POSTERIOR CONVERGENCE GIVEN THE MEAN

by

B. Clarke and J.K. Ghosh
Purdue University and the Indian Statistical Institute

Technical Report #91-51

Department of Statistics
Purdue University

September 1991

◇

Posterior Convergence Given the Mean

B. Clarke and J. K. Ghosh*
Purdue University and the Indian Statistical Institute*

Abstract

For various applications, one wants to know the asymptotic behavior of the posterior density of a finite dimensional real parameter given the mean of the data rather than the full data set. Such a result is only nontrivial when the mean is not sufficient and examples are easy to construct. Here we show that for lattice valued random variables $X_1, \ldots, X_n, \ldots$, assumed independently and identically distributed, that the posterior density $w(\theta|\bar{X})$ - where $\bar{X}$ is the mean - is asymptotically normal in an L^1 sense, and we identify the location and the asymptotic variance as a function of the unknown parameter θ.

§1 Introduction

For ease of exposition suppose we have data from a standardized test, and the data is believed to be from a known finite dimensional parametric family. The parameters of interest, together with any parameters of secondary importance, may represent latent traits in the sense of Holland and Rosenbaum (1986). We represent them as $\theta = (\theta_1, \ldots, \theta_d)$, where the entries are real, distributed according to the bounded density $w(\theta_1, \ldots, \theta_d)$ taken with respect to Lebesgue measure. The goal is to estimate θ for each examinee based on the examinee's answers to n multiple choice questions, termed test items.

The worth of an examinee's answer to the i^{th} test item is represented as a random variable X_i, taking values in a regular k-dimensional minimal lattice L, with common step length ℓ. Now the joint distribution of the two stage experiment is

$$w(\theta)p_\theta(X_1), \ldots, p_\theta(X_n),$$

where $p_\theta(X_i)$ is the probability mass function for the i^{th} test-item. We assume that the probability mass functions $p_{\theta,i}$ are identical; this is very hard to ensure in practice but does not invalidate the interpretation. For brevity we write $X^n = (X_1, \ldots, X_n)$, and denote the parameter space by $\Omega \subset \mathbb{R}^d$.

Given that the number of test items is sufficiently large it is of interest to ask whether estimation of θ can be based on summary statistics such as the mean, $\bar{X}$. In this context estimation often means looking at highest posterior density (HPD) regions of $w(\theta|\bar{X})$. As a consequence the asymptotic behavior of the posterior density of the parameter given the mean is of interest. Aside from the convenience of such a procedure, it is already virtually always used informally. Furthermore, it is desirable from a modelling standpoint since $\bar{X}$ can be interpreted as the marking scheme.

It has already been conjectured that $w(\theta|\bar{X})$ should be asymptotically normal. Indeed, if $\bar{X}$ is sufficient then $w(\theta|\bar{X}) = w(\theta|X^n)$ by the factorization criterion so the usual asymptotic normality results apply. In such cases $w(\theta|\bar{X})$ converges to a normal with location related to $\bar{X}$, and standard error given by the second moments of the X_i's. In some cases this may be derived from earlier work. Le Cam (1953) proves a version of the desired result for the maximum likelihood estimator in place of the mean and Doksum and Lo (1990) establish a form of the result for location families. Their result allows for estimators other than the mean.

It is easy to produce examples of probability mass functions which have identical conditional distributions for $(X_i|\theta)$ but still do not admit $\bar{X}$ as a sufficient statistic. In the simplest examples $d = k = 1$ and the random variables take one of three values, zero, one and two. The probability mass function for n outcomes is

$$p^n_{\theta_1}(X^n) = \frac{e^{n\theta_1(\alpha\bar{X}_{0,1}+\beta\bar{X}_2)}}{(1+e^{\alpha\theta_1}+e^{2\beta\theta_1})^n}, \tag{1.1}$$

in which α, $\beta > 0$, with $\alpha \neq \beta$, are assumed known, and $\bar{X}_{0,1} = \sum_{i=1}^{n} X_i\chi_{\{X_i=0,1\}}$, $\bar{X}_2 = \sum_{i=1}^{n} X_i\chi_{\{X_i=2\}}$. This family of mass functions is of exponential form so $(\alpha/n)\bar{X}_{0,1}+(\beta/n)\bar{X}_2$ is sufficient for θ_1, but different from $\bar{X}$.

One can force an educational testing interpretation on (2): Each item is worth two points, and partial credit is awarded for one of the wrong answers. The α and β are slope parameters indicating how the probability of giving a wrong answer, say that corresponding to the zero or 1 or the right answer, say that corresponding to two, changes with increasing θ_1. The test items are identical in terms of the demand placed on the examinee and $\bar{X}$ is the examinees score.

When $\bar{X}$ is not sufficient conventional Bernstein-von Mises results (see Le Cam (1958), Bickel and Yahav (1969), Walker (1969); there are many others) do not apply, however, we still expect some form of normality to be achieved asymptotically, with larger standard error since $\bar{X}$ represents less information than the full data set. The consequence is that the standard error of the limiting normal of $w(\theta|\bar{X})$ should be different from the standard error of $w(\theta|X^n)$ and the discrepancy between them tracks the degree to which $\bar{X}$ fails to be sufficient.

The result in Section 2 applies to the mass function in (1.1). In general, denote expectation under $p^n_{\theta^d_1}$ by E_θ; so that the mean of any X_i is the k-dimensional vector $\mu(\theta)$ and its variance is the $k \times k$ matrix $\Sigma(\theta_1)$. In this case $k = 1$ and it is easy to find $\mu(\theta_1)$ and $\Sigma(\theta_1)$. So the asymptotic variance which is $n\mu'(\theta_1)^t\Sigma^{-1}(\theta_1)\mu'(\theta_1)$ can be evaluated explicitly. The location of the limiting normal is $\mu^{-1}(\bar{X})$. Since the derivative $\mu'(\theta)$ is obviously not identically zero μ^{-1} can be well defined on a neighborhood of almost every value of the parameter. As the other hypotheses of Theorem 2.1 are satisfied we have that

$$w(\theta|\bar{X}) - n(\theta;\theta_0,\hat{\theta}) \overset{L_1(P_{\theta_0})}{\longrightarrow} 0$$

where θ_0 is a fixed value of θ_1, θ is a variable of integration, $\hat{\theta} = \mu^{-1}(\bar{X})$ and the limiting normal is

$$n(\theta;\theta_0,\hat{\theta}) = \frac{\sqrt{n\mu'(\theta_0)^2\Sigma^{-1}(\theta_0)}e^{-(n/2)\mu'(\theta_0)^2\Sigma^{-1}(\theta_0)(\theta-\hat{\theta})^2}}{\sqrt{2\pi}}.$$

Thus HPD confidence regions for $(\theta|\bar{X})$ can be approximately determined from the $N(\hat{\theta}, \Sigma(\theta_0)/\mu'(\theta_0)^2)$ distribution.

In addition to such applications in estimation it has been shown that an appropriate form of asymptotic normality of $w(\theta|\bar{X})$ has applications to testing independence of test items, see Junker (1991). Also, Ackerman (1991) assumes such a result for the purpose of evaluating the influence of dimensionality of a parameter on test item bias. The limitation to the $k = d$ case is consistent with the psychometric orthodoxy which strongly favors a unidimensional parameter and test items which take values in a unidimensional lattice.

The latter application brings out a methodological implication. In general the asymptotic variance is $nJ_\mu(\theta_0^t)\Sigma^{-1}(\theta_0)J_\mu(\theta_0)$ where $J_\mu(\theta_0)$ is the $k \times d$ Jacobian matrix of $\mu(\theta)$ regarded as a function $\mu : \Omega \to L$. It is seen that the resulting normal is nondegenerate only when $d \le k$. The consequence is that if the data is strongly multidimensional then test items must be combined if normality is to hold asymptotically. That is, if all d parameters are essential to the modelling procedure then test items, commonly taking values in a one-dimensional lattice with finitely many points having positive probability, must be grouped together in sets of at least d items so that the dimension of the lattice is also at least d.

The structure of the paper is as follows. In Section 2 we state and prove our results for the case of independent and identical lattice valued random variables with a compact parameter space. There are two main results: The first guarantees the uniformity of a local limit theorem so that mixtures of distributions can be approximated; the second gives the desired result assuming that the number of parameters equals the dimension of the lattice. In section 3 we give a generalization to noncompact parameter spaces by controlling tail behavior of the posterior, and consider the case that the number of parameters is strictly less than the dimension of the lattice.

