Inference from the Product of Marginals
of a Dependent Likelihood

Bertrand S. Clarke
Department of Statistics
Purdue University
West Lafayette IN 47907

Brian W. Junker *
Department of Statistics
Carnegie Mellon University
Pittsburgh PA 15213

Abstract

Suppose the true density generating data $\underset{\sim}{x}^n = (x_1, \ldots, x_n)$ is in a parametric family denoted $\nu^n (\underset{\sim}{x}^n \mid \theta_1)$, where θ is a real parameter, but that $\nu^n (\underset{\sim}{x}^n \mid \theta_1)$ is not known in detail. One may try to model the data by using a different conditional distribution $q^n (\underset{\sim}{x}^n \mid \theta_1) = \prod_{i=1}^n q_i (x_i \mid \theta_1)$, which assumes independence even when this is not valid. Here we take the q_i's to be the marginals from ν^n since this is an optimal choice. The independence density q^n can be used to obtain a q^n-based MLE, $\hat{\theta}_q$, and a q^n-based posterior, w_q. We examine the performance of $\hat{\theta}_q$ and w_q under ν^n in two situations. The first situation assumes no extra structure on ν^n, only that it satisfies some laws of large numbers. The second situation assumes that ν^n may be realized as a mixture over nuisance parameters of some underlying higher dimensional conditional independence model. Under our conditions, none of the parameters in this conditional independence model need "fade out" as n increases to get consistency of estimators based on q^n.

Assessing convergence in $\nu^n (\underset{\sim}{x}^n \mid \theta_1)$, we find that w_q is consistent and asymptotically normal in both cases, with asymptotic variance unchanged from what one would expect if the data were generated by an independence model. The asymptotic distribution of $\hat{\theta}_q$ need not be normal nor be scaled by the Fisher information. Consequently, posterior inference based on the product of marginals is different from MLE-type inference derived from the product of marginals. This analysis is distinct from the usual "estimation in the presence of nuisance parameters" analysis in that we are interested in estimators based on the product of marginals q^n, not estimators based on the true dependent likelihood ν^n.

1   Introduction and motivations

It is widely believed that the two paradigms, likelihood based inference and posterior based inference, are philosophically different but asymptotically the same except in pathological cases. Here we examine two estimators, one based on maximizing a likelihood and the other based on a posterior density, which are derived from an independence model applied to dependent data. When we examine their performance in the true (dependence) model substantial asymptotic differences emerge even when typical parametric families are used. Specifically, the asymptotic variance of the likelihood based estimator may be inflated—although it may be estimated—and the posterior estimator fails to reflect the dependence present in the true model.

An important aspect of the results here is that for certain estimation procedures nuisance parameters—which empiricists will assert are always present—need not disappear asymptotically. Indeed, empiricists would argue that we never know the true density, and even if we can regard it as a member of a parametric family nuisance parameters will characterize the lack of independence and lack of identicality which cannot be assumed to attenuate even with extremely large sample sizes. Modeling all of these dependencies and nonidenticalities is fundamentally impossible so we are always using approximations which we hope are adequate. The present methods are an attempt to take these legitimate concerns into account: we estimate on the basis of an independence density (known to be wrong) but evaluate its performance in the true density, perhaps involving persistent dependency, persistent lack of identicality and persistent nuisance parameters. We obtain general results suggesting the typical behavior of our procedures and give examples which indicate their feasibility.

Our formulation of this problem is motivated by the statistical theory of standardized educational tests. In trying to measure a construct such as social adjustment, job satisfaction, school math achievement, it is common to make a set of n observations on each individual. Often, the trait is then quantified as a latent (unobservable) random variable Θ_1. A numerical value is assigned to each observation used to measure Θ_1, giving rise to random variables $\underset{\sim}{X}^n = (X_1, \ldots, X_n)$. [Replications of $(\Theta_1, X_1, \ldots, X_n)$ across individuals are considered to be i.i.d.; and we will denote outcomes of random variables with the corresponding lower case

letter.]

The "ideal" model often proposed for such data is a mixture

$$m(\underset{\sim}{x}^n) = \int q^n \left(\underset{\sim}{x}^n \mid \theta_1 \right) dF(\theta_1) \tag{1}$$

where F is the distribution of Θ_1, and $q^n(\underset{\sim}{x}^n|\theta_1)$ factors as

$$q^n \left(\underset{\sim}{x}^n \mid \theta_1 \right) = \prod_{i=1}^n q_i(x_i|\theta_1). \tag{2}$$

The goal is to infer each individual's θ_1 from his $\underset{\sim}{x}^n$, based on the right-hand side of (2).

Conditional independence models such as (2) are often assumed even though they are only approximately valid. Suppose the correct formulation is

$$m(\underset{\sim}{x}^n) = \int \nu^n \left(\underset{\sim}{x}^n \mid \theta_1 \right) dF(\theta_1), \tag{3}$$

where the conditional model for $\underset{\sim}{X}^n$ given θ_1 is some dependent $\nu^n \left(x^n \mid \theta_1 \right)$ whose structure is not known in detail. We give conditions under which asymptotic inference (as $n \to \infty$) based on the product of one-dimensional marginals where $q_i(x_1|\theta_1) = \int N^n(\underset{\sim}{x}^n|\theta_1)dx_1 \ldots dx_{i-1} \, dx_{i+1} \ldots dx_n$, may still be successful.

There are two ways in which ν^n may arise in applications. In the first, nuisance parameters prevent ν^n from factoring. If we regard Θ_1 as the first coordinate in a vector valued parameter $\underset{\sim}{\Theta}_1^d = (\Theta_1, \Theta_2, \ldots, \Theta_d)$, the "ideal" density in (2) should be replaced by

$$p^n \left(\underset{\sim}{x}^n \mid \underset{\sim}{\theta}_1^d \right) = \prod_{i=1}^n p_i \left(x_i \mid \underset{\sim}{\theta}_1^d \right). \tag{4}$$

Hence,

$$\nu^n \left(\underset{\sim}{x}^n \mid \theta_1 \right) = \int p^n \left(\underset{\sim}{x}^n \mid \underset{\sim}{\theta}_1^d \right) dF(\underset{\sim}{\theta}_2^d|\theta_1). \tag{5}$$

Typically d is so large that we cannot simply estimate $\hat{\underset{\sim}{\theta}}_1^d$ and then "throw away" $\hat{\underset{\sim}{\theta}}_2^d$. Moreover, unidimensional models are strongly favored in psychometrics applications. Indeed, the practitioner will sometimes concede that (4) is correct, but continue to use a wrong unidimensional model of the form (2) on the grounds that it is not far wrong; for several perspectives on this see Drasgow and Parsons (1983), Harrison (1986), Reckase (1979), Reckase, Ackerman and Carlson (1988), Wang (1986, 1987), and Yen (1984). Results similar to ours, but

assuming (2) is the correct likelihood, have been produced by many authors, e.g. recently Chang (1991) in the educational measurement setting.

