POSTERIOR NORMALITY GIVEN THE MEAN WITH APPLICATIONS

by

B. Clarke
Purdue University

and

J. K. Ghosh
Indian Statistical Institute
and Purdue University

Technical Report #91-68

Department of Statistics
Purdue University

December 1991

Posterior Normality Given the Mean with Applications

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 $w(\theta|\bar{X})$, the posterior density of a parameter given the mean of the data rather than the full data set. Here we show that $w(\theta|\bar{X})$ is asymptotically normal in an L^1 sense, and we identify the asymptotic variance as a function of the variances of the random variables, modified by the gradient of the mean. The location of the limiting normal is also identified. The main results are proved assuming that $X_1, \ldots, X_n, \ldots$ are independent, but not necessarily identically distributed. Our result may be used to construct approximate HPD sets for the parameter which is of use in the statistical theory of standardized tests. It may also be used to show that the covariance between two test items conditioned on the mean is asymptotically nonpositive and, consequently, to base a hypothesis test of independence on a quantity which does not explicitly use the parametric family. Finally, our result leads to an explanation of why the size distribution of sand grains on a riverbed may tend to normality if sampled sufficiently far from the source.

§1 Introduction

Suppose we want to estimate a finite dimensional parameter $(\theta_1, \ldots, \theta_d)$, distributed according to the density $w(\theta_1, \ldots, \theta_d)$, with respect to Lebesgue measure, where w is continuous and bounded. We have data in the form of $X_1, ..., X_n$ where each X_i takes values in a k-dimensional minimal lattice L assumed to be regular with common step length ℓ. Now the joint density is

$$w(\theta_1^d) p_{\theta_1^d}(X_1), \ldots, p_{\theta_1^d}(X_n),$$

where $p_{\theta_1^d}(X_i)$ is the probability mass function for X_i. For brevity we write $X^n = (X_1, \ldots, X_n)$, and denote the parameter space by $\Omega \subset \mathbb{R}^d$.

Given that n is sufficiently large it is of interest to look at highest posterior density (HPD) regions of $w(\theta_1^d | \bar{X})$, the posterior density given a summary statistic such as the mean $\bar{X}$. 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 modeling standpoint since $\bar{X}$ often arises naturally. For instance, in educational testing problems it is the student's score on a multiple choice test.

It has already been proved that $w(\theta_1^d | \bar{X})$ is asymptotically normal when the X_i's are identically and independently distributed (IID), see Clarke and Ghosh (1991), hereafter referred to as CG. Even when they are not identical, if $\bar{X}$ is sufficient then $w(\theta_1 | \bar{X}) = w(\theta_1 | X^n)$ by the factorization criterion so conventional asymptotic normality results in the independent not identical (INID) case apply. However, there are examples where the X_i's are not identical and $\bar{X}$ is not sufficient. One is the Rasch model, see Lindsay, Clogg and Grego (1991). We still expect $w(\theta | \bar{X})$ to be asymptotically normal even though existing results do not apply. In addition, it is expected that the asymptotic variance will not be the Fisher information since $\bar{X}$ is not the MLE in general. The discrepancy between standard error of the limiting normal of $w(\theta_1 | \bar{X})$ and the Fisher information tracks the degree to which $\bar{X}$ fails to be sufficient.

One methodological implication of our results is the following. Denote the mean and variance of any X_i by the k-dimensional vector $\mu_i(\theta_1^d)$ and the $k \times k$ matrix $\Sigma_i(\theta_1)$

respectively. So, the mean of $\bar{X}$ is the mean of the μ_i's, $\bar{\mu}^n$, and its variance is the sum of the Σ_i's, $n\bar{\Sigma}$. We show that the asymptotic variance is the inverse of $nJ_{n,\mu}(\theta_0)^t\bar{\Sigma}^{-1}(\theta_0)J_{n,\mu}(\theta_0)$ where $J_{n,\mu}(\theta_0)$ is the $k \times d$ Jacobian matrix of $\bar{\mu}_n(\theta)$, regarded as a function from $\Omega \rightarrow \mathbb{R}^d$. Consequently, it is seen that the resulting normal is nondegenerate only when $d \leq k$.

In addition to applications in estimation, asymptotic normality of $w(\theta|\bar{X})$ has implications for testing the independence of test items. In Junker (1991), a heuristic argument suggested that a hypothesis test of the independence of test items i and j could be based the behavior of

$$COV(X_i, X_j|\bar{X}), \tag{1.1}$$

provided it is nonpositive. We give conditions under which expression (1.1) is asymptotically nonpositive for independent random variables. Note that expression (1.1) is a manifest quantity – that is, it can be calculated from the data without reference to the underlying parametric family. This supports Junker's program of characterizing the desired properties of standardized tests, namely unidimensionality and local asymptotic discrimination, in terms of manifest quantities. As a result, it is possible to test whether the desired properties are satisfied directly from the data.

Another application of the result is for the evaluation of the influence of dimensionality on test item bias. This was used in Ackerman (1991).

It is a curiosity that, as outlined in Section 4, our results below can also be used to explain why the size distribution of sand grains on a riverbed tends to normality when one samples sufficiently far from the source. Due to sorting through transportation, the observed size distribution downstream will show concentration around a typical grain size, say $\hat{\theta}$, and a quadratic, i.e. normal approximation will be available. Data and more details may be found in Ghosh (1988) and Ghosh, Mazumder, and Sengupta (1991).

The structure of the paper is as follows. In Section 2 we state and prove our results for the case of INID lattice valued random variables. There are three results: The first guarantees that a local limit theorem is uniformly good over compact sets in the parameter space; the second gives the desired result when the parameter space is compact; the third gives an extension to noncompact parameter spaces. In Section 3 we explain the relevance of (1.1) to educational testing and then use the results of Section 2 to show that (1.1) is asymptotically nonpositive. This requires a technical lemma which is used to show that $E(X_i|\theta)$is a good approximation for $E(X_i|\bar{X},\theta)$. The result then follows by using a standard identity and examining posterior covariances. In Section 4 we briefly explain the application to grain size distribution.

§2 Asymptotic normality of the posterior given the mean

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 prove a local limit theorem so as to approximate $p_\theta(\bar{X})$ by a normal density uniformly in θ. This implies $m(\bar{X})$ is well approximated by a mixture of normals. Assessing convergence under a fixed member of the parametric family, p_{θ_0}, forces the posterior to concentrate on a shrinking neighborhood of θ_0, allowing identification of the asymptotic variance. The proof that $w(\theta|\bar{X})$ is asymptotically normal follows by obtaining an upper bound on the L^1 distance between $w(\theta|\bar{X})$ and the target normal denoted $n(\theta; \theta_0, \hat{\theta})$. This upper bound has three terms which tend to zero. They come from adding and subtracting two densities each of which is a step toward the desired normal.