§2 Compact parameter space

Recall that by Bayes rule we can write the joint density for $(\Theta, \bar{X})$ as

$$w(\theta)p_\theta(\bar{X}) = w(\theta|\bar{X})m(\bar{X}) \tag{2.1}$$

where $w(\theta|\bar{X})$ is the posterior density for Θ given $\bar{X}$ and $m(\bar{X})$ is the mixture of densities

$$m(\bar{X}) = \int_\Omega w(\theta)p_\theta(\bar{X})d\theta. \tag{2.2}$$

We show that $w(\theta|\bar{X})$ is asymptotically normal in an L^1 sense, and identify the location and scale of the asymptotic normal. Our main assumptions are on moments, and on the characteristic function $f(\theta,t) = E_\theta \exp i(t, X_1)$.

The intuition behind the proof is to approximate $p_\theta(\bar{X})$ by a normal density uniformly in θ. This implies that $m(\bar{X})$ can be well approximated by a mixture of normals. Since convergence is assessed under a fixed member p_{θ_0} of the parametric family, the posterior density for θ concentrates on a shrinking neighborhood located at $\hat{\theta}$, which converges to θ_0, allowing identification of the asymptotic variance.

Central to the proof is a three term upper bound on the L^1 distance between $w(\theta|\bar{X})$ and the target normal denoted $n(\theta;\theta_0,\hat{\theta})$. The three terms result from using three normal approximations. The first is the target normal itself

$$n(\theta;\theta_0,\hat{\theta}) = \frac{\sqrt{|nJ_\mu^t(\theta_0)\Sigma^{-1}(\theta_0)J_\mu(\theta_0)|}\, e^{-(\frac{n}{2})(\theta-\hat{\theta})^t J_\mu^t(\theta_0)\Sigma^{-1}(\theta_0)J_\mu(\theta_0)(\theta-\hat{\theta})}}{(2\pi)^{d/2}}, \tag{2.3}$$

where $|\cdot|$ denotes the determinant, and $\hat{\theta} = \mu^{-1}(\bar{X})$, which need only be defined near θ_0. Note that $\hat{\theta}$ is not the estimate obtained by maximizing the likelihood of $\bar{X}$ but a convenient approximation which suffices for our purpose. Typically μ^{-1} will only make sense when k and d are equal since the image of $\mu: \Omega \to \mathbb{R}^k$ is a d dimensional surface. Although we require in this section that $k = d$, we distinguish between the dimension of the parameter space and the lattice for the sake of generality. Note that in (2.3) the variance is no longer $\Sigma(\theta_0)$, but a modification depending on the parametrization. In places where the slope of μ changes rapidly as a function of the true value the variance increases, where μ is relatively constant the variance in fact decreases.

The second normal approximation is obtained from a local limit theorem, proved to hold uniformly over compact sets in the parameter space. It is well known that the density of $\bar{X}$ can be approximated by a sum whose leading term is a normal density and successive terms are normal densities multiplied by polynomials. The rate at which the distance in supremum norm between $p_\theta(\bar{X})$ and its closest approximation of this type tends to zero depends on the number of moments assumed to exist. One such result can be found in Bhattacharya and Rao (1976, §22) henceforth referred to as BR. We write

$$q_{\theta,r,n}(\bar{X}) = \frac{\ell}{n^{k/2}} \sum_{i=1}^{r} \frac{f_i(\sqrt{n}(\bar{X} - \mu(\theta)))}{n^{(i-1)/2}} \varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X} - \mu(\theta))) \tag{2.4}$$

for the r term approximation to $p_\theta(\bar{X})$, where f_i is a polynomial of degree $3r$ in k variables and $\varphi_{\Sigma(\theta)}$ is the normal density with mean 0 and variance $\Sigma(\theta)$.

The third normal approximation is a variant on (2.4), to wit,

$$q_{\theta,\theta_0,r,n}(\bar{X}) = \frac{\ell}{n^{k/2}} \sum_{i=1}^{r} \frac{f_i(\sqrt{n}(\bar{X} - \mu(\theta)))}{n^{(i-2)/2}} \varphi_{\Sigma(\theta_0)}(\sqrt{n}(\bar{X} - \mu(\theta))) \tag{2.5}$$

in which the variance matrix is evaluated at θ_0. The approximations (2.4) and (2.5) are used to define mixtures with respect to θ are denoted

$$m_r(\bar{X}) = \int_\Omega w(\theta) q_{\theta r}(\bar{X}) d\theta, \tag{2.6}$$

and

$$m_{r,\theta_0}(\bar{X}) = \int_\Omega w(\theta) q_{\theta,\theta_0,r}(\bar{X}) d\theta, \tag{2.7}$$

respectively, where, for brevity we have omitted the n's on densities.

Shrinking neighborhoods in the sample space and in the parameter space are essential to the proof. We denote them

$$U_{n,\theta_0} = \{X^n : ||\bar{X} - \mu(\theta_0)|| \leq \frac{k_n}{\sqrt{n}}\} \tag{2.8}$$

and

$$U'_{n,\theta_0} = \{\theta : ||\mu(\theta) - \mu(\theta_0)|| \le \frac{k'_n}{\sqrt{n}}\} \tag{2.9}$$

where $k_n/\sqrt{n}$, $k'_n/\sqrt{n} \to 0$ and $||\cdot||$ is a norm on the lattice L, assumed to be embedded in k-dimensional real space. The defining condition in (2.8) can be equivalently expressed as $||\mu(\hat{\theta}) - \mu(\theta_0)|| \le k_n/\sqrt{n}$. To permit upper bounds, Taylor expansions can be used to obtain sets containing U_{n,θ_0} and U'_{n,θ_0}. The defining conditions become $||\hat{\theta}-\theta_0|| \le k_n/\alpha\sqrt{n}$ and $||\theta - \theta_0|| \le k'_n/\alpha\sqrt{n}$ where $\alpha = \inf ||\nabla\mu(\theta')||$ and the infimum is over θ' in a ball of radius ε centered at θ_0. Again, we typically drop the subscript n's on U and U'. The rates of shrinkage of the neighborhood that are seen to be most useful are $k_n = c(\ell n\ n)^{1/2}$ and $k'_n = c'(\ell n\ n)^{1/2}$, where c', $c > 0$ and $c' - c > 0$. It will be seen that choosing $c' - c$ large enough gives the desired convergence.

First we state and prove a uniform version of Theorem 22.1 in BR.

Proposition 2.1: For $r \ge 1$ suppose that

$$E_\theta ||X_1 - \mu(\theta)||^{r+2}$$

is continuous as a function of $\theta \in K$ compact. Assume also that the function

$$f(\theta, t) = E_\theta e^{i(t,X_1)}$$

for $t \in \mathbb{R}^k$ is jointly continuous in its two arguments. Then provided that $\Sigma(\theta)$ is positive definite on K

$$\sup_{\theta\in K}\sup_{\alpha\in L}(1 + ||\frac{\alpha - n\mu(\theta)}{\sqrt{n}}||^{r+1})|p_\theta(\frac{\alpha}{n}) - q_{\theta r}(\frac{\alpha}{n})| = O(\frac{1}{n^{(k+r)/2}}) \tag{2.10}$$

Proof: For fixed θ we have the desired rate: From BR we have that

$$\sup_{\alpha}(1 + ||\frac{\alpha - n\mu(\theta)}{\sqrt{n}}||^{r+1})|p_\theta(\frac{\alpha}{n}) - q_{\theta r}(\frac{\alpha}{n})| = o(\frac{1}{n^{(k+r-1)/2}})$$

and

$$\sup_{\alpha}(1 + ||\frac{\alpha - n\mu(\theta)}{\sqrt{n}}||^{r+1})|p_\theta(\frac{\alpha}{n}) - q_{\theta r+1}(\frac{\alpha}{n})| = o(\frac{1}{n^{(k+r)/2}})$$

By the triangle inequality we have

$$
\begin{aligned}
&(1+||\frac{\alpha-n\mu(\theta)}{\sqrt{n}}||^{r+1})|p_\theta(\frac{\alpha}{n})-q_{\theta r}(\frac{\alpha}{n})| \\
&\quad\le (1+||\frac{\alpha-n\mu(\theta)}{\sqrt{n}}||^{r+1})|p_\theta(\frac{\alpha}{n})-q_{\theta r+1}(\frac{\alpha}{n})| \\
&\qquad + (1+||\frac{\alpha-n\mu(\theta)}{\sqrt{n}}||^{r+1})|q_{\theta r+1}(\frac{\alpha}{n})-q_{\theta r}(\frac{\alpha}{n})| \\
&\quad\le o(\frac{1}{n^{k+r/2}})+(1+||\frac{\alpha-n\mu(\theta)}{\sqrt{n}}||^{r+1})\frac{f_{r+1}(\sqrt{n}(\bar{X}-\mu(\theta))}{n^{k/2}n^{r/2}}\varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X}-\mu(\theta)))
\end{aligned}
$$

in which the last factor is the $r+1$ term in the normal expansion. The product $f_{r+1}(\sqrt{n}(\bar{X}-\mu(\theta)))\varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X}-\mu(\theta)))$ can be bounded above by a constant. (Indeed $e^{-n||\bar{X}-\mu(\theta)||^2}\ ||\sqrt{n}(\bar{X}-\mu(\theta))||^{3r}$ is maximized at $\sqrt{n}||\bar{X}-\mu(\theta)|| = \pm\sqrt{3r/2}$ which gives $e^{-(3/2)r}(3r/2)^{3r/2}$.)