Alternatively the dependent model $\nu^n\left(\underset{\sim}{x}^n \mid \theta_1\right)$ may also arise when meaningful secondary traits $\underset{\sim}{\Theta}_2^d$ fail to exist. The correct model is (3) in which ν^n does not factor and does not arise by mixing out nuisance parameters as in (5). Reiser (1989) suggests that the section of the NIMH Diagnostic Interview Schedule measuring major depressive disorder may fall into this case.

These two cases require distinct analyses, both of which are presented in this paper. In both cases we generate estimators from q^n, the conditional independence likelihood closest to the correct dependence model ν^n, and assess their behavior in ν^n. There is some evidence that this "best possible" case is approximately achieved in certain applications (cf. Wang, 1987). We also restrict our attention to cases in which Θ_1 or Θ_1^d, respectively, have continuous distributions $\omega(\theta_1)$ or $\omega(\theta_1^d)$. On the other hand, $\underset{\sim}{x}^n$ may be discrete or continuous.

For inference based on q^n, it is more straightforward to first suppress consideration of the "full model" $p^n\left(\underset{\sim}{x}^n \mid \theta_1^d\right)$. Stout (1987, 1990) has developed a criterion for binary data $\underset{\sim}{x}^n$ called *essential independence* which identifies θ_1 as the "dominant latent trait". This may be interpreted more generally as imposing a law of large numbers (LLN's) on ν^n,

$$\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} a_i(X_i) \middle| \theta_1\right) = 0, \tag{6}$$

for all bounded sequences of functions $\{a_i(\cdot) : i = 1, \ldots, \infty\}$. Condition (6) was applied by Junker (1991) in the educational measurement setting to the analysis of maximum likelihood estimators for θ_1 based on q^n when in fact $\nu^n\left(\underset{\sim}{x}^n \mid \theta_1\right)$ is dependent. Equation (6) imposes conditions only on the dependent likelihoods $\nu^n\left(\underset{\sim}{x}^n \mid \theta_1\right)$.

The estimators we study are derived from q^n, which is a misspecification of ν^n. We establish, both with and without assuming the existence of a full model p^n, consistency and asymptotic distribution theory for $\hat{\theta}_q$ which maximizes q^n, and for the posterior density $\omega_q(\theta_1|\underset{\sim}{x}^n)$ constructed as though q^n were the correct likelihood. Even though obtaining consistency is straightforward, the asymptotic distribution of $\hat{\theta}_q$ is a mixture of normals, whereas the posterior distribution $\omega_q(\theta_1|\underset{\sim}{x}^n)$ continues to be normal with the "independence-based" location and scale parameters.

In a model for paragraph comprehension tests we explicitly obtain the asymptotic distributions of $\hat{\theta}_q$ and ω_q; in this case, $\hat{\theta}_q$ continues to be normal with an inflated variance which can be estimated from the data. The results here all have obvious extensions to the case in which θ_1 is really of fixed dimension $d \geq 1$.

In the i.i.d. case Berk (1966) characterized the asymptotic carrier of the posterior distribution under a wrong-model analysis. Yamada (1976) extended this characterization to more general dependence structures. Here, we focus on situations in which the correct model involves some form of dependence while the incorrect model assumes independence.

In Section 2, making assumptions only on ν^n we show that the q^n-based MLE $\hat{\theta}_q$ is ν^n-consistent for θ_1. Also, we show $\omega_q(\theta_1|\underset{\sim}{x}^n)$ centered at $\hat{\theta}_q$ and scaled according to the q^n-based empirical Fisher information is asymptotically normal. Our techniques for the posterior are based on Laplace's method. A consequence is that asymptotic posterior normality in the wrong model is insensitive to the true dependence structure of the data. Indeed, Chen (1985) shows that the success or failure of Laplace's method in for posterior normality is an analytic property of the model, not the stochastic structure of the data. Our conditions overlap considerably with those used in Kass, Tierney and Kadane (1990) to define "Laplace-regularity".

In Section 3 we establish ν^n-consistency of $\hat{\theta}_q$ without LLN assumptions on ν^n, and show that $\hat{\theta}_q$ converges in distribution to a mixture of normals. However, $\omega_q(\theta_1|\underset{\sim}{x}^n)$, the formal posterior of Θ_1 given $\underset{\sim}{x}^n$, is still asymptotically normal, with the same q^n-based centering and scaling as before.

In Section 4 we consider some further implications of our results. In Appendix A we show that the product of marginals q^n is the optimal approximation to the dependent likelihood ν^n. In Appendix B, proofs are sketched for the major results of the paper.

2   Direct analysis of q^n under ν^n

In this section, only the dependent measure $\nu^n \left(\cdot \mid \theta_1 \right)$, its one-dimensional marginals $q_i \left(x_i \mid \theta_1 \right)$, and the product measure $q^n \left(\underset{\sim}{x}^n \mid \theta_1 \right) = \prod_{i=1}^{n} q_i \left(x_i \mid \theta_1 \right)$ are used. The law governing $\underset{\sim}{X}^n$ is at all times $\nu^n \left(\underset{\sim}{x}^n \mid \theta_1 \right)$; but the likelihood we will analyze is $q^n \left(\underset{\sim}{x}^n \mid \theta_1 \right)$.

2.1 Consistency of the wrong-model MLE $\hat{\theta}_q$

Let $\hat{\theta}_q$ be the MLE from $q^n\left(\underset{\sim}{X}^n \mid \theta_1\right)$. Define

$$L_n(\theta) = \log q^n\left(\underset{\sim}{X}^n \mid \theta\right) = \sum_{i=1}^{n} \log q_i(X_i|\theta) \tag{7}$$

and

$$D_n(\theta_1,\theta) \equiv \frac{1}{n}[L_n(\theta_1) - L_n(\theta)] = \frac{1}{n}\sum_{i=1}^{n} \log \frac{q_i(X_i|\theta_1)}{q_i(X_i|\theta)}.$$

Under ν^n, $E\left[D_n(\theta_1,\theta)|\,\theta_1\right] = (1/n)D\left(q_{\theta_1}^n \,\middle\|\, q_\theta^n\right)$ where D is the Kullback-Leibler distance. For each $t \in \Omega_{\Theta_1}$, define $B_\delta(t) \equiv \{\theta \in \Omega_{\Theta_1} : \; |\theta - t| < \delta\}$. In the present context, the key assumptions for a Wald-style proof of consistency can be stated as follows.