To introduce our approximations we require some notation. We denote the sum of the first n outcomes by $S_n(X) = \sum_{i=1}^{n} X_j$, with mean $\mu^n(\theta) = E_\theta S^n(X) = \sum_{j=1}^{n} \mu_j(\theta)$, where $\mu_j(\theta) = E_\theta X_j$. Analogously, we write $\Sigma^n(\theta) = \sum_{j=1}^{n} \Sigma_j(\theta)$ where $\Sigma_j(\theta) = \mathrm{Var}_\theta X_j$. The average mean is $\bar{\mu}^n(\theta) = (1/n)\mu^n(\theta)$; the average variance is $\bar{\Sigma} = \bar{\Sigma}^n(\theta) = (1/n)\Sigma^n(\theta)$. We write $\bar{J}_{\mu,n}(\theta) = \nabla\bar{\mu}_n(\theta)$. Outcomes of random variables are denoted by the appropriate lower case letter. Where no confusion will result we omit subscripts, for instance dependence on n. To define the location of the limiting normal we require the following.

Definition 2.1: A sequence of functions $< f_n(\theta) > |_{n=1}^{\infty}$ is locally invertible at θ_0 if and only if there is a neighborhood N_{θ_0} of θ_0 so that for all n

$$f_n|_{N_{\theta_0}} : \;\; N_{\theta_0} \longrightarrow f_n(N_{\theta_0})$$

is invertible, for $\theta \in N_{\theta_0}^c$ we have that $f_n(\theta) \in f_n(N_{\theta_0})^c$ and the set $\bigcap_{n=1} f_n(N_{\theta_0})$ contains an open set around $\lim_{n\to\infty} f_n(\theta_0)$, assumed to exist.

Now, the target normal is

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

where $\hat{\theta} = (\bar{\mu}^n)^{-1}(\bar{X})$ near θ_0 since $\bar{\mu}^n$ is assumed to be locally invertible at θ_0, and $|\cdot|$ denotes the determinant. Note that the variance in (2.3) is no longer $\bar{\Sigma}(\theta_0)$, but a modification depending on the parametrization. In places where the slope of $\bar{\mu}^n$ changes rapidly as a function of the true value the variance increases, where $\bar{\mu}^n$ is relatively constant the variance in fact decreases. Note also that $(\bar{\mu}^n)^{-1}$ is well defined only when k and d are equal. We allow k to be formally different from d so that their roles can be distinguished. In fact, a result is possible in the case that $k > d$ provided that $\hat{\theta}$ is redefined as in CG.

We remark that the limiting normal can be expressed (under further hypotheses) in terms of an asymptotic variance $J_\mu(\theta)^t\Sigma^{-1}(\theta)J_\mu(\theta)$ where $J_n(\theta) = \lim_{n\to\infty} \bar{J}_{\mu,n}(\theta)$ and $\Sigma(\theta) = \lim_{n\to\infty} \bar{\Sigma}(\theta)$, provided the limits are well defined. However, this form is not as useful for applications.

To obtain the limiting form (2.3) we use two normal approximations. 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 between $p_\theta(\bar{X})$ and its normal approximation of r terms tends to zero in supremum norm 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

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

for the r term approximation to $p_\theta(\bar{X})$, where f_i is a polynomial of degree $3r$ in k variables and $\varphi_{\bar{\Sigma}(\theta)}$ is the normal density with mean 0 and variance $\bar{\Sigma}(\theta)$. A variant on (2.4) is

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

in which the variance matrix is evaluated at θ_0. Mixtures of the densities in (2.4) and (2.5) 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.

Our first result is an INID analog of Theorem 22.1 of BR. To obtain it, we use characteristic function (CF) arguments. The CF of X_j is

$$f_j(\theta, t) = E_\theta e^{i(t, X_j)}.$$

Since the X_j's take values in a common lattice their CF's have a common fundamental domain which we denote by $\mathcal{F}^*$. Central to the statement and proof of the result is a proper subset E_1 of $\mathcal{F}^*$,

$$E_1 = \{t \in \mathbb{R}^k : ||t|| \le C\},$$

where C is a constant. Let K be a compact set in the parameter space. We require that C satisfies the following.

Assumption A: (i) On $\sqrt{n}E_1$ we can use the expansion given in Theorem 9.9 of BR modified in the same way as Theorem 9.12 of BR. (ii) For $t \in \sqrt{n}E_1$ we have that

$$\sup_{j=1}^{n} |f_j(\theta, \frac{\bar{\Sigma}^n(\theta)^{-1/2} t}{\sqrt{n}}) - 1| \le 1/2.$$

(iii) For $\delta(\theta) = \sup_{j \in \mathbb{N}} \sup_{t \in \mathcal{F}^* - E_1(\theta)} |f_j(\theta, t)|$ we have that

$$\delta_K = \sup_{\theta \in K} \delta(\theta) < 1.$$

Assumption B: For $r \ge 1$ suppose that on K

$$g(\theta) = \sup_n \frac{1}{n} \sum_{j=1}^{n} E_\theta ||X_j||^{r+2}$$

exists and is bounded.

Assumption C: We have that on K

$$\eta_1 Id \leq \bar{\Sigma}(\theta) \leq \eta_2 Id$$

for some $\eta_1, \eta_2 > 0$ and all n, where Id is the d by d identity matrix.

Our result is the following.

Proposition 2.1: Under Assumptions A,B, and C we have that

$$\sup_{\theta\in k}\sup_{\alpha\in L}\left(1+||\frac{\alpha-\bar{\mu}^n(\theta)}{\sqrt{n}}||^{r+1}\right)|p_\theta(\alpha/n)-q_{\theta r}(\alpha/n)| \\ = O(\frac{1}{n^{(k+r)/2}}). \tag{2.8}$$

Remark: The above assumptions A, B, and C hold if the f_j's are jointly continuous in (t,θ), uniformly in j. This is the case in the Rasch model and in the generalisation of that model considered by Tsutakawa and Johnson (1990). More generally, suppose X_i is distributed according to a probability function $p(x_i,\theta,\alpha_i)$ where the dependence on i is only in the third argument. Then, assumptions A, B, and C hold if (i) $p(x_i,\theta,\alpha)$ is a continuous function of (θ,α), which ranges over a fixed compact set; (ii) the moments $E_{(\theta,\alpha)}|X|^{r+2}$ are continuous and finite for (θ,α) in the compact set; and (iii) for some positive constants η_1 and η_2, the variance matrix $\Sigma(\theta,\alpha)$ is continuous and satisfies $\eta_1 Id \leq \Sigma(\theta,\alpha) \leq \eta_2 Id$ on the compact set.