To finish the proof it remains to show that the BR result holds uniformly over compact sets. Fix $\theta_0 \in K$. For a sufficiently small neighborhood U_{θ_0} of θ_0 the two t-sets in the proof of BR's result can be chosen so as to satisfy (i) the expansion for the characteristic function holds with uniformly small remainder and (ii) on the second t-set $f(\theta,t)$ for $\theta \in U_{\theta_0}$ is uniformly bounded away from unity.

As a consequence of (i) and (ii) it is seen that I_1 and I_2 in BR notation tend to zero uniformly at rate $o(1/n^{k+d/2})$. Also, if U_{θ_0} is small enough that the first $r+2$ moments are bounded on U_{θ_0}, then I_3 also goes to zero uniformly.

Now for each θ_0 there is a U_{θ_0} so that the normal approximation is uniformly valid there. By the Heine-Borel theorem the proof is complete. $\square$

We use the uniformity in Proposition 2.1 to prove the basic result for compact parameter spaces. It will be seen that later results follow by suitably modifying the hypotheses or extending the technique of proof.

Theorem 2.1: Let $\Omega \subset \mathbb{R}^d$ compact be the closure of an open set and suppose the X_i are all drawn from p_{θ_0} where $\theta_0 \in \Omega$, taking values in the k dimensional regular lattice L, with common step length ℓ. Assume that on Ω $\mathrm{Var}_\theta X_1 = \Sigma(\theta)$ satisfies

$$\eta_1 Id \le \Sigma(\theta) \le \eta_2 Id$$

for some $\eta_1, \eta_2 > 0$, where Id is the $k \times k$ identity matrix, and that the entries of $\Sigma(\theta)$ are continuously differentiable. Assume also that $\mu(\theta) = E_\theta X_1$ has two continuous derivatives, is locally invertible at θ_0, and its $d \times k$ derivative matrix $J_\mu(\theta)$ has rank d at $\theta = \theta_0$ where $d = k$. Then, if the hypotheses of Proposition 2.1 are satisfied with $r \geq d$ we have that

$$E_{\theta_0} \int |w(\theta|\bar{X}) - n(\theta; \theta_0, \hat{\theta})| d\theta \to 0 \tag{2.11}$$

as $n \to \infty$.

Remark 1) If $d < k$ then the technique of proof breaks down and so must be modified, see Section 3.

Remark 2) We use K to denote a positive constant for bounding purposes not in general the same from occurrence to occurrence.

Proof: We proceed in four steps. The first step is to obtain lower bounds on $m_r(\bar{X})$ and $\chi_U |m(\bar{X}) - m_r(\bar{X})|$, and note a straightforward upper bound on (2.11) which has 3 terms. The following 3 steps will deal with each term in turn.

Step 1, part 1: We show that there is a $K > 0$ so that

$$m_r(\bar{X}) \geq K/n^{(k+d)/2}. \tag{2.12}$$

First note that since products of the form $f_i(\sqrt{n}(\bar{X} - \mu(\theta)))\varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X} - \mu(\theta)))$ are bounded in absolute value by constants for $i \geq 2$ (as in the proof of Proposition 2.1) we can write

$$\begin{aligned} m_r(\bar{X}) &\geq K \int_\Omega \frac{e^{-(n/2)(\mu(\hat{\theta})-\mu(\theta))^t \Sigma^{-1}(\theta)(\mu(\hat{\theta})-\mu(\theta))}}{n^{k/2}} \\ &\geq \frac{K}{n^{k/2}} \int_{||\theta-\hat{\theta}|| \leq \frac{k'_n}{\sqrt{n}}} e^{-(n/2)(\mu(\hat{\theta})-\mu(\theta))^t \Sigma^{-1}(\theta)(\mu(\hat{\theta})-\mu(\theta))} d\theta \\ &\geq \frac{K}{n^{k/2}} \int_{||\hat{\theta}-\theta|| \leq \frac{k'_n}{\sqrt{n}}} e^{-(n/2)(\hat{\theta}-\theta)^t J_\mu(\tilde{\theta})^t \Sigma^{-1}(\theta) J_\mu(\tilde{\theta})(\hat{\theta}-\theta)} d\theta \end{aligned}$$

by a Taylor expansion, where $\tilde{\theta}$ lies on the straight line joining θ and $\hat{\theta}$. Since $J_\mu(\tilde{\theta})^t \Sigma^{-1}(\theta)$ $J_\mu(\tilde{\theta})$ is positive definite, bounded above and bounded away from singularity the last expression can be lower bounded by using the transformation $\varphi = \sqrt{n}(\theta - \hat{\theta})$ so as to give a positive probability from a normal and a factor $1/n^{d/2}$ from the Jacobian, thus proving (2.12).

Step 1, Part 2: We show that

$$\chi_U \left| m(\bar{X}) - m_r(\bar{X}) \right| \leq K \chi_U (k_n')^d / n^{(k+r+1)/2} \tag{2.13}$$

where χ_U is the indicator function for U. The left hand side can be bounded above by

$$\begin{aligned}
&\chi_U \int_{U'} |p_\theta(\bar{X}) - q_{\theta,r}(\bar{X})| w(\theta) d\theta + \chi_U \int_{U'^c} |p_\theta(\bar{X}) - q_{\theta,r}(\bar{X}) w(\theta) d\theta \\
&\leq \chi_U \left[\sup_{\theta \in U'} w(\theta) \right] \text{Vol } (U') \frac{K}{n^{(k+r)/2}} + \chi_U \int_{U'^c} |p_\theta(\bar{X}) - q_{\theta,r+1}(\bar{X})| w(\theta) d\theta \\
&\quad + \chi_U \int_{U'^c} |q_{\theta,r+1}(\bar{X}) - q_{r\theta}(\bar{X})| w(\theta) d\theta \\
&\leq \chi_U \frac{K(k_n')^{d/2}}{n^{(k+d+r)/2}} + \chi_U \frac{K}{n^{(k+r+1)/2}} \\
&\quad + \chi_U \int_{U'^c} w(\theta) \frac{|f_{r+1}(\sqrt{n}(\bar{X} - \mu(\theta)))|}{n^{r/2}} \varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X} - \mu(\theta))) d\theta \\
&\leq \chi_U \frac{K(k_n')}{n^{(k+d+1)/2}} + \chi_U \int_{V_n} w(\theta) |f_{r+1} \frac{(\sqrt{n}(\bar{X} - \mu(\theta)))}{n^{r/2}} | \varphi_{\Sigma(\theta)}(\sqrt{n}(\bar{X} - \mu(\theta))) d\theta
\end{aligned} \tag{2.14}$$

where $U'^c \subset V_n$ defined by

$$V_n = \{\theta : \; ||\mu(\theta) - \bar{X}|| \geq (c' - c) \sqrt{\frac{\ell n \; n}{n}} \}$$

using the triangle inequality since the inequalities in U and U'^c go in opposite directions.

To show (2.13) it is enough to control the integral term in the last upper bound. First note that it is bounded by

$$\chi_U \frac{K}{n^{r/2}} e^{-(n/4)\eta_1 (c'-c)^2 (\ell n \; n/n)} \int w(\theta) |f_{r+1}(\sqrt{n}(\bar{X} - \mu(\theta)))| \varphi_{\frac{1}{2}\Sigma(\theta)}(\sqrt{n}(\bar{X} - \mu(\theta))) d\theta. \tag{2.15}$$

The product $f_{r+1}\varphi_{\frac{1}{2}\Sigma(\theta)}$ is uniformly bounded by a constant as in Proposition 2.1 so the integral factor can be absorbed into K. The exponential factor is $1/n^{(\eta_1/4)(c'-c)^2}$ so choosing c' large enough gives (2.13).