Assumption C1. For each θ_1 and $t \neq \theta_1$, there exists $c(t) > 0$, such that

$$\lim_{n\to\infty} P[D_n(\theta_1,t) > c(t)|\,\theta_1] = 1.$$

Assumption C2. For all $t \neq \theta_1$ and all $\xi > 0$, there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\left[\inf_{\theta \in B_\delta(t)} D_n(t,\theta) \geq -\xi \,\middle|\, \theta_1\right] = 1.$$

Assumption C3. There exist $c_\Delta > 0$, such that for all $\delta > 0$ and Δ sufficiently large (depending on δ), $\liminf_{n\to\infty} P\left[\inf_{|\theta|>\Delta} D_n(\theta_1,\theta) > c_\Delta \,\middle|\, \theta_1\right] \geq 1 - \delta$.

Under C1, C2 and C3 the wrong model MLE $\hat{\theta}_q$ is consistent:

Proposition 2.1 *Under Assumptions C1 through C3, for all $\epsilon > 0$ and all $\delta > 0$, there exists $\gamma = \gamma(\epsilon,\delta) > 0$ such that*

$$\liminf_{n\to\infty} P\left[\inf_{\theta \notin B_\epsilon(\theta_1)} \frac{1}{n}[L_n(\theta_1) - L_n(\theta)] \geq \gamma \,\middle|\, \theta_1\right] \geq 1 - \delta \tag{8}$$

and hence the formal MLE $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$ as $n \to \infty$ (where "$\overset{\nu^n}{\to}$" denotes convergence in ν^n-probability).

The proof is based on the usual Wald compactness argument. It is useful to identify more readily interpretable sufficient conditions for C1 and C2. Essentially, we require a LLN for $D_n(\theta_1, \theta)$, whose summands need not be bounded.

Lemma 2.1 *Suppose*

(a) *For each $t \neq \theta_1$ there exists $\beta(t) > 0$ such that $\liminf_{n \to \infty} (1/n) D\left(q_{\theta_1}^n \middle\| q_t^n\right) \geq \beta(t)$;*

(b) *As $n \to \infty$, $D_n(\theta_1, t) - (1/n) D\left(q_{\theta_1}^n \middle\| q_t^n\right) \overset{\nu^n}{\to} 0$.*

Then Assumption C1 holds.

Condition (a) can be seen to be a kind of minimum information or identifiability condition. In Section 2.4 we will see that for typical binary response data, (6) implies (b).

Lemma 2.2 *Suppose that, for all $t \neq \theta_1$ there exists $\delta_t > 0$ such that*

(a) *$\forall\, \xi > 0\ \exists\, \delta \in (0, \delta_t)$, such that $\liminf_{n \to \infty} \inf_{\theta \in B_\delta(t)} E\left[D_n(t, \theta) \middle| \theta_1\right] > -\xi$;*

(b) *$\forall\, \xi > 0\ \exists\, \delta \in (0, \delta_t)$ such that $\lim_{n \to \infty} P\left[\sup_{\theta \in B_\delta(t)} |D_n(t, \theta) - E\left[D_n(t, \theta) \middle| \theta_1\right]| < \xi \middle| \theta\right]$*
    $= 1$.

Then Assumption C2 holds.

Note that $E\left[D_n(t, \theta) \middle| \theta_1\right] = (1/n)[D(q_{\theta_1}^n \| q_\theta^n) - D(q_{\theta_1}^n \| q_t^n)] \to 0$ as $\theta \to t$; hence (a) is a locally uniform one-sided version of this continuity condition on the map $\theta \mapsto q_\theta^n$. Similarly, (b) is a locally uniform version of the WLLN $D_n(t, \theta) - E\left[D_n(t, \theta) \middle| \theta_1\right] \overset{\nu^n}{\to} 0$

2.2 The asymptotic distribution of $\hat{\theta}_q$

For inference we must know the asymptotic distribution of $\sqrt{n}\,(\hat{\theta}_q - \theta_1)/\sigma_n$, for some appropriate scale term σ_n. Taylor expansion gives

$$\sqrt{n}\,(\theta_1 - \hat{\theta}_q) = \frac{\sqrt{n}\,\overline{L}_n'(\theta_1)}{\overline{J}_n(\tilde{\theta}_1)}, \tag{9}$$

where $\overline{L}_n'(\theta_1) = (1/n)\partial L_n(\theta_1)/\partial \theta_1$, and $\overline{J}_n(\tilde{\theta}_1) \equiv -(1/n)\partial^2 \log q^n(\underline{x}^n | \tilde{\theta}_1)/\partial \theta_1{}^2$ for some $\tilde{\theta}_1 \in \{\theta : |\theta - \theta_1| < |\hat{\theta}_q - \theta_1|\}$. Assumptions such as those in Section 2.3 (see especially Assumption PN1 and Assumption PN3 below) guarantee that $\overline{J}_n(\tilde{\theta}_1)$ will behave well; the main

problem is the behavior of $\sqrt{n}\,\overline{L}'_n(\theta_1)$ under ν^n. General conditions for asymptotic normality for dependent sums have been established by Dvoretzky (1972), Iosifescu and Theodorescu (1969), Cox and Grimmett (1984), and Newman and Wright (1982). Applications to item response models are considered by Junker (1988, 1991).

2.3   Posterior asymptotics

We now turn to the possibility of basing inference for θ_1 on the formal posterior distribution

$$\omega_q\left(\theta_1 \mid \underset{\sim}{x}^n\right) = \frac{q^n\left(\underset{\sim}{x}^n \mid \theta_1\right)\omega(\theta_1)}{\int_{-\infty}^{\infty} q^n\left(\underset{\sim}{x}^n \mid \theta\right)\omega(\theta)\,d\theta} \tag{10}$$

where $\omega(\theta_1)$ is the prior density of θ_1. Of course the true posterior distribution is

$$\omega_\nu\left(\theta_1 \mid \underset{\sim}{x}^n\right) = \frac{\nu^n\left(\underset{\sim}{x}^n \mid \theta_1\right)\omega(\theta_1)}{\int_{-\infty}^{\infty} \nu^n\left(\underset{\sim}{x}^n \mid \theta\right)\omega(\theta)\,d\theta}.$$

The main result, Theorem 2.1, is that $\omega_q((\theta_1 - \hat{\theta}_q)/\sigma_n | \underset{\sim}{x}^n)$ is asymptotically normal, in the sense of Walker (1969). We make the following regularity assumptions.

Assumption PN1. Let $I_i(\theta_1) = E\left[\left(\partial \log q_i(X_i|\theta_1)/\partial\theta_1\right)^2 \middle| \theta_1\right]$ and $\overline{I}_n(\theta) = (1/n)\sum_1^n I_i(\theta)$. Then there exist $0 < \epsilon_{\theta_1} \le M_{\theta_1} < \infty$ such that $\epsilon_{\theta_1} \le \overline{I}_n(\theta_1) \le M_{\theta_1}$, for all large n.