Proof: First we show that

$$\sup_{\theta\in k}\sup_{\alpha\in L}\left(1+||\frac{\alpha-\bar{\mu}^n(\theta)}{\sqrt{n}}||^{r+1}\right)|p_\theta(\alpha/n)-q_{\theta r}(\alpha/n)| \\ = o(\frac{1}{n^{(k+r-1)/2}}). \tag{2.9}$$

The CF of $S^n(X)$ is

$$\tilde{f}^n(\theta,t)=\prod_{j=1}^{n}f_j(\theta,t)$$

and the CF of $Y_n=(S^n(X)-\mu^n(\theta))/\sqrt{n}$ is

$$f^n(\theta,t)=\tilde{f}^n(\theta,t/\sqrt{n})e^{-i(t/\sqrt{n},\mu^n(\theta))}.$$

By the inversion formula we have

$$P(S^n(X)=\xi)=\frac{\ell}{(2\pi)^k}\int_{\mathcal{F}^*}\tilde{f}(\theta,t)e^{-i(t,\xi)}dt.$$

Using $t=t'/\sqrt{n}$ we obtain

$$\begin{aligned}P(S^n(X)=\xi)&=\frac{\ell}{(2\pi)^k}\frac{1}{n^{k/2}}\int_{\sqrt{n}\mathcal{F}^*}\tilde{f}(\theta,t/\sqrt{n})e^{-i(t/\sqrt{n},\xi)}dt\\&=\frac{\ell}{(2\pi)^k}\frac{1}{n^{k/2}}\int_{\sqrt{n}\mathcal{F}^*}f(\theta,t)e^{-i(t,\frac{\xi-\mu^n(\theta)}{\sqrt{n}})}dt\end{aligned}$$

from which we see for $Y_n=(1/\sqrt{n})(S^n(X)-n\bar{\mu}^n)$ that

$$P(Y_n=y_{n\xi})=\frac{1}{(2\pi)^k}\frac{1}{n^{k/2}}\int_{\sqrt{n}\mathcal{F}^*}f^n(\theta,t)e^{-(t,y_{n\xi})}dt.$$

By differentiation we obtain

$$Y_{n\xi}^{\beta}P_\theta(Y_{n\xi})=\frac{\ell(-i)^{|\beta|}}{(2\pi)^k n^{k/2}}\int_{\sqrt{n}\mathcal{F}^*}[D^\beta f^n(\theta,t)]e^{-i(t,Y_{n\xi})}dt$$

for vectors $\beta=(\beta_1,\ldots,\beta_k)$ where $\beta_i\geq 0$ are integers summing to $|\beta|\leq r+2$ and D^β denotes the differentiation operator $(D_{t_1})^{\beta_1},\ldots,(D_{t_k})^{\beta_k}$. Vectors raised to powers β mean that each entry in the vector is raised to the corresponding entry in β.

We have a similar result for the Fourier transform of $q_{\theta r}(\bar{X})$, which we denote $\tilde{q}_{\theta r}(t)$, namely,

$$Y_{n\xi}^{\beta}q_{\theta r}(Y_{n\xi})=\frac{\ell(-i)^{|\beta|}}{(2\pi)^k n^{k/2}}\int_{\mathbb{R}^k}[D^\beta\tilde{q}_{\theta r}(t)]e^{-i(t,Y_{n\xi})}dt,$$

where

$$\tilde{q}_{\theta r}(t)=\sum_{j=1}^{r}n^{-(j-1)/2}\tilde{P}_j(it:\{\chi_\nu\})e^{-||t||^2/2}.$$

The $\tilde{P}_j(it:\{\chi_\nu\})$'s are polynomials with coefficients depending on cumulants χ_ν.

Now we have the upper bound

$$\begin{aligned}&|Y_{n\xi}^{\beta}(p_\theta(Y_{n\xi})-q_{\theta r}(Y_{n\xi}))|\\&\quad\leq\frac{K}{n^{k/2}}\left[\int_{\sqrt{n}E_1}|D^\beta(f^n(\theta,t)-\tilde{q}_{\theta r}(t))|dt\right.\\&\qquad+\int_{\sqrt{n}\mathcal{F}^*-\sqrt{n}E_1}|D^\beta f^n(\theta,t)|dt\\&\qquad\left.+\int_{\mathbb{R}^k-\sqrt{n}E_1}|D^\beta\tilde{q}_{\theta r}(t)|dt\right].\end{aligned}\tag{2.10}$$

Since the domain of integration excludes a ball with radius increasing as $\sqrt{n}$, the presence of the exponential factor implies that the last integral tends to zero at rate $O(e^{-nr'})$ for some $r' \geq 0$. The middle integral tends to zero at an exponential rate also: After differentiating $f^n(\theta, t)$ and observing that the exponential factor has norm 1 one can transform back to $\mathcal{F}^* - E_1$. The product $\tilde{f}(\theta, t)$ can be bounded from above by $O(\delta_K^n)$, in which $\delta_K < 1$.

The first integral in (2.10) requires Theorem 9.12 in BR, which is based on Theorems 9.9 and 9.10, also in BR. Examination of the proofs of those theorems shows that our assumptions give an upper bound for the integral of order $o(1/n^{r/2})$ uniformly in θ. Now (2.10) gives (2.8) by the same triangle inequality argument as was used in the proof of Proposition 2.1 in CG. $\square$

We remark that we can dispense with Assumption A(i) by making use of the other assumptions with Theorem 9.11 and Lemma 14.3 in BR. The first integral in (2.10)is broken up into two parts, say I_{11} and I_{12}, the range of integration for I_{11} being the set $n^{r/2(r+2)}E_1$. Using the moment assumptions it can be seen that A(i) holds by use of the expansions given in Theorem 9.11 of BR (modified as in Theorem 9.12 BR). These facts imply $I_{11} = o(n^{-r/2})$. To bound one part of the integrand in I_{12} we can use the estimate given in Lemma 14.3 of BR on the set $\sqrt{n}E_1 - n^{r/2(r+2)}E_1$. For the part of the integrand involving a normal density as a factor we can use a direct argument similar to that used for the third term in (2.10)).

To set up the statement and proof of our next result requires more notation. We define shrinking neighborhoods in the sample space and in the parameter by

$$U_n = \{X^n : \ |\bar{\mu}_n(\hat{\theta}) - \bar{\mu}_n(\theta_0)| \leq \frac{k_n}{\sqrt{n}}\}$$
$$U'_n = \{\theta : \ |\bar{\mu}_n(\theta) - \bar{\mu}_n(\theta_0)| \leq \frac{k'_n}{\sqrt{n}}\}$$

where $k_n/\sqrt{n}$, $k'_n/\sqrt{n} \to 0$ and $||\cdot||$ is a norm on L. Both sets are dependent on θ_0 although this has been suppressed in the notation. To work with these sets we will require that the error terms in Taylor expansions are small uniformly in n. We use the following definition.

Definition 2.2: A sequence of functions $< g_n(\theta) > |_{n=1}^{\infty}$ is uniformly Taylor expandable at θ_0 if and only if (1): each g_n is continuously differentiable on an open set N_{θ_0} containing

θ_0,(2): there are α, $\beta > 0$ so that for all n and all $\theta \in N_{\theta_0}$

$$\beta > ||\nabla g_n(\theta)|| > \alpha;$$

and (3): on N_{θ_0}, ∇g_n has maximal rank.