Step 1, part 3: We upper bound the L^1 distance in (2.11) by the sum

$$E_{\theta_0} \int |\frac{w(\theta)p_\theta(\bar{X})}{m(\bar{X})} - \frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})}|d\theta \tag{2.16}$$

$$+ E_{\theta_0} \int |\frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})} - \frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{\theta_o r}(\bar{X})}|d\theta \tag{2.17}$$

$$+ E_{\theta_0} \int |\frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{\theta_0 r}(\bar{X})} - n(\theta; \theta_0, \hat{\theta})|d\theta. \tag{2.18}$$

Step 2, part 1: We use (2.12) and (2.13) to obtain a lower bound for $\chi_U m(\bar{X})$:

$$\begin{aligned} \chi_U m(\bar{X}) &\geq \chi_U (m_r(\bar{X}) - |m(\bar{X}) - m_r(\bar{X})|) \\ &\geq \chi_U K/n^{(k+d)/2}, \end{aligned} \tag{2.19}$$

provided $r \geq d$.

Step 2, part 2: Expression (2.16) equals

$$E_{\theta_0}\chi_{U^c} \int |\frac{w(\theta)p_\theta(\bar{X})}{m(\bar{X})} - \frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})}|d\theta \tag{2.20}$$

$$+ E_{\theta_0}\chi_U \int |\frac{w(\theta)p_\theta(\bar{X})}{m(\bar{X})} - \frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})}|d\theta. \tag{2.21}$$

We show that expression (2.20) tends to zero by the CLT: For n large enough the first term in the sum which gives $q_{\theta r}(\bar{X})$ dominates so that $q_{\theta r}(\bar{X})$ is positive everywhere (see the proof of (2.12)). As a result (2.20) is upper bounded by $E_{\theta_0}\chi_U c(\int w(\theta)p_\theta(\bar{X})/m(\bar{X})d\theta + \int w(\theta)q_{\theta r}(\bar{X})/m_r(\bar{X})d\theta)$ which is less than $2P_{\theta_0}(U^c)$ and so goes to zero. For expression (2.21) we use (2.12) and (2.13) directly as well as the bound $\chi_U \int |p_\theta(\bar{X}) - q_{\theta r}(\bar{X})|w(\theta)d\theta \leq \chi_U K(k_n)^d/n^{(k+r+1)/2}$ derivable from (2.14) and (2.15). By adding and subtracting $q_{\theta r}(\bar{X})/m(\bar{X})$ we have

$$\begin{aligned} \chi_U \int w(\theta)|\frac{p_\theta(\bar{X})}{m(\bar{X})} - \frac{q_{\theta r}(\bar{X})}{m_r(\bar{X})}|d\theta &\leq \chi_U \int w(\theta)\frac{|p_\theta(\bar{X}) - q_{\theta r}(\bar{X})|}{m(\bar{X})}d\theta \\ &\quad + \chi_U \int \frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})}\frac{|m(\bar{X}) - m_r(\bar{X})|}{m(\bar{X})}d\theta \\ &\leq \chi_U K\frac{(k_n)^d n^{(k+d)/2}}{n^{(k+r+1)/2}} \\ &\quad + \chi_U \frac{K(k_n')^d n^{(k+d)/2}}{n^{(k+r+1)/2}}, \end{aligned}$$

so that applying E_{θ_0} to both sides gives an upper bound on (2.21) which tends to zero since $r \geq d$.

Step 3, part 1: Next we show (2.17) tends to zero. We upper bound it by

$$E_{\theta_0}\chi_U \int_{U'} w(\theta)|\frac{q_{\theta r}(\bar{X})}{m_r(\bar{X})} - \frac{q_{\theta\theta_0 r}(\bar{X})}{m_{r\theta_0}(\bar{X})}|d\theta \tag{2.22}$$

$$+ E_{\theta_0}\chi_U \int_{U'^c} \frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})} \tag{2.23}$$

$$+ E_{\theta_0}\chi_U \int_{U'^c} \frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{r\theta_0}(\bar{X})} d\theta \tag{2.24}$$

$$+ E_{\theta_0}\chi_{U^c} \int |\frac{w(\theta)q_{\theta r}(\bar{X})}{m_r(\bar{X})} - \frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{r\theta_0}(\bar{X})}|d\theta. \tag{2.25}$$

Step 3, part 2: Three of the four terms in the last upper bound are easy to control. Term (2.25) tends to zero by the same reasoning as was used for (2.20): the triangle inequality allows us to use 2 as an upper bound for the integral and gives the convergence to zero.

Terms (2.23) and (2.24) are easy also. Note that by reasoning similar to that used to prove (2.12) one can prove

$$m_{r\theta_0}(\bar{X}) \geq K/n^{(k+d)/2}. \tag{2.26}$$

By use of (2.26) and (2.12), to prove that (2.23) and (2.24) go to zero it is enough to show

$$E_{\theta_0}\chi_U \int_{U'^c} q_{\theta r}(\bar{X})d\theta = o(\frac{1}{n^{(k+d)/2}}) \tag{2.27}$$

and

$$E_{\theta_0}\chi_U \int_{U'^c} q_{\theta\theta_0 r}(\bar{X})d\theta = o(\frac{1}{n^{(k+d)/2}}). \tag{2.28}$$

we see that the absolute values of the left hand sides of (2.27) and (2.28) are upper bounded by a sum of r terms of the form of the second term in (2.14). The same technique as was used to control it can again be used. The result is that for $c' - c$ large enough expressions (2.27) and (2.28) can be forced to go to zero at any rate of the form $o(1/n^\alpha)$ for $\alpha > 0$.

Step 3, part 3: Expression (2.22) is the only problematic term left in dealing with (2.17). Our technique will be similar to that used for (2.21).

By adding and subtracting $q_{\theta\theta_0 r}(\bar{X})/m_r(\bar{X})$ and using (2.12) we see that (2.22) is upper bounded by

$$\begin{aligned}
&E_{\theta_0}\chi_U \int_{U'} w(\theta)\frac{|q_{\theta r}(\bar{X}) - q_{\theta\theta_0 r}(\bar{X})||}{m_r(\bar{X})}d\theta \\
&\quad + E_{\theta_0}\chi_U \int_{U'} \frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{r\theta_0}(\bar{X})}\frac{|m_r(\bar{X}) - m_{r\theta_0}(\bar{X})|}{m_r(\bar{X})}d\theta \\
&\leq Kn^{(k+d)/2}[E_{\theta_0}\chi_U \int_{U'} w(\theta)|q_{\theta r}(\bar{X}) - q_{\theta\theta_0 r}(\bar{X})|d\theta \\
&\quad + E_{\theta_0}\chi_U |m_r(\bar{X}) - m_{r\theta_0}(\bar{X})|] \\
&\leq Kn^{(k+d)/2}[2E_{\theta_0}\chi_U \int_{U'} w(\theta)|q_{\theta r}(\bar{X}) - q_{\theta\theta_0 r}(\bar{X})|d\theta \\
&\quad + E_{\theta_0}\chi_U \int_{U'^c} w(\theta)|q_{\theta r}(\bar{X}) - q_{\theta\theta_0 r}(\bar{X})|d\theta]
\end{aligned}$$

we note that the second term in brackets goes to zero by use of (2.27) and (2.28).

For the first term, it is enough to show

$$\chi_U\chi_{U'}|q_{\theta r}(\bar{X}) - q_{\theta\theta_0 r}(\bar{X})| = 0(\chi_U\chi_{U'}\frac{(k_n + k_n')^{3r}}{n^{(k+1)/2}}) \tag{2.29}$$

for then the integration will give a factor of $K(k_n'/\sqrt{n})^d$ so that term will tend to zero also. Since f_i has degree $3(i-1)$ and on the intersection of U_n and U', $||\sqrt{n}(\bar{X}-\mu(\theta))|| \leq k_n+k_n'$ so we have that the left hand side of (2.29) is bounded above by

$$\begin{aligned}
&\frac{K\chi_U\chi_{U'}}{n^{k/2}}\sum_{i=1}^{r}\frac{|f_i(\sqrt{n}(\bar{X}-\mu(\theta)))|}{n^{(i-1)/2}} \\
&\qquad |e^{-(\frac{n}{2})(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta)(\bar{X}-\mu(\theta))} - e^{-(n/2)(\bar{X}-\mu(\theta))\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}| \\
&\leq \frac{K\chi_U\chi_{U'}}{n^{k/2}}(k_n+k_n')^{3(r-1)}(\frac{n}{2})||\bar{X}-\mu(\theta)||^2||\Sigma^{-1}(\theta)-\Sigma^{-1}(\theta_0)||,
\end{aligned}$$

in which we have used the elementary inequality $|e^{-x} - e^{-y}| \leq |x-y|$ to control the difference of exponentials and used norm inequalities on the upper bound resulting from that inequality. Using the restriction to U and U' again we obtain the bound $K\chi_U\chi_U(k_n+k_n')^{3(r-1)}||\Sigma^{-1}(\theta) - \Sigma^{-1}(\theta_0)||/n^{k/2}$. Since all Euclidean norms are equivalent, we can replace the matrix norm with any norm. We choose the norm which sums the absolute values of the entries. Each term in that sum admits a Taylor expansion which can be bounded from above by $(k_n'/\sqrt{n})$ times a positive constant. There are only finitely many constants so taking the maximum gives an upper bound $K(k_n'/\sqrt{n}) \leq K(k_n + k_n')/\sqrt{n}$ which finishes the proof of (2.29).