Assumption PN2. $\int \partial^2 q_i(x|\theta_1)/\partial\theta_1^2 dx = 0$.

Assumption PN3. $M_{\epsilon,i}(x,\theta_1) = \sup_{\theta \in B_\epsilon(\theta_1)} \left|\partial^2 \log q_i(x|\theta)/\partial\theta^2 - \partial^2 \log q_i(x|\theta_1)/\partial\theta_1^2\right|$ is bounded uniformly in x and i, for small $\epsilon > 0$, ; and for $\overline{M}_n(\epsilon,\theta_1) = (1/n)\sum_1^n M_{\epsilon,i}(X_i,\theta_1)$,

$$\lim_{\epsilon \to 0}\limsup_{n \to \infty} E\left[\overline{M}_n(\epsilon,\theta_1)\middle|\theta_1\right] = 0.$$

Assumption PN4. The prior density $\omega(\theta)$ is positive and continuous throughout a small neighborhood of θ_1.

Theorem 2.1 *Assume (6) and the conclusion of Proposition 2.1. Under the additional assumptions PN1 through PN4,*

$$\sigma_n = \{-L''_n(\hat{\theta}_q)\}^{-1/2} \ge 0.$$

Then, for all $a < b$,

$$\int_{\hat\theta_q + a\sigma_n}^{\hat\theta_q + b\sigma_n} \omega_q\left(\theta \mid \underset{\sim}{X}^n\right) d\theta \overset{\nu^n}{\to} \Phi(b) - \Phi(a) \tag{11}$$

as $n \to \infty$, where $\Phi(\cdot)$ is the the standard normal c.d.f.

Proof. Under the stated assumptions it is straightforward to show that for $\theta_q^* = \hat\theta_q + r(\theta_1 - \hat\theta_q)$, where $r \in [0,1]$, and $B_\epsilon(\theta_1)$ as in Assumption PN3, then for all $\xi > 0$ there exists ϵ sufficiently small that

$$\lim_{n\to\infty} P\left[\sup_{\{r:\theta_q^* \in B_\epsilon(\theta_1)\}} \left|\frac{1}{n}L_n''(\theta_q^*) + \overline{I}_n(\theta_1)\right| < \xi \,\middle|\, \theta_1\right] = 1. \tag{12}$$

In particular,

$$(1/n)L_n''(\hat\theta_q) + \overline{I}_n(\theta_1) \overset{\nu^n}{\to} 0. \tag{13}$$

For the proof of the Theorem, break up the integral in (11) as follows:

$$
\begin{aligned}
\int_{\hat\theta_q + a\sigma_n}^{\hat\theta_q + b\sigma_n} \omega_q(\theta|\underset{\sim}{X}^n)d\theta
&= \frac{\int_{\hat\theta_q + a\sigma_n}^{\hat\theta_q + b\sigma_n} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta}{\int_{-\infty}^{\infty} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta} \\
&= \frac{\int_{\hat\theta_q + a\sigma_n}^{\hat\theta_q + b\sigma_n} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta}{\left[\int_{B_\epsilon(\theta_1)} + \int_{B_\epsilon(\theta_1)^c}\right] q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta} \\
&\equiv \frac{I_3}{I_1 + I_2},
\end{aligned}
$$

with ϵ to be determined below. The three integrals can be dealt with by modifications of the techniques found in Walker (1969), leading to

Fact 1: For all $\xi > 0$, there exists ϵ small enough that

$$\lim_{n\to\infty} P\left[|I_1/\{\sigma_n q^n(\underset{\sim}{X}^n|\hat\theta_q)\} - (2\pi)^{1/2}\omega(\theta_1)| < \xi \,\middle|\, \theta_1\right] = 1.$$

Fact 2: For each fixed $\epsilon > 0$,

$$I_2/\{\sigma_n q^n(\underset{\sim}{X}^n|\hat\theta_q)\} \overset{\nu^n}{\to} 0.$$

Fact 3: For all $\xi > 0$,

$$\lim_{n\to\infty} P\left[|I_3/\{\sigma_n q^n(\underset{\sim}{X}^n|\hat\theta_q)\} - (2\pi)^{1/2}\omega(\theta_1)[\Phi(b) - \Phi(a)]| < \xi \,\middle|\, \theta_1\right] = 1. \square$$

The LLN assumption (6) only enters into the proofs of Proposition 2.1 and the subsidiary results (12) and (13). Thus we *could* replace (6) with the subsidiary results.

A straightforward modification gives consistency of the posterior mean and higher posterior moments. For example, we have

Corollary 2.1 *In addition to the hypotheses of the Theorem 2.1, suppose*

$$\int_{-\infty}^{\infty} |t| q^n \left(\underset{\sim}{X}^n \mid t \right) \omega(t)\, dt < \infty \tag{14}$$

with ν^n-probability tending to 1 as $n \to \infty$. Then $E_q\left[\Theta_1 \mid \underset{\sim}{X}^n\right] \overset{\nu^n}{\to} \theta_1$ under ν^n.

2.4 Examples from item response theory

Item response theory, IRT, seeks to estimate an examinee's latent trait θ_1 from his responses to n individual items (questions) on a standardized multiple-choice questionnaire. Suppose each observable variable x_i has k_i values $\xi_{i1}, \ldots, \xi_{ik_i}$ where the k_i's are uniformly bounded. In practice the model often used to analyze the data is the product of marginals $\prod_{i=1}^{n} q_i\left(x_i \mid \theta_1\right)$, where $q_i\left(x_i \mid \theta_1\right) = \prod_{j=1}^{k_i} P_{ij}(\theta_1)^{Y_{ij}}$, and $Y_{ij} = 1_{\{X_i = \xi_{ij}\}}$. In many settings, the curves $P_{ij}(\theta_1)$ are considered well-enough estimated that they are taken to be known. In large-scale educational testing, for example, Wang (1986, 1987) argues that when the full model $p^n(\underset{\sim}{x}^n | \theta_1, \theta_2)$ applies, the popular response curve fitting program LOGIST produces stable estimates of $q_i(x_i|\tau)$, where τ is an appropriate one-dimensional projection of (θ_1, θ_2).