The defining conditions in U_n and U'_n can be equivalently expressed as $||\hat{\theta} - \theta_0|| \leq k_n/\alpha\sqrt{n}$ and $||\theta - \theta_0|| \leq k'_n/\alpha\sqrt{n}$ where $\alpha = \inf ||\nabla \mu_n(\theta')||$ and the infimum is over θ' in a ball of radius ε centered at θ_0. The rates of shrinkage of the neighborhood that are seen to be most useful are $k_n = c\sqrt{\ell n\ n}$ and $k'_n = c'\sqrt{\ell n\ n}$, where c', $c > 0$ and $c' - c > 0$. Obtaining the desired convergence requires choosing $c' - c$ large enough.

Our next result makes use of the properties of the summands in (2.4), particularly the polynomial form of the functions $f_{i,\theta}$ and the fact that the coefficients are continuous functions of θ. Without further remark we use results from BR §7 and §9. Now, for compact parameter spaces proving the desired result will only require examination of the steps in the proof of Theorem 2.1 in CG.

Theorem 2.1: Assume the hypotheses of Proposition 2.1 are satisfied and that w is positive at θ_0. Assume also that

$$< \bar{\mu}_n(\theta) > |_{n=1}^{\infty}, \quad < \bar{\Sigma}^{-1}(\theta) > |_{n=1}^{\infty}, \quad < \bar{J}^t_{\mu,n}(\theta)\Sigma^{-1}(\theta_0)\bar{J}_{\mu,n}(\theta) >$$

are uniformly Taylor expandable and that $< \bar{\mu}_n(\theta) > |_{n=1}^{\infty}$ is locally invertible at θ_0. Finally, suppose there is a neighborhood N_{θ_0} of θ_0 and $\alpha, \beta > 0$ so that for θ, $\theta' \in N_{\theta_0}$ we have that

$$\beta Id \geq \bar{J}_{\mu,n}(\theta)\bar{\Sigma}^{-1}(\theta')\bar{J}_{\mu,n}(\theta) \geq \alpha Id, \tag{2.11}$$

uniformly in n. Then if Ω is compact we have that

$$E_{\theta_0} \int_{\Omega} |w(\theta|\bar{X}) - n(\theta; \theta_0, \hat{\theta})| d\theta \to 0, \tag{2.12}$$

as $n \to \infty$, where $n(\theta; \theta_0, \hat{\theta})$ is as in (2.3).

Remark: If we replace θ_0 where it occurs in the target density n with $\hat{\theta}$ and apply Sheffe's theorem (see BR pg.6) it is seen that the result continues to hold. As a result the dominated convergence theorem implies that posterior normality holds if the mode of convergence is changed to expectation with respect to $m(\bar{X})$.

Proof: In reviewing the proof of Theorem 2.1 in CG, it can be seen that most of the steps go through with only cosmetic changes. For instance, we use $m_r(\bar{X})$ and m_{r,θ_0} as defined in (2.6) and (2.7) rather than their IID analogs. Also, we replace $\mu(\theta)$, $\Sigma(\theta)$ and $J_\mu(\theta)$ by $\bar{\mu}_n(\theta)$, $\bar{\Sigma}(\theta)$, and $\bar{J}_{\mu,n}(\theta)$. There are, however, steps where the modifications are not merely a matter of notation. They are step 1, part 1; step 3, part 3; and step 4, parts 3 and 5. It will be seen that they follow largely by the uniform Taylor expandability and local invertiblity assumptions on sequences of functions.

For step 1, part 1, (2.11) ensures that the last inequality in proving the extension to (2.12) in CG continues to hold. Part 2 relies on the properties of (2.4), U_n and U_n' as before. The product $|f_{i,\theta}(\sqrt{n}(\bar{X} - \bar{\mu}_n(\theta)))\varphi_{\bar{\Sigma}(\theta)}(\sqrt{n}(\bar{X} - \bar{\mu}_n(\theta)))$ remains bounded by a constant, for n large enough and θ in a compact set. Part 3 only requires cosmetic changes.

Step 2 continues to hold, subject to cosmetic changes, once step 1 is extended. Part 1 is obvious. Part 2 only requires that one observes $P_{\theta_0}(U^c)$ tends to zero by the moment conditions.

Step 3 uses the assumptions on $< \bar{\Sigma}^{-1}(\theta) > |_{n=1}^{\infty}$. Part 1 is unchanged, and part 2 follows by the same techniques as before. The main difference occurs in part 3: the uniform Taylor expandability of $< \bar{\Sigma}^{-1}(\theta) > |_{n=1}^{\infty}$ can be used to obtain the appropriate analog of (2.29) in CG.

Step 4 requires a bit more observation. While parts 1 and 2 continue to hold, part 3 requires the local invertibility and uniform Taylor expandability of $< \bar{\mu}_n(\theta) > |_{n=1}^{\infty}$ to ensure the INID analog of (2.33) in CG goes to zero by straightforward modifications of the earlier technique. Part 4 is again cosmetic. Part 5, the last one, requires that the Laplace integration in (2.38) of CG and the bounding of the difference in the exponents in (2.40) of CG be generalized. The latter is covered by the uniform Taylor expandability of $< \bar{J}_{\mu,n}^t(\theta)\bar{\Sigma}^{-1}(\theta_0)$ $\bar{J}_{\mu,n}(\theta) > |_{n=1}^{\infty}$. The former follows by the same techniques as before. (One observes that (2.11) controls the analog to (2.39) in CG.) So, the earlier proof has been adopted to give a proof of Theorem 2.1. □

We remark 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. Furthermore, Proposition 2.1 can be extended to give a sequence of approximations in which the errors decrease as the number of terms included increases. In Theorem 2.1 we have only used a one term expansion. Higher order correction terms can be obtained by a more careful analysis of expression (2.30) in CG using the same techniques.

It is of interest to generalize one step further so as to obtain a result in the case of noncompact parameter spaces. Our technique of proof will be to reduce the result to the compact case. Thus we define two mixtures, 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 probability with density w. Our result is Theorem 2.2.

Theorem 2.2: Assume the hypotheses of Theorem 2.1. In addition, assume that for some $\delta > 0$, $\bigcap_{n=1}^{\infty} \bar{\mu}_n^{-1}(\beta(\mu(\theta_0), \delta))$ contains a nonvoid open set around θ_0 and that the moment generating functions $g_{\theta,i}(t)$ for the X_i's are defined on a common open interval around zero for θ in an open set around θ_0. Suppose that $\inf_{i \in \mathbb{N}} \inf_{\theta \in K} g_{\theta,i}(t)$ is bounded away from unity for $|t| > 0$, where K is a compact set with nonvoid interior around θ_0. Then, if a common large deviation result holds for θ outside K we have that

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

where $n(\theta; \theta_0, \hat{\theta})$ is as in (2.3).