Step 4, part 1: In this final step we show that (2.18) goes to zero. We start by bounding (2.18) from above by a sum of 5 terms, two of which are easy. Our bound is

$$E_{\theta_0}\chi_U \int_{U'} |\frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{\theta_0 r}(\bar{X})} - \frac{w(\theta)q_{\theta\theta_0 1}(\bar{X})}{m_{\theta_0 1}(\bar{X})}|d\theta \tag{2.30}$$

$$+ E_{\theta_0}\chi_U \int_{U'} |\frac{w(\theta)q_{\theta\theta_0 1}(\bar{X})}{m_{\theta_0 1}(\bar{X})} - n(\theta;\theta_0,\hat{\theta})|d\theta \tag{2.31}$$

$$+ E_{\theta_0}\chi_U \int_{U'^c} \frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{\theta_0,r}(\bar{X})}d\theta \tag{2.32}$$

$$+ E_{\theta_0}\chi_U \int_{U'^c} n(\theta;\theta_0,\hat{\theta})d\theta \tag{2.33}$$

$$+ E_{\theta_0}\chi_{U^c} \int |\frac{w(\theta)q_{\theta\theta_0 r}(\bar{X})}{m_{\theta_0 r}(\bar{X})} - n(\theta;\theta_0,\hat{\theta})|d\theta \tag{2.34}$$

Step 4, part 2: The easy terms are (2.32) and (2.34). The first is the same as (2.24) the other follows from the CLT since $q_{\theta\theta_0 r}(\bar{X})$ is positive for n large enough; this is the same argument as for (2.25) and (2.20).

Step 4, part 3: The next easiest term is (2.33). Since μ is invertible on a neighborhood of θ_0 for any $\eta > 0$ there is an $\varepsilon > 0$ so that

$$|\mu(\theta) - \mu(\theta_0)| < \varepsilon \Rightarrow |\theta - \theta_0| < \eta$$

and also

$$|\mu(\hat{\theta}) - \mu(\theta_0)| < \varepsilon \Rightarrow |\hat{\theta} - \theta_0| < \eta.$$

For such a choice of ε we write (2.33) as

$$E_{\theta_0}\chi_U \int_{|\mu(\theta)-\mu(\theta_0)|>\varepsilon} n(\theta;\theta_0,\hat{\theta})d\theta \tag{2.35}$$

$$+ E_{\theta_0}\chi_U \int_{\varepsilon \geq |\mu(\theta)-\mu(\theta_0)| \geq k_n'\sqrt{n}} n(\theta;\theta_0,\hat{\theta})d\theta \tag{2.36}$$

For (2.36) restriction to U and to the domain of integration gives that $|\theta - \hat{\theta}| \leq 2\eta$ so we can use a Taylor expansion and the triangle inequality to obtain

$$\begin{aligned} ||\nabla\mu(\tilde{\theta})|||\hat{\theta} - \theta| &\geq |\mu(\hat{\theta}) - \mu(\theta)| \\ &\geq (k_n' - k_n)/\sqrt{n} = (c' - c)\sqrt{\frac{\ell n\ n}{n}} \end{aligned}$$

for some $\tilde{\theta}$ lying on the straight line joining θ and $\hat{\theta}$. By the continuity of the derivative we have that

$$(\hat{\theta} - \theta) \geq k(c' - c)\sqrt{\frac{\ell n\ n}{n}}.$$

(Choose, for instance, $K = \sup_{\theta' \in \mathbf{B}(\theta_0, \eta)} ||\nabla \mu(\theta')||$.)

Now (2.36) is bounded by

$$K e^{-(n/4)K(c'-c)^2(\ell n\ n/n)} E_{\theta_0} \chi_U \int e^{-(n/4)(\theta-\hat{\theta})^t J(\theta_0)^t \Sigma^{-1}(\theta_0) J(\theta)(\theta-\hat{\theta})} d\theta, \tag{2.37}$$

which tends to zero.

For (2.35) we use a variant of the last argument. We note that local invertibility implies that given $\varepsilon > 0$ there is an $\eta > 0$ so that

$$|\mu(\theta) - \mu(\theta_0)| > \varepsilon \Rightarrow |\theta - \theta_0| > \eta$$

By restriction to U we have that $\hat{\theta}$ and θ_0 are close so we Taylor expand to get that there is a $K > 0$ so that

$$K|\hat{\theta} - \theta_0| \leq k_n/\sqrt{n}$$

(Choose $K = \inf_{\theta' \in \mathbf{B}(\theta_0, \eta)} ||\nabla \mu(\theta')||$.) Again by the triangle inequality

$$|\hat{\theta} - \theta| \geq \eta - (K k_n/\sqrt{n}) \geq \eta/2.$$

So, in this case we still get a bound much like (2.37). As a result (2.35) goes to zero.

Step 4, part 4: Write expression (2.30) as

$$E_{\theta_0} \chi_U \int_{U'} \frac{w(\theta) q_{\theta\theta_0 1}(\bar{X})}{m_{\theta_0 1}(\bar{X})} \left| 1 - \frac{1 + \sum_{i=2}^{r} f_i(\sqrt{n}(\bar{X} - \mu(\theta)))/n^{(i-1)/2}}{1 + \frac{\sum_{i=2}^{r} \int \frac{f_i(\sqrt{n}(\bar{X}-\mu(\theta)))}{n^{(i-1)/2}} w(\theta) \varphi_{\Sigma(\theta_0)}(\sqrt{n}(\bar{X}-\mu(\theta))) d\theta}{\int w(\theta) \varphi_{\Sigma(\theta_0)} \sqrt{n}((\bar{X}-\mu(\theta))) d\theta}} \right| d\theta$$

It is enough to upper bound the factor in absolute value bars by a function which is $o(1)$ since by enlarging the domain of integration the rest is bounded by 1.

By the restriction to U and U' each f_i is bounded by $K(k_n + k_n')^{3(i-1)}$ which is of lower order than $n^{(i-1)/2}$ and so the summation in the numerator goes to zero.

There are $r-1$ terms in the summation in denominator. The i^{th} one ($i=2,\ldots,r$) can be bounded from above by breaking the integral into two pieces, one over U' the other over U'^c, and using (2.26). Also, on the term with restriction to U and U' we can bound f_i. The result is

$$\frac{K(k_n+k_n')^{3(i-1)}}{n^{(i-1)/2}}\frac{\int_{U'} w(\theta)\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\bar{X}-\mu(\theta)))d\theta}{\int w(\theta)\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\bar{X}-\mu(\theta)))d\theta}$$
$$+\frac{Kn^{(k+d)/2}}{n^{(i-1)/2}}\int_{U'^c} w(\theta)f_i(\sqrt{n}(\bar{X}-\mu(\theta)))\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\bar{X}-\mu(\theta)))d\theta.$$

Obviously the first term tends to zero. The second term also tends to zero by the same technique as used on the right hand term of (2.14). So, (2.30) tends to zero.

Step 4, final part: At last we deal with (2.31). We bound it by adding and subtracting

$$\frac{w(\theta)e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}}{w(\theta_0)(2\pi)^{d/2}|nJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}$$

and

$$\frac{e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}}{(2\pi)^{d/2}|nJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}.$$

Our upper bound on (2.31) is now

$$E_{\theta_0}\chi_U|1-\frac{\int w(\theta)e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}d\theta}{(2\pi)^{d/2}w(\theta_0)|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}| \tag{2.38}$$

$$+E_{\theta_0}\chi_U\int_{U'}|\frac{w(\theta)}{w(\theta_0)}-1|\frac{e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}}{(2\pi)^{d/2}|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}d\theta \tag{2.39}$$

$$+E_{\theta_0}\chi_U\int_{U'}\frac{|e^{-(n/2)(\bar{X}-\mu(\theta))\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}-e^{-(n/2)(\theta-\hat{\theta})^tJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)(\theta-\hat{\theta})}|}{(2\pi)^{d/2}|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}d\theta \tag{2.40}$$

For (2.39) we note that by Taylor expanding on the restricted domain there is a positive definite matrix M so that $(\hat{\theta}-\theta)^tJ^t(\tilde{\theta})\Sigma^{-1}(\theta_0)J(\tilde{\theta})(\hat{\theta}-\theta)\geq(\hat{\theta}-\theta)^tM(\hat{\theta}-\theta)$ As a result (2.39) is bounded from above by

$$K\sup_{\theta\in U'}|\frac{w(\theta)}{w(\theta_0)}-1|E_{\theta_0}\chi_U\int_{U'}n^{d/2}e^{-(n/2)(\hat{\theta}-\theta)^tM(\theta-\hat{\theta})}d\theta$$

in which the integral is finite and by the continuity of w the bound goes to zero.