Stout's notions of *essential independence* and *essential unidimensionality* (Stout, 1987, 1990; Junker, 1988, 1991) provide conditions under which neglecting nuisance factors is reasonable. Traditional analysis of educational tests is based on averages of item response scores $\overline{A}_n = (1/n)\sum_1^n A_i$, where $A_i = \sum_{j=1}^{k_i} a_{ij} Y_{ij}$, subject to

$$\begin{aligned} &\text{(a)} \quad \exists M < \infty: \ -M \leq a_{i1} \leq a_{i2} \leq \ldots \leq a_{ik_i} \leq M, \ \forall\, i; \text{ and} \\ &\text{(b)} \quad \liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a_{ik_i} - a_{i1} > 0. \end{aligned} \tag{15}$$

Assumption (6) applies directly to such scores, and directly generalizes Stout's definition of *strong essential independence* for binary data (Definition 3.5, Stout, 1990). Note also that, because the X_i's are bounded, (6) implies (b) of Lemma 2.1, as long as the $P_{ij}(\theta_1)$

are bounded away from 0 and 1. Since estimation of θ_1 is the goal, some sort of minimum information condition is needed. Let $\overline{A}_n(\theta_1) = E\left[\overline{A}_n \middle| \theta_1\right]$. In the educational testing context it is natural to assume that, for every set of item scores satisfying (15) and every θ_1, there is an interval $B = B_\delta(\theta_1)$ and an $\epsilon > 0$ such that

$$\liminf_{n\to\infty} \frac{\overline{A}_n(t) - \overline{A}_n(\theta_1)}{t - \theta_1} \geq \epsilon, \ \forall t \in B, \ t \neq \theta_1. \tag{16}$$

This generalizes Stout's "local asymptotic discrimination," LAD, condition for binary items (Stout, 1990, Definition 3.8).

Proposition 2.2 *Suppose that* EI *and* LAD *hold, and that the response curves* P_{ij} *satisfy*

$$\text{For each } t, \ 0 < \inf_{i,j} P_{ij}(t) \leq \sup_{i,j} P_{ij}(t) < 1; \tag{17}$$

$$P_{ij}(t) \text{ is continuous at each } t, \text{ uniformly in } i \text{ and } j \tag{18}$$

and suppose Assumption C3 holds. Then the "wrong model MLE" $\hat{\theta}_q$ *is* ν^n-*consistent for* θ_1, *as* $n \to \infty$.

The proof is a routine verification of the conditions of Lemmas 2.1 and 2.2. The inequality $D\left(f \middle\| g\right) \geq (1/4)\left[\int |f(\xi) - g(\xi)|\, d\xi\right]^2$ (Csiszar, 1975) is helpful in verifying (a) of Lemma 2.1.

Example 2.1 Assumption C3 may often be verified directly. Consider the case of binary response data, in which $k_i \equiv 2$, $\xi_{i1} \equiv 0$ and $\xi_{i2} \equiv 1$. A commonly used model for the response curves is

$$P_{i2}(\theta_1) = c_i + (1 - c_i)\frac{1}{1 + \exp\{-a_i(\theta_1 - b_i)\}},$$

and $P_{i1}(\theta_1) = 1 - P_{i2}(\theta_1)$. Then $D_n(\theta_1, \theta) = (1/n)\sum_1^n t_i(\theta_1) - t_i(\theta)$, where

$$t_i(\theta) = X_i \log \frac{c_i + e^{a_i(\theta - b_i)}}{1 - c_i} - \log\left[1 + \frac{c_i + e^{a_i(\theta - b_i)}}{1 - c_i}\right].$$

Hence

$$\lim_{\theta\to\infty} -t_i(\theta) = \begin{cases} 0, \text{ if } X_i = 1, \\ \infty, \text{ if } X_i = 0; \end{cases}$$

$$\lim_{\theta\to-\infty} -t_i(\theta) = -\log c_i^{X_i}(1 - c_i)^{1 - X_i},$$

and we see that Assumption C3 holds as long as $P[X_i = 1 \,\forall i | \theta_1] = P[X_i = 0 \,\forall i | \theta_1] = 0$; this in turn follows from (6) and (17), which has a natural interpretation in terms of the a_i's, b_i's and c_i's. $\square$

Proposition 2.3 *Suppose, in addition to the assumptions of Proposition 2.2, that*

$$\frac{\partial^2}{\partial \theta^2} \log P_{ij}(\theta) \text{ is bounded pointwise in } \theta, \text{ uniformly in } i \text{ and } j. \tag{19}$$

Then, in the sense of (11),

$$\mathcal{L}_q \left\{ \left. \frac{\Theta_1 - \hat{\theta}_q}{\sigma_n} \right| \underset{\sim}{x}^n \right\} \overset{\nu^n}{\to} N(0, 1).$$

The main part of the proof, which is omitted, is a routine verification of Assumption PN1 and Assumption PN3. In the following example the asymptotic MLE and posterior distributions are different; this is an explicit case in which interval estimates for θ_1 based on the q^n-likelihood are wider than intervals based on the q^n-posterior.

Example 2.2 Consider binary responses X_1, X_2, X_3, ..., having the same response curve $P_{i2}(\theta_1) = \theta_1$ (so that $P[X_i = 1 \,|\, \theta_1] \equiv \theta_1$). Suppose that the items are arranged in successive groups of g_o items as X_1, X_2, ..., X_{g_o}; X_{g_o+1}, X_{g_o+2}, ..., X_{2g_o}; etc., such that different groups of g_o items are independent of one another, given θ, and items within a single group are positively correlated, given θ_1, and with

$$\text{Corr}\,(X_i, X_j | \theta_1) = \begin{cases} c \text{ if } X_i \text{ and } X_j \text{ are in the same group,} \\ 0 \text{ if not,} \end{cases}$$

for some fixed $c \in (0, 1]$. This ν^n might model a paragraph comprehension test in which several paragraphs are presented and g_o questions are asked for each paragraph. Here, θ_1 represents a trait common to all the items, which we might wish to think of as reading comprehension; and the nonzero correlations are induced by specific knowledge about the subject matter of the paragraph at hand, for example. This situation is also considered by Stout (1990), Junker (1991) and Wainer and Lewis (1990).

The conditions of Proposition 2.3 are easily verified, for $\theta_1 \in (0,1)$. Hence, in the sense of (11),

$$\mathcal{L}_q \left\{ \sqrt{n} \frac{\Theta_1 - \hat{\theta}_q}{\sqrt{\hat{\theta}_q (1 - \hat{\theta}_q)}} \middle| \underline{x}^n \right\} \overset{\nu^n}{\to} N(0,1),$$

where the standard error $\sqrt{\hat{\theta}_q (1 - \hat{\theta}_q)}$ is calculated directly from $L_n''(\hat{\theta}_q)$ as in the statement of Theorem 2.1, using $\hat{\theta}_q = \bar{x}_n$.