Proof: The structure and techniques of the proof of Theorem 3.1 in CG continue to be valid. It is enough to deal with the INID analogs of (3.2), (3.3) and (3.4) in CG. The INID analog of expression (3.2) in CG goes to zero by Theorem 2.1. The remaining quantities (3.3) and (3.4) in CG go to zero provided that

$$\frac{\int_{K^c} w(\theta) p(\bar{X}|\theta) d\theta}{\int_K w(\theta) p(\bar{X}|\theta) d\theta} \xrightarrow{P_{\theta_0}} 0 \tag{2.13}$$

in the INID case. By the reasoning in CG, to show (2.13) it is enough to show that

$$P_{\theta_0}(m_{K^c}(\bar{X}) e^{nr} \geq P_{\theta_0}(\bar{X})) \leq e^{-nr'}, \tag{2.14}$$

for some $r,\ r' > 0$.

Choose K to be compact with nonvoid interior, contained in

$$\bigcap_{n=1}^{\infty}\{\theta:\ |\bar{\mu}_n(\theta)-\bar{\mu}_n(\theta_0)|<\delta\}. \tag{2.15}$$

On K we have that $|\bar{\mu}_n(\theta)-\bar{\mu}(\theta_0)|<\delta$.

We upper bound (2.14) by

$$P_{\theta_0}(|\bar{X}-\bar{\mu}_n(\theta_0)|>\delta/2)+P_{\theta_0}(|\bar{X}-\bar{\mu}_n(\theta_0)|<\delta/2, m_{K^c}(\bar{X})e^{nr}\geq P_{\theta_0}(\bar{X})). \tag{2.16}$$

The first term in (2.16) is of order $O(e^{-nr''})$ for some $r''>0$, see Chernoff (1952). The second term is bounded from above by

$$\sum_{\substack{|\bar{X}-\bar{\mu}_n(\theta_0)|<\delta/2\\ e^{nr}m_{K^c}>P_{\theta_0}}} P_{\theta_0}(\bar{X}).$$

For $\theta\in K^c$, we have $|\bar{X}-\bar{\mu}_n(\theta)|>\delta/2$ and we use $e^{nr}m_{k^c}>P_{\theta_0}$ to obtain the upper bound

$$e^{nr}\int_{K^c} w_{K^c}(\theta)P_\theta(|\bar{X}-\mu(\theta)|>\delta/2)d\theta$$

which is of order $O(e^{-n\rho})$ for some $\rho>0$, by the large deviation assumption when r is small enough. As a result, (2.14) holds. $\Box$

We remark that these results can be extended to the case that $d<k$. The main difference in hypotheses is that stronger moment conditions must be assumed. We have not done this here since the case $d=k=1$ is the most important for applications to educational testing: the psychometric orthodoxy strongly favors unidimensional parameters for discrimination purposes and test items are typically examined on an individual basis.

§3 Implications for testing independence of test items

In this section we make use of Proposition 2.1 and Theorem 2.1 to obtain a result which has implications for item response theory (IRT), the statistical theory of standardized tests. Specifically, in Junker (1991), the condition

$$\text{Cov}\ (X_i,X_j|\bar{X})\leq 0, \tag{3.1}$$

for $i \neq j$, is studied as a verifiable condition that can be used to imply unidimensionality and local asymptotic discrimination, the two main hypotheses of IRT. See also Joag-Dev and Proschan (1983). Expression (3.1) is a "manifest" condition in that it can be estimated from the data alone; it does not explicitly involve the structure of the unobservable parametric family. By contrast, "latent" conditions do explicitly use the parameter. It is argued in Junker (1991) that in IRT settings one should base inference on manifest quantities as far as possible.

Here, we use the results of Section 2 to obtain an asymptotic form of (3.1). This is useful for two applications. The first, and more important, is that one can base a test of the independence of items i and j on the convergence of Cov $(X_i, X_j|\bar{X})$ to a nonpositive number. The other is that it can be used to obtain a partial converse to a characterization result for tests which satisfy strict unidimensionality and are locally asymptotically discriminating; for definitions see Junker (1991). Stating what exactly the test is, and proving the characterization are of a specialised nature which we do elsewhere. Here we restrict our attention to obtaining a general asymptotic version of (3.1).

In the proof of Theorem 3.1 below we make use of an identity in Junker (1991). Effectively, this reduces the quantity in (3.1) to

$$\text{Cov}\ (E(X_i|\bar{X},\theta),\ E(X_j|\bar{X},\theta)|\bar{X}). \tag{3.2}$$

We show that (3.2) tends to zero as n increases. Note that the quantity in (3.2) is zero when the X_i's are IID. Indeed, it is straight forward to see that for any i

$$E(X_i|\bar{X},\theta) = \frac{1}{n}\sum_{i=1}^{n} E(X_i|\bar{X},\theta) = E(\bar{X}|\bar{X},\theta) = \bar{X},$$

which is close to $E(X_i|\theta)$. That is,

$$|E(X_i|\bar{X},\theta) - E(X_i|\theta)| \tag{3.3}$$

is small in probability. We obtain a version of (3.3) for use in the INID case; this is given in Proposition 3.1.

We begin with a lemma which will be used to control the difference between $p(X_i|S_n,\theta)$ and $p(X_i|\theta)$. In the course of the proof we use Proposition 2.1 twice, once for the density of

S_n and once for the density of $S_n - X_i$. We denote their one-term normal approximations by $q = q_{\theta n}$ and $q_i = q_{\theta n_i}$. For brevity we write $\bar{\Sigma}_i = \frac{1}{n-1} \sum_{j \neq i} \Sigma_j(\theta)$. In addition, we assume that the X_i's take values in a finite range, that their variances are uniformly bounded above and below by constant multiples of the d by d identity matrix, and that the set U_n is re-expressed as $U_{n,s}(\theta) = \{s : \sqrt{n}|\frac{s}{n} - \bar{\mu}^n(\theta)| \leq c\sqrt{\ell n\ n}\}$. Letting x denote a fixed value of X_i we have the following.

Lemma 3.1: Assume the hypotheses of Proposition 2.1 hold on a compact set K for $r = 1$. Then there is an $\xi > 0$ so that

$$\sup_{\theta \in K} \sup_{s} |\frac{P_\theta(S_n - X_i = s - x)}{P_\theta(S_n = s)} - 1|\chi_{U_{n,s}(\theta)} = O(\frac{1}{n^\xi}), \tag{3.4}$$

where χ_A is the indicator function for the set A.

Proof: Observe that by Proposition 2.1 we have that

$$P_\theta(S_n - X_i = s - x) = q_i + T_i \tag{3.5}$$

and

$$P_\theta(S_n = s) = q + T \tag{3.6}$$

where the T_i and T are error terms from the $r = 1$ term normal approximation satisfying

$$\sup_{s-x} |T_i|,\ \sup_{s} |T| = O(\frac{1}{n^{(k+1)/2}}), \tag{3.7}$$

uniformly for $\theta \in K$.