For expression (2.40) we use techniques similar to those used for (2.29). By the same elementary inequality we obtain the upper bound

$$Kn^{d/2}E_{\theta_0}\chi_U \int_{U'} n||(\theta-\hat{\theta})||^2||J(\tilde{\theta})^t\Sigma^{-1}(\theta_0)J(\tilde{\theta}) - J(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)||d\theta$$

after Taylor expansion of μ, where $\tilde{\theta}$ is on the straight line joint θ and $\hat{\theta}$. By reasoning used in the proof that (2.36) goes to zero we have that $||\sqrt{n}(\theta-\hat{\theta})|| \le K(k_n+k_n')^2$. Also since we have restricted to U and U' the norm of the difference of matrices can be controlled by a Taylor expansion.

Let $f(\psi) = J(\psi)^t\Sigma^{-1}(\theta_0)J(\psi)$, then

$$f(\psi) - f(\theta_0) = (\psi-\theta_0)\nabla f(\tilde{\psi})$$

for some $\tilde{\psi}$ between ψ and θ_0. By continuity of the derivative we have that

$$\sup_{\psi\in B(\theta_0,(k_n+k_n')/\sqrt{n})} ||f(\psi)-f(\theta_0)|| \le \frac{(k_n+k_n')}{\sqrt{n}}K.$$

Using the last two inequalities in the last upper bound gives

$$Kn^{d/2}E_{\theta_0}\chi_U \int_{U'} \frac{(k_n+k_n')^3}{\sqrt{n}}d\theta$$

which goes to zero.

Finally, we control (2.38), by a straight forward Laplace integration. Observe that the integral in (2.38) is the same as

$$\chi_U \int_{U'} w(\theta)e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}d\theta$$
$$+\chi_U \int_{U'^c} w(\theta)e^{-(n/2)(\bar{X}-\mu(\theta))^t\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta_0))}d\theta \qquad (2.41)$$

For a lower bound we drop the second term and Taylor expand in the first to obtain

$$\chi_U(w(\theta_0)-\varepsilon)(\int_{\mathbb{R}^d} - \int_{U'^c})e^{-(1+\varepsilon)(n/2)(\theta-\hat{\theta})^tJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)(\theta-\hat{\theta})}d\theta$$

since $J(\tilde{\theta})^t\Sigma(\theta_0)J(\tilde{\theta}) \ge (1+\varepsilon)J^t(\theta_0)\Sigma(\theta_0)J(\theta_0)$ for $\tilde{\theta}$ between $\hat{\theta}$ and θ_0 given the restriction to U and U'. The integration can be performed and bounded from below by a product of a factor which tends to as ε goes to zero and $w(\theta_0)(2\pi)^{d/2}/|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}$.

For an upper bound of the same form observe that the second term in (2.41) can be bounded above by a function of the form k/n^α where $\alpha > 0$ is an increasing function of $c' - c$, similar to the second term in (2.14). The first term in (2.41) admits an upper bound similar in form to the lower bound just noted. In fact one again obtains a product of $w(\theta_0)(2\pi)^{d/2}/|nJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}$ with a function that goes to one as ε goes to zero.

The result is that there are functions f_1 and f_2 such that $f_1(\varepsilon)$, $f_2(\varepsilon) \to 1$ as $\varepsilon \to 0$, and on U

$$f_1(\varepsilon) \leq \frac{\int w(\theta)e^{-(n/2)(\bar{X}-\mu(\theta))\Sigma^{-1}(\theta_0)(\bar{X}-\mu(\theta))}d\theta}{(w(\theta_0)(2\pi)^{d/2})/|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}} \leq f_2(\varepsilon)$$

Using the last pair of inequalities it is seen that (2.38) tends to zero also.

This finishes the proof. $\square$

We remark that the asymptotic variance is based on $\Sigma(\theta)$ not the Fisher information. This is due to the fact that we are locating at $\hat{\theta}$ which is not the MLE based on the full data. This is consistent with what one expects from uniformizing the local limit theorem which gives $\Sigma(\theta)$ is the variance.

Another remark is that for applications one typically requires the parametric family defined for a parameter space Ω which contains the support of w as a proper subset.

Local limit theorems, and other types of asymptotic results usually give a series of approximations which the errors decrease with the number of terms included increases. The same is true here although we have confined the result to only giving the first term. Higher order correction terms can be deduced from a more careful analysis. Indeed, it term (2.30) which expresses the restriction to one term.

§3 Noncompact parameter spaces and the case $d \neq k$

Our next result gives an extension of Theorem 2.1 to noncompact parameter spaces. Our technique of proof will be to reduce the result to the compact case. Thus we define two mixture, one over a compact set K, the other over its complement. They are

$$M_K(\bar{X}) \int_K \frac{w(\theta)}{W(K)} P_\theta(\bar{X}) d\theta,$$
$$M_{K^c}(\bar{X}) = \int_{K^c} \frac{w(\theta)}{W(K^c)} P_\theta(\bar{X}) d\theta,$$

where W is the prior probability with density w.

Again we use local invertibility of μ at θ_0. Recall that this means there is an open set O containing θ_0 so that the function $\mu|_O : O \to \mu(O)$ is invertible and that for $\theta \in O^c$, $\mu(\theta) \in \mu(O)^c$. Our result for noncompact parameter spaces is the following.

Theorem 3.1: Assume the hypotheses of Theorem 2.1, including $k = d$. In addition assume that there is an $\eta \geq 0$ so that for all θ outside an open set around θ_0 the moment generating function for P_θ is finite on an open neighborhood centered at zero of radius at least η. Then,

$$E_{\theta_0} \int |w(\theta|\bar{X}) - n(\theta; \theta_0, \hat{\theta})| d\theta \to 0. \tag{3.1}$$

Remark: Assuming that the moment generating function is finite is more than we actually require for (3.1) to hold. However, for applications in educational testing, the random variables assume finitely many values so such assumptions are easily satisfied.

Proof: Let K be a compact set, to be specified shortly. We use the normal density restricted to K and denote the normalized restriction by n_K Write $W_K(\theta) = w(\theta)|_K / W(K)$ and observe that (3.1) is

$$E_{\theta_0} \int_K |\frac{w(\theta)p(\bar{X}|\theta)}{\int_K w(\theta)p(\bar{X}|\theta)d\theta + \int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta} - n(\theta; \theta_0, \hat{\theta})| d\theta$$
$$+ E_{\theta_0} \int_{K^c} |w(\theta|\bar{X}) - n(\theta; \theta_0, \hat{\theta})| d\theta$$
$$\leq E_{\theta_0} \int_K |\frac{w_K(\theta)p(\bar{X}|\theta)}{m_K(\bar{X})(1 + \frac{\int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta}{\int_K w(\theta)p(\bar{X}|\theta)d\theta})} - n(\theta; \theta_0, \hat{\theta})| d\theta$$
$$+ E_{\theta_0} \int_{K^c} w(\theta|\bar{X}) + n(\theta; \theta_0, \hat{\theta}) d\theta$$
$$\leq E_{\theta_0} \int_K |\frac{w_K(\theta)p(\bar{X}|\theta)}{n_K(\bar{X})} - n(\theta; \theta_0, \hat{\theta})| d\theta \tag{3.2}$$

$$+ E_{\theta_0} \int_K \frac{w_K(\theta)p(\bar{X}|\theta)}{m_K(\bar{X})} |1 - \frac{1}{1 + \frac{\int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta}{\int_K w(\theta)p(\bar{X}|\theta)d\theta}} |d\theta \tag{3.3}$$

$$+ E_{\theta_0} W(K^c|\bar{X}) + E_{\theta_0} N(K^c; \theta_0, \hat{\theta}). \tag{3.4}$$