One may also easily verify the conditions of Proposition 2.2. Also, because of the block-dependent structure, it is trivial to obtain an asymptotic normality result for $\hat{\theta}_q \equiv \overline{X}_n$; we see that

$$\sqrt{n} \left(\hat{\theta}_q - \theta_1 \right) \sim AN(0, \sigma^2)$$

where $\sigma^2 = \theta_1 (1 - \theta_1)[1 + c(g_0 - 1)]$ is somewhat inflated over the anticipated asymptotic variance $\theta_1 (1-\theta_1)$ under q^n. Thus the q^n-based MLE is consistent and asymptotically normal, but has a larger asymptotic variance. $\square$

It remains to show that this inflated variance can be estimated. Observe that the marginal distributions of X_i given θ_1 are known and $T_n = \frac{1}{n} \sum_1^n X_i$ is a consistent estimator for θ_1 (under ν_{θ_1}); more generally, T_n might be any consistent estimator of θ_1. We are interested in estimating

$$
\begin{aligned}
\text{Var} \left(\sqrt{n}(T_n - \theta_1) | \theta_1 \right) &= \frac{1}{n} \sum_{i=1}^n \text{Var} \left(X_i | \theta_1 \right) \\
&\quad + \frac{1}{n} \sum_{i \neq j} [E(X_i X_j | \theta_1) - E(X_i | \theta_1) E(X_j | \theta_1)] \\
&= \theta_1 (1 - \theta_1) - (n - 1)\theta_1^2 + \frac{1}{n} \sum_{i \neq j} E(X_i X_j | \theta_1).
\end{aligned}
$$

Since θ_1 can be estimated by T_n we show how to estimate the product moments on the right hand side.

Recall that we have m i.i.d. observations of the vector $(X_1, \ldots, X_n)$; m represents the number of examinees and n represents the test length. For the kth examinee, $k = 1, \ldots, m$

we have a value of T_n, $t_{k,n} = \frac{1}{n} \sum\limits_{i=1}^{n} x_{ki}$. Let $h > 0$ and consider the kernel regression estimator

$$\mu_{ij}(t,h) = \frac{(1/mh) \sum\limits_{k=1}^{m} K(\frac{t-t_{k,n}}{h}) X_{ki} X_{kj}}{(1/mh) \sum\limits_{k=1}^{m} K(\frac{t-t_{k,n}}{h})}$$

for $E(X_i X_j | \theta_1)$, where K is a bounded kernel [for instance a $N(0,1)$ density] and t is a dummy variable.

The intuition is as follows. For fixed n, if m increases, the denominator should converge to $\omega_n(\theta_1)$, the density of T_n. Then, as n increases, the consistency of T_n should imply that $\omega_n(\theta_1)$ converges to $\omega(\theta_1)$. For the denominator, again look at what happens when n is held fixed. As m increases the numerator goes to $\int x_i x_j f(x_i, x_j, t_n) dx_i dx_j$ where f is the joint density for its arguments and the integral is with respect to counting measure. As n increases, the consistency of T_n implies this integral converges to $\int x_i x_j f(x_i, x_j, \theta_1) dx_i dx_j$. Consequently, the ratio of the two limits gives $E(X_i X_j | \theta_1)$.

Verifying this intuition requires ensuring that the discrete-valued T_n converges to a random variable with a continuous density and obtaining appropriate rates on $n, m \to \infty$ and $h \to 0$. This can be done; indeed, it can be shown that the numerator and denominator converge to their respective limits in an L^2 sense so that their ratio, $\mu_{ij}(t,h)$ converges in probability to $E(X_i X_j | \theta_1)$. The main sufficient conditions for this to be true are that $mh^2 \to \infty$, that T_n is asymptotically unbiased for all θ_1 and that Var T_n tends to zero.

The important implication of this line of reasoning is that the inflated variance characterizing the asymptotic distribution of the q-MLE can be estimated.

3 Analysis of q^n under ν^n using the full model p^n

In many settings it is natural to assume that there exists a "full model" $p^n(\underset{\sim}{x}^n | \underset{\sim}{\theta}_1^d)$ from which ν^n and q^n can be derived. If we simply apply the results of Section 2 to the full model case, we see that, under finite moment conditions, the LLN assumptions on ν^n imply that the effect of the nuisance parameters $\underset{\sim}{\theta}_2^d$ in the full model $p^n\left(\cdot | \underset{\sim}{\theta}_1^d\right)$ disappears asymptotically.

Consider (6) in this context, which asserts that the left side of the identity

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}a_i(X_i)\bigg|\theta_1\right) = \text{Var}\left(E\left[\frac{1}{n}\sum_{i=1}^{n}a_i(X_i)\bigg|\theta_{\sim 1}^{d}\right]\bigg|\theta_1\right) + E\left[\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}a_i(X_i)\bigg|\theta_{\sim 1}^{d}\right)\bigg|\theta_1\right]$$

tends to zero as $n \to \infty$, for bounded $a_i(\cdot)$. Also, the second term on the right will tend to zero by the weak law of large numbers for p^n. Hence the remaining term tends to zero, from which we may conclude, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left\{E\left[a_i(X_i)|\theta_{\sim 1}^{d}\right] - E\left[a_i(X_i)|\theta_1\right]\right\}\right| < \epsilon\bigg|\theta_1\right] = 1.$$

As a result, for $\omega\left(\theta_{\sim 2}^{d}\,|\,\theta_1\right)$-almost every $\theta_{\sim 2}^{d}$, the first moment is asymptotically free of $\theta_{\sim 2}^{d}$. This argument extends to higher moments as well.

However in the development below we do not make LLN assumptions on $\nu_{\theta_1}^{n}$. In particular this means that in some cases the nuisance parameter(s) $\theta_{\sim 2}^{d}$ need not "fade away" to produce meaningful results.

3.1   Consistency of $\hat{\theta}_q$

Our first result is an extension of Proposition 2.1 in the context of p^n. With the notation exactly as in Section 2.1 we assume:

Assumption C1′. For all $t \neq \theta_1$, there exists $c(t) \equiv c(t; \theta_{\sim 1}^{d}) > 0$ such that

$$\lim_{n\to\infty} P\left[D_n(\theta_1, t) > c(t)|\theta_{\sim 1}^{d}\right] = 1.$$

Assumption C2′. For all $t \neq \theta_1$, and all $\xi > 0$, there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\left[\inf_{\theta\in B_\delta(t)} D_n(t, \theta) > -\xi\bigg|\theta_{\sim 1}^{d}\right] = 1.$$

Assumption C3′. There exist $c_\Delta \equiv c_\Delta(\theta_{\sim 1}^{d}) > 0$, such that for all $\delta > 0$ and Δ sufficiently large (depending on δ), $\liminf_{n\to\infty} P\left[\inf_{|\theta|>\Delta} D_n(\theta_1, \theta) > c_\Delta\big|\theta_{\sim 1}^{d}\right] \geq 1 - \delta$.