Now, consider the left hand side of (3.4) for fixed θ. Add and subtract q_i/q and use the triangle inequality to obtain the upper bound

$$|\frac{q_i + T_i}{q + T} - \frac{q_i}{q}|\chi_{U_n(\theta)} + |\frac{q_i}{q} - 1|\chi_{U_{n,s}(\theta)}. \tag{3.8a, b}$$

Apart from $\chi_{U_{n,s}(\theta)}$, expression (3.8a) is, after adding and subtracting qT, bounded from above by

$$\begin{aligned} |\frac{(q_i + T_i)q - q_i(q + T)}{q(q + T)}| &\leq \frac{|T_i - T|}{q + T} + \frac{|T(q_i - q)}{q(q + T)} \\ &= O(\frac{1}{n^{(k+1)/2}})\frac{1}{q + T} + O(\frac{1}{n^{(k+1)/2}})\frac{(q_i - q)}{q(q + T)} \end{aligned} \tag{3.9}$$

On $U_{n,s}(\theta)$ we have that there is an $\varepsilon > 0$ so that

$$q, q+T \geq O(\frac{1}{n^{(k+\varepsilon)/2}}). \qquad (3.10a, b)$$

In fact, ε may be chosen as small as desired by using small enough c in the definition of $U_{n,s}$. Using (3.10a,b) we upper bound (3.9) by

$$O(n^{(\varepsilon-1)/2}) + O(n^{(k-1+2\varepsilon)/2})|q_i - q|. \qquad (3.11)$$

Apart from $\chi_{U_{n,s}(\theta)}$ expression (3.8b) is $|q_i - q|/q$ which is bounded from above by

$$O(n^{(k+\varepsilon)/2})|q_i - q| \qquad (3.12)$$

using (3.10a). Now (3.8) is bounded from above by the sum of (3.11) and (3.12).

Since the same quantity, $|q_i - q|$, appears in both expressions, we obtain an upper bound for it. Let K denote a positive constant not necessarily the same from occurrence to occurrence. Also, let $m_i = \sum_{i \neq j} \mu_j(\theta)$. By adding and subtracting

$$|\bar{\Sigma}|^{-1/2} e^{-(n-1)(\frac{s-x}{n-1} - \frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1} - \frac{m_i}{n-1})},$$

we have that

$$\begin{aligned} |q - q_i| \leq \frac{K}{n^{k/2}} |\bar{\Sigma}|^{-1/2} |&e^{-n(s/n - \bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(s/n - \bar{\mu}^n(\theta))} \\ &- e^{-(n-1)(\frac{s-x}{n-1} - \frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1} - \frac{m_i}{n-1})}| \\ + \frac{K}{n^{k/2}} &||\bar{\Sigma}|^{-1/2} - |\bar{\Sigma}_i|^{-1/2}| \\ &|e^{-(n-1)(\frac{s-x}{n-1} - \frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1} - \frac{m_i}{n-1})}|. \end{aligned} \qquad (3.13)$$

Expression (3.13) is clearly $O(n^{-k/2})$, which is enough to ensure that expression (3.11) tends to zero, provided that we choose ε small enough to force $-1 + 2\varepsilon < 0$. (The first term in (3.11) already goes to zero.) It remains to show that (3.12) goes to zero. This requires that we obtain a faster rate of convergence to zero for (3.13).

First note that

$$||\bar{\Sigma}|^{-1/2} - |\bar{\Sigma}_i|^{-1/2}| = O(1/n). \qquad (3.14)$$

Inequality (3.14) follows by noting that $|\bar{\Sigma}|$ and $|\bar{\Sigma}_i|$ are controlled by the hypotheses on the variances of the X_i's. Indeed, take a common denominator, multiply and divide by $|\bar{\Sigma}|^{1/2} + |\bar{\Sigma}_i|^{1/2}|$, bound the denominator from below, and remove and bound the common factor $|\bar{\Sigma}_i|$ to obtain the bound

$$K||\bar{\Sigma}_i{}^{-1}\bar{\Sigma}| - 1|.$$

Apply the identity $\bar{\Sigma} = ((n-1)/n)\bar{\Sigma}_i + (1/n)\bar{\Sigma}_i$, add and subtract $|((n-1)/n)Id|$, and use the triangle inequality. One term is $O(1/n)$ immediately, the other term is seen to be $(1/n)$ by Taylor expanding the determinant function at the identity.

Now, if we use (3.13) in (3.12) we can note that $e^{-x} \leq 1$, for $x \geq 0$ so that one of the resulting terms goes to zero at rate $O(1/n^{1-\varepsilon/2})$. The other term is is bounded above (on $U_{n,s}$) by

$$O(n^{\varepsilon/2})|e^{-n(\frac{s}{n}-\bar{\mu}^n(\theta))\Sigma^{-1}(\frac{s}{n}-\bar{\mu}^n(\theta))} - e^{-(n-1)(\frac{s-x}{n-1}-\frac{m_i}{n-1})\Sigma_i^{-1}(\frac{s-x}{n-1}-\frac{m_i}{n-1})}| \tag{3.15}$$

Since ε can be made arbitrarily small, and $|\bar{\Sigma}|^{-1/2}$ is bounded by assumption, we can use the fact that $|e^{-x} - e^{-y}| \leq |x-y|$ to see that, on $U_{n,s}(\theta)$, it is enough to show

$$|n(\frac{s}{n}-\bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(\frac{s}{n}-\bar{\mu}^n(\theta))-(n-1)(\frac{s-x}{n-1}-\frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1}-\frac{m_i}{n-1})| = o(\frac{1}{n^{\xi}}) \tag{3.16}$$

for some $\xi > 0$.

We first manipulate the left hand side of (3.16) into a more useful form. Note that it is

$$\begin{aligned}|n(\frac{s}{n} - \bar{\mu}^n(\theta))|\bar{\Sigma}^{-1}(\frac{s}{n} - \bar{\mu}^n(\theta)) - \frac{n^3}{(n-1)^2}(\frac{s-x}{n} - \frac{m_i}{n})\bar{\Sigma}^{-1}(\frac{s-x}{n} - \frac{m_i}{n})& \\ + n(\frac{s-x}{n-1} - \frac{m_i}{n-1})(\bar{\Sigma}^{-1} - \bar{\Sigma}_i^{-1})(\frac{s-x}{n-1} - \frac{m_i}{n-1})& \\ + (\frac{s-x}{n-1} - \frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1} - \frac{m_i}{n-1})|&\end{aligned} \tag{3.17}$$