By Theorem 2.1, expression (3.2) tends to zero. Also, expression (3.3) equals $E_{\theta_0}|1 - 1/(1 + \int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta / \int_K w(\theta)p(\bar{X}|\theta)d\theta)|$. Since the quantity is absolute value bars is bounded by zero from below and 1 from above, expression (3.3) will tend to zero if we show that

$$\frac{\int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta}{\int_K w(\theta)p(\bar{X}|\theta)d\theta} \xrightarrow{P_{\theta_0}} 0. \tag{3.5}$$

First we show that

$$\exists\, r, r' > 0 P_{\theta_0}(m_{K^c}(\bar{X})e^{nr} \geq p_{\theta_0}(\bar{X})) \leq e^{-nr'}. \tag{3.6}$$

So, we choose the compact set K to be

$$K = \{\theta : |\mu(\theta) - \mu(\theta_0)| \leq \delta\},$$

for some small $\delta > 0$. Intersecting the event in (3.6) with $\{|\bar{X} - \mu(\theta_0)| > \delta/2\}$ and its complement gives an upper bound on the probability in (3.6) as

$$\begin{aligned} &P_{\theta_0}|\bar{X} - \mu(\theta_0)| > \delta/2 \\ &\quad + P_{\theta_0}(|\bar{X} - \mu(\theta_0)| < \delta/2, m_{k^c}(\bar{X})e^{nr} > P_{\theta_0}(\bar{X})) \\ &\leq e^{-nr''} + \sum_{\substack{|\bar{X} - \mu(\theta)| < \delta/2 \\ e^{nr} m_{k^c} > P_{\theta_0}}} P_{\theta_0}(\bar{X}), \end{aligned}$$

for some choice of $r'' > 0$. We apply one of the conditions on the summands and use the fact that for $\theta \in K^c$

$$|\bar{X} - \mu(\theta)| \geq |\mu(\theta_0) - \mu(\theta)| - |\mu(\theta_0) - \bar{X}| \geq \delta/2$$

to give the upper bound

$$\begin{aligned} &e^{-nr''} + e^{nr} \int_{K^c} w_{K^c}(\theta) \sum_{|\bar{X} - \mu(\theta_0)| < \delta/2} P_\theta(\bar{X}) d\theta \\ &\leq e^{-nr''} + e^{nr} \int_{K^c} w_{K^c}(\theta) P_\theta(|\bar{X} - \mu(\theta)| > \delta/2) d\theta. \end{aligned}$$

Since there are k components in $\bar{X}$ and $\mu(\theta)$ which we denote $\bar{X}_{(i)}$ and $\mu_{(i)}(\theta)$ the union of events bound gives

$$e^{-nr''} + e^{nr} \sum_{i=1}^{k} \int_{K^c} w_{K^c}(\theta) P_\theta(|\bar{X}_{(i)} - \mu_{(i)}(\theta)| > \frac{\delta}{2k}) d\theta.$$

Each of the k terms in the summation is exponentially small since

$$\sup_{\theta \in K^c} P_\theta(|\bar{X}_{(i)} - \mu_{(i)}(\theta)| > \frac{\delta}{2k}) \le e^{-nr'}$$

for some $r' > 0$, see Chernoff (1952). As a result (3.6) holds.

Now we can show (3.5). Let $\varepsilon > 0$ and intersect with the event in (3.6) and its complement. We obtain

$$\begin{aligned} &P_{\theta_0}\left(\left(\frac{\int_{K^c} w(\theta)p(\bar{X}|\theta)}{p_{\theta_0}(\bar{X})} d\theta\right)\left(\frac{p_{\theta_0}(\bar{X})}{\int_K w(\theta)p(\bar{X}|\theta)d\theta}\right) > \varepsilon\right) \\ &\le P_{\theta_0}\left(\frac{\int_{K^c} w(\theta)p(\bar{X}|\theta)d\theta}{p_{\theta_0}(\bar{X})} > e^{-nr}\right) \\ &\quad + P_{\theta_0}(p_{\theta_0}(\bar{X}) > \varepsilon e^{nr} m_K(\bar{X})). \end{aligned}$$

We use (3.6) to control the first term. In the second term, intersect again with U and U^c so as to apply (2.19). The resulting upper bound is

$$P_{\theta_0}(U^c) + P_{\theta_0}(p_{\theta_0}(\bar{X}) > ke^{nr}/n^{(k+d)/2}).$$

Since both terms go to zero (3.5) holds and (3.3) tends to zero.

Now the last two terms are easy. The first term in (3.4) is bounded between zero and one, and dominated by the ratio in (3.5) which goes to zero.

For the other term we note it is bounded by

$$\begin{aligned} &KE_{\theta_0}\chi_{\{|\mu(\hat{\theta})-\mu(\theta_0)|<\delta/2\}} \int_{K^c} n^{d/2} e^{-n(\theta-\hat{\theta})^t J_\mu^t(\theta_0)\Sigma_{(\theta_0)}^{-1} J_\mu(\theta_0)(\theta-\hat{\theta})} d\theta \\ &\quad + KE_{\theta_0}\chi_{\{|\mu(\hat{\theta})-\mu(\theta_0)|>d/2\}}. \end{aligned}$$

The second term goes to zero by consistency of $\bar{X}$ for $\mu(\theta_0)$. The first term is the same as (2.35) and so goes to zero also. $\square$

In the statement and proof of the theorem we have allowed k and d to be different. One can observe that if $d > k$ then the resulting normal will have a singular variance matrix. On the other hand if $d < k$ then μ^{-1} does not make sense and consequently neither does $\hat{\theta}$ as it was defined.

The desired result can be proved by centering at the estimator

$$\hat{\theta} = \arg\min_{\theta'} ||\overline{X} - \mu(\theta')||_{\Sigma(\theta_0)}, \tag{3.7}$$

which reduces to $\mu^{-1}(\overline{X})$ when $d = k$. Our final result is the following.

Theorem 3.2: Assume the hypothesis of Theorem 2.1 and Theorem 3.1. Then for $k > d$ there exists r so large that (3.1) continues to hold for the estimator $\hat{\theta}$ in (3.7).

Proof: We indicate how to modify the proof in the compact case for $d = 1$ and general $k > 1$. Extension to larger values of d and noncompactness follows straightforwardly.

Step 1, part 1: Note that by adding and subtracting $\mu(\hat{\theta})$ in the exponent we obtain

$$m_r(\overline{X}) \geq \frac{k}{n^{k/2}} \int e^{-(n/2)||\overline{X}-\mu(\hat{\theta})||_\theta^2 - (n/2)||\mu(\hat{\theta})-\mu(\theta)||_\theta^2 - (\frac{n}{2})(\overline{X}-\mu(\hat{\theta}))\Sigma^{-1}(\theta)(\mu(\hat{\theta})-\mu(\theta))} d\theta \tag{3.8}$$

where $||\cdot||_\theta$ indicates the inner product wrt $\Sigma(\theta)^{-1}$. On U we have that $||\overline{X} - \mu(\hat{\theta})||_{\theta_0} \leq k_n/\sqrt{n}$ and if we use the implicit function theorem we can assert the existence of a solution h to the equation

$$L(\hat{\theta}) = \Sigma(\overline{X}_i - \mu_i(\hat{\theta}))\mu_j(\hat{\theta})\sigma^{ij}(\theta_0) = 0$$

where $\hat{\theta} = h(\overline{X})$, $\theta_0 = h(\mu(\theta_0))$ and $\sigma^{ij}(\theta_0)$ are the entries of $\Sigma^{-1}(\theta_0)$. As a result $|\hat{\theta} - \theta_0| \leq K||\overline{X} - \mu(\theta_0)||_{\theta_0} \leq K\sqrt{\frac{\ell n\ n}{n}}$. If we cut the domain of integration down to $|\theta - \hat{\theta}| \leq \frac{k_n'}{\sqrt{n}}$ then by the triangle inequality $|\theta - \theta_0| \leq K\sqrt{\ell n\ n/n}$. By Taylor expanding we then obtain that

$$\Sigma^{-1}(\theta) \cong (1 + \varepsilon_n)\Sigma^{-1}_{(\theta_0)} \tag{3.9}$$

where $\varepsilon_n = O\left(\sqrt{\frac{\ell n\ n}{n}}\right)$ and $\cong$ means the LHS is bounded above and below by expressions of the form of the RHS.

Next we note that the third term in the exponent of (3.8) is negligible compared to the other two, at least when restricted to U: From (3.9) it is enough to examine

$n(\overline{X}-\mu(\hat{\theta}))\Sigma(\theta_0)^{-1}(\mu(\hat{\theta})-\mu(\theta))$. Taylor expanding μ at $\hat{\theta}$ and using $L(\hat{\theta})=0$ gives that the third term is

$$K\ n(\theta-\hat{\theta})^2||\overline{X}-\mu(\hat{\theta})||_{\theta_0}$$

which is seen to be of order $O(\ell n\ n/\sqrt{n})$ for some $\alpha>0$. As a result we have on U that

$$m_r(\overline{X}) \geq \frac{Ke^{-(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{n^{(k+d)/2}}.$$

Step 1, parts 2 and 3 are unchanged.