Proposition 3.1 *Under Assumption C1′ through Assumption C3′, for all $\epsilon > 0$ and all $\delta > 0$ there exists $\gamma > 0$ such that*

$$\liminf_{n \to \infty} P\left[\inf_{\theta \notin B_\epsilon(\theta_1)} D_n(\theta_1, \theta) \geq \gamma \,\middle|\, \theta_1^d \right] \geq 1 - \delta. \tag{20}$$

The proof, which is omitted, is identical to that of Proposition 2.1, except that all probabilities and expectations are conditional on θ_1^d, not θ_1. Note that laws of large numbers may be expected to hold in the independence model p^n; arguments about the plausibility of Assumption C1′ and C2′ reduce to verifying the appropriate moment conditions (cf. e.g. Theorem 5.2.3 of Chung, 1974). Also, the $c(t)$ in Assumption C1′ and the c_Δ in Assumption C3′ depend on θ_2^d. This suggests what kind of uniformity argument to make, to obtain ν^n-consistency from p^n-consistency. Example 3.1 illustrates Corollary 3.1.

Corollary 3.1 *Suppose that, for any compact set $\mathcal{K} \subset \operatorname{supp} \omega(\theta_2^d | \theta_1)$, $\inf_{\theta_2^d \in \mathcal{K}} c(t) > 0$ and $\inf_{\theta_2^d \in \mathcal{K}} c_\Delta > 0$. Then for all ϵ and all δ, there exists γ such that*

$$\liminf_{n \to \infty} P\left[\inf_{\theta \notin B_\epsilon(\theta_1)} D_n(\theta_1, \theta) \geq \gamma \,\middle|\, \theta_1 \right] \geq 1 - \delta.$$

Proof. Let δ' be so small and $\mathcal{K}$ so large that $w(\mathcal{K}|\theta_1)(1 - \delta') > 1 - \delta$. Then apply Proposition 3.1 and use Fatou's lemma to obtain a lower bound on the left hand side. □

3.2 Asymptotic distribution of $\hat{\theta}_q$

For the discussion of asymptotic distribution properties of the $\hat{\theta}_q$, it is convenient to consider again the Taylor expression (9).

Assumption AN1′. Let $\gamma_n(\theta_1^d) = \sqrt{n}\, E\left[\overline{L}_n'(\theta_1) \,\middle|\, \theta_1^d \right]$, and assume there exist functions $\sigma_n^2(\theta_1^d) > 0$ such that

$$\frac{\sqrt{n}\, \overline{L}_n'(\theta_1) - \gamma_n(\theta_1^d)}{\sigma_n(\theta_1^d)} \sim AN(0, 1)$$

under $p^n\left(\cdot \,\middle|\, \theta_1^d \right)$.

Assumption AN2′. There exists $\epsilon = \epsilon(\theta_1^d) > 0$ such that $M_{\epsilon,i}(x, \theta_1) = \sup_{\theta \in B_\epsilon(\theta_1)} |\partial^2 \log q_i(x_i|\theta)/\partial\theta^2 - \partial^2 \log q_i(x_i|\theta_1)/\partial\theta_1^2|$ is dominated by some p^n-integrable function, uniformly in i, and for $\overline{M}_n(\epsilon, \theta_1) = (1/n) \sum_{i=1}^n M_{\epsilon,i}(X_i, \theta_1)$, $\lim_{\epsilon \to 0} \limsup_{n \to \infty} E\left[\overline{M}_n(\epsilon, \theta_1) \,\middle|\, \theta_1^d \right] = 0$.

Assumption AN3′. $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$ under $\nu^n\left(\cdot \mid \theta_1\right)$.

Note that, since p^n is a product measure, Assumption AN1′ is a fairly mild and natural assumption requiring only, say, the Lindeberg-Feller conditions on the summands of $\overline{L}'_n(\theta_1)$.

Proposition 3.2 *Let z_α be the standard normal cutoff, $\alpha = \Phi(z_\alpha)$ for $Z \sim N(0,1)$, and assume Assumption AN1′ through Assumption AN3′. Then*

1. For all t,

$$\lim_{n\to\infty} P\left[\sqrt{n}\,(\theta_1 - \hat{\theta}_q) \le t \,\Big|\, \theta_1\right] - E\left[\Phi\left(\frac{t\overline{J}_n(\theta_1) - \gamma_n(\theta_1, \Theta_2^d)}{\sigma_n(\theta_1, \Theta_2^d)}\right)\Big|\,\theta_1\right] = 0.$$

2. For all $\alpha \in (0,1)$, and any "centering" and "scale" terms $b(\theta_1)$ and $c(\theta_1)$,

$$\lim_{n\to\infty} P\left[\frac{\sqrt{n}\,(\overline{L}'_n(\theta_1) - b(\theta_1))}{c(\theta_1)} \le z_\alpha \,\Big|\, \theta_1\right] - E\left[\Phi\left(\frac{z_\alpha c(\theta_1) - (\gamma_n(\theta_1, \Theta_2^d) - \sqrt{n}\,b(\theta_1))}{\sigma_n(\theta_1, \Theta_2^d)}\right)\Big|\,\theta_1\right] = 0.$$

Proof. To see Part 1, note that from (9)

$$P\left[\sqrt{n}\,(\theta_1 - \hat{\theta}_q) \le t \,\Big|\, \theta_1\right] = E\left[P\left[\frac{\sqrt{n}\,(\overline{L}'_n(\theta_1) - E\left[\overline{L}'_n(\theta_1)\big|\theta_1^d\right])}{\sigma_n(\theta_1^d)} \le \frac{t\overline{J}_n(\tilde{\theta}_1) - \gamma_n(\theta_1^d)}{\sigma_n(\theta_1^d)}\,\Big|\,\theta_1^d\right]\Big|\,\theta_1\right]$$

and Part 1 follows using Assumptions AN1′, AN2′ and AN3′. Part 2 follows in a similar fashion, using Assumptions AN1′ and AN3′. □

Part 1 shows the distortion of the usual confidence intervals based on $\hat{\theta}_q$. Part 2 shows what happens if we try to force centering and scaling terms which depend only on θ_1. If we insist on having a "standard" asymptotic normality result, we are faced with investigating the stability and fixed points of integral operators, as $n \to \infty$:

$$\alpha = \lim_{n\to\infty} E\left[\Phi\left(\frac{z_\alpha \overline{J}_n(\theta_1) - \gamma_n(\theta_1, \Theta_2^d)}{\sigma_n(\theta_1, \Theta_2^d)}\right)\Big|\,\theta_1\right]$$

$$\alpha = \lim_{n\to\infty} E\left[\Phi\left(\frac{z_\alpha c(\theta_1) - (\gamma_n(\theta_1, \Theta_2^d) - \sqrt{n}\,b(\theta_1))}{\sigma_n(\theta_1, \Theta_2^d)}\right)\Big|\,\theta_1\right]$$

It is suggestive to consider the easy case in which we may interchange limit and expectation. For example, in Part 2, we would require

$$z_\alpha = \lim_{n\to\infty} \frac{z_\alpha c(\theta_1) - (\gamma_n(\theta_1^d) - \sqrt{n}\,b(\theta_1))}{\sigma_n(\theta_1^d)}$$

for all z_α; clearly this requires $c(\theta_1)/\sigma_n(\theta_1^d) \to 1$ and $\sqrt{n}(E[\overline{L}'_n(\theta_1)|\theta_1^d] - b(\theta_1))/\sigma_n(\theta_1^d) \to 0$.