Since $m_i = \mu^n(\theta) - \mu_i(\theta)$ and $\frac{s-x}{n} - \frac{m_i}{n} = (s/n - \bar{\mu}^n(\theta)) - \frac{1}{n}(\mu_i - x)$ (3.17) is

$$\begin{aligned}|n(\frac{s}{n} - \bar{\mu}^n(\theta))|\bar{\Sigma}^{-1}(\frac{s}{n} - \bar{\mu}^n(\theta)) - \frac{n^3}{(n-1)^2}(\frac{s}{n} - \bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(\frac{s}{n} - \bar{\mu}^n(\theta))& \\ + \frac{2n^3}{(n-1)^2}(\frac{s}{n} - \bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(\frac{\mu_i - x}{n})&\end{aligned}$$

$$
\begin{aligned}
& - \frac{n^3}{(n-1)^2}(\frac{\mu_i - x}{n})\bar{\Sigma}^{-1}(\frac{\mu_i - x}{n}) \\
& + n(\frac{s-x}{n-1} - \frac{m_i}{n-1})(\bar{\Sigma}^{-1} - \bar{\Sigma}_i^{-1})(\frac{s-x}{n-1} - \frac{m_i}{n-1}) \\
& + (\frac{s-x}{n-1} - \frac{m_i}{n-1})\bar{\Sigma}_i^{-1}(\frac{s-x}{n-1} - \frac{m_i}{n-1})|
\end{aligned}
$$

$$
\begin{aligned}
\leq |n - \frac{n^3}{(n-1)^2}|(\frac{s}{n} - \bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(\frac{s}{n} - \bar{\mu}^n(\theta)) + \frac{1}{\sqrt{n}}(\frac{2n^2}{(n-1)^2})|\sqrt{n}(\frac{s}{n} - \bar{\mu}^n(\theta))\bar{\Sigma}^{-1}(\mu_i - x)| \\
+ \frac{n}{(n-1)^2}(\mu_i - x)\bar{\Sigma}^{-1}(\mu_i - x) + n|(\frac{s-x}{n-1} - \frac{m_i}{n-1})(\bar{\Sigma}^{-1} - \bar{\Sigma}_i^{-1})(\frac{s-x}{n-1} - \frac{m_i}{n-1})| \\
+ \frac{1}{n}(\sqrt{n}(\frac{s-x}{n-1} - \frac{m_i}{n-1}))\bar{\Sigma}_i^{-1}(\sqrt{n}(\frac{s-x}{n-1} - \frac{m_i}{n-1})) \qquad (3.18)
\end{aligned}
$$

Now, for fixed x, on $U_{n,s}(\theta)$ we have that

$$
\begin{aligned}
|\frac{s-x}{n-1} - \frac{m_i}{n-1}| &= |\frac{n}{n-1}(\frac{s}{n} - \bar{\mu}^n(\theta)) + \frac{n}{n-1}(\mu_i - x)/n| \\
&\leq K|\frac{s}{n} - \bar{\mu}^n(\theta)| + K(\frac{1}{n})|\mu_i - x| \\
&\leq K(\sqrt{\frac{\ell n\ n}{n}} + \frac{|\mu_i - x|}{n})
\end{aligned}
$$

So, squaring and using $(x+y)^2 \leq 2(x^2+y^2)$ gives

$$
|\frac{s-x}{n-1} - \frac{m_i}{n-1}|^2 \leq K\ (\frac{\ell n\ n}{n} + \frac{|\mu_i - x|^2}{n^2}) \qquad (3.19)
$$

on $U_{n,s}(\theta)$. Here, we have used the fact that the norm defined by $\bar{\Sigma}^{-1}$ is dominated by a constant times the Euclidean norm. We use (3.19) and the restriction to $U_{n,s}(\theta)$ to upper bound (3.18). (We also use other trivial bounds to handle the rational functions of n which occur.) Our bound is

$$
\begin{aligned}
& O(\frac{\ell n\ n}{n}) + |\mu_i - x|O(\frac{\ell n\ n}{n}) + |\mu_i - x|^2 O(\frac{1}{n}) \\
& + O(\ell n\ n) + \frac{|\mu_i - x|^2}{n}||\bar{\Sigma}^{-1} - \bar{\Sigma}_i^{-1}|| \\
& + O(\frac{\ell n\ n}{n} + |\mu_i - x|^2/n^2). \qquad (3.20)
\end{aligned}
$$

By reasoning similar to that used to obtain (3.14), the matrix norm in the fifth term in (3.20) is seen to be $O(1/n)$. Consequently, rearranging gives

$$
O(\frac{\ell n\ n}{n}) + |\mu_i - x|O(\sqrt{\frac{\ell n\ n}{n}}) + |\mu_i - x|^2 O(\frac{1}{n}) + (\mu_i - x)^2 O(\frac{1}{n^2}), \qquad (3.21)
$$

as an upper bound on the left hand side of (3.16), on $U_{n,s}(\theta)$. Now, expression (3.16) holds. The unformity over K is clear. □

We use the technical result in Lemma 3.1 to prove Proposition 3.1, the desired generalization of (3.3) when the X_i's assume finitely many values.

Proposition 3.1: Assume the hypotheses of Lemma 3.1. Let $\bar{X} = \frac{S}{n}$ be an element of $U_n(\theta)$. Then, there is an $\eta > 0$ so that as n increases

$$\sup_{\theta \in K} \sup_{S} |E(X_i|\bar{X}, \theta) - E(X_i|\theta)| \chi_{U_n(\theta)} = O(\frac{1}{n^\eta}). \tag{3.22}$$

Proof: Note that the left hand side of (3.22) is

$$\begin{aligned} &|\Sigma x_i P(x_i|S_n, \theta) - \Sigma x_i P(X_i|\theta)| \\ &\quad \leq \Sigma x_i P(x_i|\theta)| \frac{P_\theta(S_n - X_i = s - x_i)}{P_\theta(S_n = s)} - 1| \end{aligned} \tag{3.23}$$

For each of the finitely many values x_i, the quantity in absolute value bars in the right hand side of (3.23) is controlled by the Lemma 3.1 so the proposition is proved. □

Finally, we state the main result of this section.

Theorem 3.1: If the hypotheses of Proposition 3.1 and Theorem 2.2 are satisfied then

$$\text{Cov } (E(X_i|\bar{X}, \theta),\ E(X_j|\bar{X}, \theta))|\bar{X}| \xrightarrow{P_{\theta_0}} 0. \tag{3.24}$$

Proof: Note that

$$\text{Cov } (E(X_i|\bar{X}, \theta),\ E(X_i|\bar{X}, \theta)|\bar{X})$$

$$= \int_{B(\theta_0, \varepsilon)} \chi_U E(X_i|\theta, \bar{X}) E(X_i|\theta, \bar{X}) w(\theta|\bar{X}) d\theta \tag{3.25a}$$

$$+ \int_{B(\theta_0, \varepsilon)^c} \chi_U E(X_i|\theta, \bar{X}) E(X_j|\theta, \bar{X}) w(\theta|\bar{X}) d\theta \tag{3.25b}$$

$$+ \int \chi_{U^c} E(X_i|\theta, \bar{X}) E(X_j|\theta, \bar{X}) w(\theta|\bar{X}) d\theta \tag{3.25c}$$

$$- (\int_{B(\theta_0, \varepsilon)} \chi_U E(X_i|\theta, \bar{X}) w(\theta|\bar{X}) d\theta + \int_{B(\theta_0, \varepsilon)^c} \chi_U E(X_i|\theta, \bar{X}) w(\theta|\bar{X}) d\theta$$

$$+ \int \chi_{U^c} E(X_i|\theta,\bar{X}) w(\theta|\bar{X}) d\theta) \tag{3.26a, b, c}$$

$$(\int_{B(\theta_0,\varepsilon)} \chi_U E(X_j|\theta,\bar{X}) w(\theta|\bar{X}) d\theta + \int_{B(\theta_0,\varepsilon)^c} \chi_U E(X_i|\theta,\bar{X}) w(\theta|\bar{X}) d\theta$$
$$+ \int \chi_{U^c} E(X_j|\theta,\bar{X}| w(\theta|\bar{X}) d\theta). \tag{3.27a, b, c}$$

For terms (3.25a), (3.26a) and (3.27a) we use Proposition 3.1 to approximate the integrands with vanishing error. For terms (3.25b), (3.26b) and (3.27b) we use the fact that X_i, X_j, and χ_U are bounded. Thus their conditional expectations are bounded so the concentration of the posterior forces them to zero.