Step 2 part 1: We use the modified bound of Step 1, part 1 to obtain

$$\chi_U m(\overline{X}) \geq \frac{K\chi_U}{n^{(k+d)/2}}e^{-(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}\left(1-\frac{(k_n')^d e^{(n/2)(1-\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{n^{(r+1-d)/2}}\right)$$

Since $n||\overline{X}-\mu(\hat{\theta})||_{\theta_0}\leq c^2\ell n\ n$, r can be chosen large enough to ensure the second term in parentheses goes to zero.

Step 2, part 2: Expression (2.20) is no problem and it is seen that (2.21) goes to zero by noting that

$$\chi_U\int\frac{w(\theta)|P_\theta(\overline{X})-q_{\theta r}(\overline{X})|}{m(\overline{X})}d\theta+\chi_U\int\frac{w(\theta)q_{\theta r}(\overline{X})}{m_r(\overline{X})}\frac{|m(\overline{X})-m_r(\overline{X})|}{m(\overline{X})}d\theta$$
$$\leq K\chi_U\Big(\frac{k_n^d}{n^{(k+r+d)/2}}\cdot\frac{n^{(k+d)/2}}{e^{-(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}+\frac{k_n'^d}{n^{(k+r+1)}}\frac{n^{(k+d)/2}}{e^{-(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}\Big)$$

which goes to zero for r large enough.

Step 3, part 1 remains unchanged.

Step 3, part 2: Showing that analogs of (2.23) and (2.24) go to zero can be readily done. It is enough to show that

$$E_{\theta_0}\chi_n\int_{U'}q_{\theta r}(\overline{X})\frac{n^{(k+d)/2}}{e^{-(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}d\theta\longrightarrow 0 \qquad (3.10a)$$

$$E_{\theta_0}\chi_U\int_{U'}q_{0\theta_0 r}(\overline{X})\frac{n^{(k+d)/2}}{e^{-(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}d\theta\longrightarrow 0 \qquad (3.10b)$$

Since the analog to (2.26), $m_{r\theta_0}(\overline{X})\geq\frac{ke^{-(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{n^{(k+d)/2}}$ can be derived by the same technique as in the modified step 1, part 1.

Now, for both cases it is enough to note that on U $n||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0} \le c^2 \ell n\ n$, and one obtains from the other part of either of the integrands bounds of the form $n^{-k(c'-c)^2}$. It is enough to choose $c'-c$ large enough.

Step 3, part 3: It is enough to show

$$Kn^{(k+d)/2}(2E_{\theta_0}\chi_U \int_{U'} w(\theta)|q_{\theta r}(\overline{X})q_{\theta\theta_0 r}(\overline{X})|d\theta e^{(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}$$
$$+ E_{\theta_0}\chi_U \int_{U'^c} w(\theta)|q_{\theta r}(\overline{X}) - q_{\theta\theta_0 r}(\overline{X})|d\theta e^{(n/2)(1+\varepsilon_n)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}) \qquad (3.11)$$

goes to zero. By the reasoning in part 2, (3.10a) and (3.10b) can be used to control the 2nd term in (3.11). For the first term we observe that the extra exponential factor is bounded above by $e^{(1+\varepsilon_n)c^2\ell n\ n} \le n^{(1+\varepsilon_n)c^2} \le n^{3c^2/2} \le n^{1/4}$, for n large enough and c small enough. The proof of this part in §2 gave a bound of the form $k(\ell n\ n)^{3r}/\sqrt{n}$ and even for r arbitrarily large the extra $n^{1/4}$ does not alter the convergence to zero.

Step 4, part 1 is unchanged.

Step 4, part 2: Use the result from the modified version of step 3, part 2.

Step 4, part 3 is unchanged.

Step 4, part 4: It is enough to show that for $r \ge c \ge 2$

$$\frac{K(k_n+k'_n)^{3(i-1)}}{n^{(i-1)/2}}\chi_U \frac{\int w(\theta)\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\overline{X}-\mu(\theta)))d\theta}{\int w(\theta)\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\overline{X}-\mu(\theta)))d\theta}$$
$$+\frac{Kn^{(k+d)/2}}{n^{(i-1)/2}}\chi_U \int_{U'^c} w(\theta)f_i(\sqrt{n}(\overline{X}-\mu(\theta)))\varphi_{\Sigma(\theta_0)}(\sqrt{n}(\overline{X}-\mu(\theta)))d\theta \cdot e^{(n/2)||\overline{X}-\mu(\hat{\theta})||_{\theta_0}}$$

goes to zero. This is obvious for the first term. Since $n||\overline{X}-\mu(\hat{\theta})||^2 \le c^2\ell n\ n$ the second term can be controlled by choosing c' large enough as in the remarks after expression (2.14).

Step 4, final part: To control the analog of (2.31) we add and subtract

$$\frac{w(\theta)e^{-(n/2)||\overline{X}-\mu(\theta)||^2_{\theta_0}}e^{(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{w(\theta_0)(2\pi)^{d/2}|nJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}$$

and

$$\frac{e^{-(n/2)||\overline{X}-\mu(\theta)||^2_{\theta_0}}e^{(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{(2\pi)^{d/2}|nJ(\theta_0)^t\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}$$

so that we must control

$$E_{\theta_0}\chi_U|1-\frac{e^{(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}\int w(\theta)e^{-(n/2)||\overline{X}-\mu(\theta)||^2_{\theta_0}}d\theta}{(2\pi)^{d/2}w(\theta_0)|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}} \tag{3.12}$$

$$+E_{\theta_0}\chi_U\int_{U'}|\frac{w(\theta)}{w(\theta_0)}-1|\frac{e^{-(n/2)||\overline{X}-\mu(\theta)||^2_{\theta_0}}e^{(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}}{(2\pi)^{d/2}|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}d\theta \tag{3.13}$$

$$+E_{\theta_0}\chi_U\int_{U'}\frac{|e^{(n/2)||\overline{X}-\mu(\hat{\theta})||^2_{\theta_0}}e^{-(n/2)||\overline{X}-\mu(\theta)||^2_{\theta_0}}-e^{-(n/2)(\theta-\hat{\theta})^tJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)(\theta-\hat{\theta})}|}{(2\pi)^{d/2}|nJ^t(\theta_0)\Sigma^{-1}(\theta_0)J(\theta_0)|^{-1/2}}d\theta, \tag{3.14}$$

the analogs of (2.38), (2.39) and (2.40). For (3.13) we use the fact that

$$n||\overline{X}-\mu(\theta)||^2_{\theta_0}=n||\overline{X}-\mu(\hat{\theta})||^2+n||\mu(\hat{\theta})-\mu(\theta)||^2_{\theta_0}+O(\frac{(\ell n\ n)^\alpha}{\sqrt{n}}) \tag{3.15}$$

so as to obtain the upper bound

$$o(1)E_{\theta_0}\chi_U\int_{U'}n^{d/2}e^{-(n/2)||\mu(\hat{\theta})-\mu(\theta)||^2_{\theta_0}}$$

which goes to zero since the integral gives a constant. (The modulus of continuity gives the $o(1)$.) For (3.14) we use (3.15) so as to reduce it to the analog of (2.29), as in step 4 final part in §2. (Choose $c>0$ small enough.)

For the last term (3.12), Laplace integration gives the desired convergence to zero, by use of (3.15) again.

REFERENCES

Ackerman, T. A. (1991). An examination of the effect of multidimensionality on parallel forms construction. Presented at annual meeting of the National Council on Measurement in Education, Chicago.

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions.* Krieger, Malabar.

Bickel, P. and Yahav, J. (1969). Some contributions to the asymptotic theory of Bayes' solutions. *Z. Wahrsch. verw. Gebeite* **11** 257-276.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, **23**, 493-507.

Doksum, K. A. and Lo, A. Y. (1990). Consistent and robust Bayes procedures for location based on partial information. *Ann. Math. Statist.*, **18**, 443-453.

Holland, P. W. and Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.*, **14**, 1523–1543.

Junker, B. W. (1991). Conditional association, essential independence, and monotone unidimensional item response models. Private communication.

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

Le Cam, L. (1958). Les proprietes asymptotiques des solution de Bayes. *Publ. Inst. Statist. Univ. Paris*, **7**, 18–35.

Walker, A. M. (1969). On the asymptotic behavior of posterior distributions. *J. Roy. Statist. Soc., Ser. B*, **31**, 80–88.