omega\{\eta^{-1}(t)\}^{-1/2} D\eta\{\eta^{-1}(t)\}^{-1}\{\theta - \eta^{-1}(t)\}\right] m_{T_n}(t) \Big| dt d\theta. \quad (20)$$

Change variables from t to $r = \sqrt{n}\Omega(\theta)^{-1/2}\{t - \eta(\theta)\}$ in the inside integral and adopt the notation $t(r) = n^{-1/2}\Omega(\theta)^{1/2} r + \eta(\theta)$ to write (20) as

$$\int\int \Big| g_{1,n}(r, \theta) - g_{2,n}(r, \theta) \Big| dr d\theta \quad (21)$$

where $g_{1,n}$ and $g_{2,n}$ are the joint densities

$$g_{1,n}(r, \theta) = f_{V_n}(r|\theta) w(\theta)$$
$$g_{2,n}(r, \theta) = B_n \phi_d\{A_n(\theta, r)\} m_{T_n}\{t(r)\}$$

and we use the notation

$$A_n(\theta, r) = \sqrt{n}\Omega[\eta^{-1}\{t(r)\}]^{-1/2} D\eta[\eta^{-1}\{t(r)\}]^{-1}[\theta - \eta^{-1}\{t(r)\}]$$
$$B_n = \frac{|\Omega(\theta)|^{1/2}}{|\Sigma[\eta^{-1}\{t(r)\}]|^{1/2}}.$$

Now, for each fixed θ and r , we have that $A_n(\theta, r) \to r$. Also we have that $B_n \to |D\eta(\theta)|$ since $\Sigma(\theta) = (D\eta)^{-1}(\theta)\Omega(\theta)\{(D\eta)^{-1}(\theta)\}^t$.

By assumption,

$$g_{1,n}(r, \theta) = f_{V_n}(r|\theta) w(\theta) \to \phi_d(r) w(\theta),$$

for each fixed (r, θ). Also, we get

$$m_{T_n}\{t(r)\} \to w(\theta)|D\eta^{-1}(\theta)|,$$

by using (13) with $h(\cdot) \equiv 1$. From this limit and those for A_n and B_n we get, for each fixed (r, θ),

$$g_{2,n}(r,\theta) = B_n \phi_d\{A_n(\theta, r)\} m_{T_n}\{t(r))\} \to \phi_d(r) w(\theta).$$

Next, add and subtract $\phi_d(r)w(\theta)$ in the absolute value signs in (21) and use the triangle inequality. Finally, Scheffé's theorem, see Serfling (1980, p. 17), gives that the right hand side of (21) goes to zero, as $n \to \infty$.

ACKNOWLEDGEMENTS

The authors thank Richard Lockhart and an anonymous Associate Editor for extensive and insightful feedback. Clarke was supported by NSERC Operating Grant 138122-00.

REFERENCES

T. M. Apostol (1974). *Mathematical Analysis*, Second Edition, Addison-Wesley.

J. M. Bernardo (1979). Reference posterior distributions for Bayesian inference (with Discussion). *Journal of the Royal Statististical Society B*, 40, 113–147.

R. N. Bhattacharya & R. Ranga Rao (1976). *Normal Approximation and Asymptotic Expansions.* Robert E. Krieger Publishing Company, Malabar FL.

P. Bickel & J. A. Yahav (1969). Some contributions to the asymptotic theory of Bayes solutions. *Zeitschrift fur wahrscheinlichkeitstheorie und verwandte gebiete*, 11, 257–276.

P. Billingsley (1986). *Probability and Measure*, Wiley.

B. Clarke (1999). Asymptotic normality of the posterior in relative entropy, *IEEE Transactions on Information Theory*, 45, 165–176.

B. Clarke & J. K. Ghosh (1995). Posterior normality given the mean with applications, *Annals of Statistics*, 23, 2116–2144.

B. Clarke & A. Yuan (2003). Partial Information Reference Priors: Derivation and interpretation. *Journal of Statistical Planning and Inference*, in press.

K. Doksum & A. Lo (1990). Consistent and robust Bayes procedures for location based on partial information. *Annals of Statistics*, 18, 443–453.

D. Fraser & P. McDunnough (1984). Further remarks on asymptotic normality of likelihood and conditional analyses. *The Canadian Journal of Statistics*, 12, 183–190.

S. Ghosal, J. K. Ghosh, & T. Samanta (1995). On convergence of posterior distributions, *Annals of Statistics*, 23, 2145–2152.

J. Hartigan (1983). *Bayes Theory.* Springer-Verlag, New York.

P. Hall (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

B. Junker (1993). Conditional association, essential independence, and monotone unidimensional item response models. *Annals of Statistics*, 21, 1359–1378.

L. Le Cam (1953). On the asymptotic properties of maximum likelihood and related Bayes estimates. In: *University of California Publications in Statistics*, Eds. J. Neyman, M. Loeve & O. Struve, Cambridge University Press, London, 277–329.

L. Le Cam (1958). Les propriétes asymptotique des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris*, 7, 18–35.

R. D. Reiss (1989). *Approximate Distributions of Order Statistics.* Springer-Verlag, New York.

R. J. Serfling (1980) *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

T. A. Severini (1991). On the relationship between Bayesian and non-Bayesian interval estimates. *Journal of the Royal Statistical Society B*, 53, 611–618.

T. A. Severini (1993). Bayesian interval estimates which are also confidence intervals. *Journal of the Royal Statistical Society B*, 55, 533–540.

H. Strasser (1981). Consistency of maximum likelihood and Bayes estimates. *Annals of Statistics*, 9, 1107–1113.

T. Sweeting (1992). On asymptotic posterior normality in the multiparameter case. In: *Bayesian Statistics IV: Proceedings of the Fourth Valencia International Meeting*, Eds. J. Bernardo, J. .O. Berger, A. P. Dawid & A. F. M. Smith, Oxford University Press, New York, 825–835.

A. M. Walker (1969). On the asymptotic behavior of posterior distributions, *Journal of the Royal Statistical Society B*, 31, 80–88.

A. Yuan & B. Clarke (1993). Posterior normality given an order statistic. Unpublished manuscript.

Received 12 Sept 2000
Accepted 5 April 2003

Ao YUAN: `ayuan@howard.edu`
National Human Genome Center, Howard University
Washington D.C., USA 20059

Bertrand CLARKE: `riffraff@stat.ubc.ca`
Department of Statistics, UBC
Vancouver, British Columbia, Canada V6T 1Z2