It remains to deal with terms (3.25c), (3.26c) and (3.27c) we use the local invertibility of $\bar{\mu}^n(\theta)$: Since $\bar{\Sigma}(\theta)$ is bounded above and below, we have that there is a M' so that on U^c $|\theta - \hat{\theta}| \geq M'\sqrt{\frac{\ell n\ n}{n}}$. Also, by the central limit theorem we have that, under θ_0, the probability of the set $|\hat{\theta} - \theta_0| \leq M\sqrt{\frac{\ell n\ n}{n}}$ tends to unity for any $M > 0$. By the boundedness of the integrands and the fact that the inequalities go in opposite directions we can control (3.25c), (3.26c) and (3.27c).

For instance (3.27c) is controlled in L^1 by

$$E_{\theta_0} \int \chi_{U^c} |E(X_j|\theta,\bar{X})| w(\theta|\bar{X}) d\theta \leq K(E_{\theta_0} \chi_{\{|\hat{\theta}-\theta_0| \leq M\sqrt{\frac{\ell n\ n}{n}}\}} \int \chi_{U^c} w(\theta|\bar{X}) d\theta$$
$$+ E_{\theta_0} \chi_{\{|\hat{\theta}-\theta_0| > M\sqrt{\frac{\ell n\ n}{n}}\}} \int \chi_{U^c} w(\theta|\bar{X}) d\theta)$$
$$\leq K(E_{\theta_0} \chi_{\{|\hat{\theta}-\theta_0| \leq M\sqrt{\frac{\ell n\ n}{n}}\}} \int \chi_{\{|\theta-\hat{\theta}| \geq M'\sqrt{\frac{\ell n\ n}{n}}\}} w(\theta|\bar{X}) d\theta$$
$$+ o(1))$$
$$\leq K(E_{\theta_0} \chi_{\{|\hat{\theta}-\theta_0| \leq M\sqrt{\frac{\ell n\ n}{n}}\}} \int \chi_{\{|\theta-\theta_0| \geq (M-M')\sqrt{\frac{\ell n\ n}{n}}\}} w(\theta|\bar{X}) d\theta)$$
$$+ o(1)$$

in which the integral in the last expression goes to zero by the L^1 asymptotic normality of the posterior, provided $M - M'$ is large enough. Thus (3.27c) goes to zero in P_{θ_0} probability. Terms (3.25c) and (3.26c) are similar. □

Corollary to Theorem 3.1 Assume the hypotheses of Theorem 3.1. If, in addition, the densities of the X_i's are log-concave then we have, for any fixed θ_0 and $\varepsilon > 0$, that

$$P_{\theta_0}(\text{Cov }(X_i, X_j|\bar{X}) \geq \varepsilon) \to 0. \tag{3.28}$$

Proof: By Junker's identity, see Junker (1991, §4) we have that

$$\begin{aligned}\text{Cov }(X_i, X_j|\bar{X}) &= E(\text{Cov }(X_i, X_j|\bar{X}, \theta)|\bar{X}) \\ &\quad + \text{Cov }(E(X_i|\bar{X}, \theta),\ E(X_j|\bar{X}, \theta)|\bar{X}) \qquad (3.29a, b)\end{aligned}$$

By Theorem 2.8 in Joag-dev and Proschan (1983), see also Theorem 4.1 in Junker (1991), (3.29a) is nonpositive. By Theorem 3.1, expression (3.29b) converges to zero in P_{θ_0}-probability. Thus, (3.28) follows. □

§4 Implications for grain size distributions

It is often observed that grain size distribution on a riverbed develops normality as one moves downstream from the source. Here, we give a brief sketch of an argument as to why this is so, based on Theorem 2.1. Details, along with data on an Indian river, are available in Ghosh (1988) and Ghosh, Mazumder, and Sengupta (1991).

Suppose we are observing at a distance s from the source. We assume transportation of grains is from the source and is over a period of T units which we break up into n discrete subintervals of a fixed length. For a grain or pebble, let θ stand for its size, X_j the net displacement in the j^{th} subinterval, $j = 1, ..., n$, $w(\theta)$ the size distribution at the source, $w(\theta|s)$ the size distribution at a distance s from the source, and $p_j(x_j|\theta)$ the probability function of X_j given θ. We assume a lattice valued X_j which will be true up to a roundoff error.

Let $s = \sum_{j=1}^{n} x_j$ be the total displacement. It is clear that

$$w(\theta|s) = \frac{w(\theta)P_\theta(S = s)}{\int w(\theta)P_\theta(S = s)d\theta},$$

i.e., $w(\theta|s)$ may be interpreted as a posterior, exactly as in Section 2. If s is large, time for transportation T, and hence n will be large also. Thus, Theorem 2.1 – the Bayesian version, not under a fixed θ_0 – applies. Essentially, this means that observation at distance s at a given time is equivalent to sorting out and keeping only those grains that can make it. The grain sizes of the latter will tend to cluster around $\hat{\theta}$ and a quadratic approximation will be available.

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.

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.

Clarke, B. and Ghosh, J.K. (1991). Posterior convergence given the mean. Technical Report 91-51, Department of Statistics, Purdue University.

Ghosh, J. K. (1988). The sorting hypothesis and new mathematical models for changes in size distribution for sand grains. *Indian Journal of Geology*, Vol. 60, No.1, 1-10.

Ghosh, J. K. ,Mazumder, G. S. and Sengupta, S. (1991). Experimental-theoretical approach to interpretation of grain size frequency distributions. In *Principles, Methods, and Applications of Particle Size Analysis.* Editor: James P. M. Syvitski, Cambridge University Press, Cambridge.

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

Joag-dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. Statist.*, **10**, 286–295.

Junker, B. W. (1991). Conditional association, essential independence, and monotone unidimensional item response models. To appear in *Ann. Statist.*

Lindsay, B. Clogg, C. and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association,* 86, 96-107.

Tsutakawa, R. K. and Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika* , Vol. 55, No. 2, 371-390.