3.3 Posterior asymptotics

Although Proposition 3.2 implies likelihood-based inference is complicated, the situation for posterior inference is more straightforward. As in Section 2.3, the principal ingredients are consistency of $\hat{\theta}_q$ for θ_1, and the approximation of the asymptotic information function with the empirical Fisher information.

Assumption PN1'. For each θ_1^d, there exists $0 < \epsilon \leq M < \infty$ such that

$$\epsilon \leq \liminf_{n\to\infty} \overline{I}_n(\theta_1; \theta_2^d) \leq \limsup_{n\to\infty} \overline{I}_n(\theta_1; \theta_2^d) \leq M,$$

where $\overline{I}_n(\theta_1; \theta_2^d) = -E\left[(1/n)L_n''(\theta_1)\,|\,\theta_1^d\right]$.

Assumption PN2'. The weak law of large numbers holds for $p^n\left(\cdot\,|\,\theta_1^d\right)$. In particular, we assume that $(1/n)L_n''(\theta_1) + \overline{I}_n(\theta_1; \theta_2^d) \overset{p^n}{\to} 0$.

Assumption PN3'. There exists $\epsilon = \epsilon(\theta_1^d) > 0$ such that $M_{\epsilon,i}(x, \theta_1) = \sup_{\theta \in B_\epsilon(\theta_1)}$ $|\partial^2 \log q_i(x_i|\theta)/\partial\theta^2 - \partial^2 \log q_i(x_i|\theta_1)/\partial\theta_1{}^2|$ is dominated by some p^n-integrable function, uniformly in i, and for $\overline{M}_n(\epsilon, \theta_1) = (1/n)\sum_{i=1}^n M_{\epsilon,i}(X_i, \theta_1)$, $\lim_{\epsilon \to 0} \limsup_{n\to\infty} E\left[\overline{M}_n(\epsilon, \theta_1)\,\Big|\,\theta_1^d\right]$ $= 0$.

Assumption PN4'. The prior density $\omega(\theta)$ is positive and continuous throughout a small neighborhood of θ_1.

Since p^n is a product measure, satisfying Assumption PN2' really just amounts to verifying appropriate moment conditions. Also, note that the "information" here is only computed one way; there is no analogue to Assumption PN2. Finally, Assumption PN3' is the same continuity condition as Assumption AN2'.

Theorem 3.1 *Assume the conclusion of Proposition 3.1, and suppose Assumption PN1' through Assumption PN4' hold. Let*

$$\sigma_n = \{-L_n''(\hat{\theta}_q)\}^{-1/2} \geq 0.$$

Then, for all $a < b$,

$$\int_{\hat{\theta}_q + a\sigma_n}^{\hat{\theta}_q + b\sigma_n} \omega_q\left(\theta\,|\,X^n\right) d\theta \overset{\nu^n}{\to} \Phi(b) - \Phi(a) \tag{21}$$

under $\nu^n\left(\cdot\,|\,\theta_1\right)$, as $n \to \infty$.

Proof. We modify the proof of Theorem 2.1 to give convergence under p^n, and then integrate over $\underset{\sim}{\theta}_2^d$ to obtain (21). First we observe analogs to the subsidiary results (12) and (13) in Theorem 2.1: Let $\theta_q^* = \hat{\theta}_q + r(\theta_1 - \hat{\theta}_q)$, $r \in [0,1]$; then for all $\xi > 0$, there exists $\epsilon > 0$ such that

$$\lim_{n \to \infty} P\left[\sup_{r:\theta_q^* \in B_\epsilon(\theta_1)} |\frac{1}{n}L_n''(\theta_q^*) + \overline{I}_n(\theta_1; \Theta_2^d)| < \xi \,\middle|\, \theta_1^d \right] = 1;$$

In particular,

$$(1/n)L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d) \overset{p^n}{\to} 0.$$

The main idea for the proof of Theorem 3.1 is to modify Facts 1, 2, 3 so as to assert convergence under $P[\cdot|\underset{\sim}{\theta}_1^d]$, obtaining

$$\int_{\hat{\theta}_q + a\sigma_n}^{\hat{\theta}_q + b\sigma_n} \omega_q\left(\theta \mid \underset{\sim}{X}^n\right) d\theta \overset{p^n}{\to} \Phi(b) - \Phi(a) \tag{22}$$

pointwise in $\underset{\sim}{\theta}_1^d$, and then integrate over $\underset{\sim}{\theta}_2^d$ to obtain (21). $\square$

Asymptotic distribution theory under ν^n involves mixing over $\underset{\sim}{\theta}_2^d$ and this means that there is no well-defined Fisher information resulting from p^n—unless it too is free of $\underset{\sim}{\theta}_2^d$. The proof of the subsidiary results in Theorem 2.1 breaks down if one tries to use the expected Fisher information under q^n, $\overline{I}_n(\theta_1) = E\left[-(1/n)L_n''(\theta_1)|\,\theta_1\right]$; and, although we can replace σ_n with $\sigma_n' \equiv \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d)^{-1/2}$, the result is of little practical value since the scaling would depend on $\underset{\sim}{\theta}_2^d$. Thus scaling with the observed Fisher information $\sigma_n = \{-L_n''(\hat{\theta}_q)\}^{-1/2}$ is essentially required.

3.4   Two normal-family examples

Example 3.1 below treats a multivariate normal p^n in which the location is a nuisance parameter, and Example 3.2 treats a multivariate normal p^n in which the scale is a nuisance parameter. In Example 3.1 we identify a rate at which the nuisance parameter must attenuate if asymptotic normality is to hold for $\hat{\theta}_q$. In Example 3.2, p^n continues to depend on θ_2 as $n \to \infty$; hence the requirement that dependence on $\underset{\sim}{\theta}_2^d$ attenuate as $n \to \infty$ is not necessary. We have kept the examples simple to allow closed-form calculation of integrals over $\underset{\sim}{\theta}_2^d$, and other needed quantities. More complicated examples require numerical computation.

Example 3.1 Suppose $\mathcal{L}(X_i|\theta_1, \theta_2) = N(\theta_2/\alpha_i, \theta_1)$, for fixed constants $\alpha_i \neq 0$, independent of one another, and $\mathcal{L}(\Theta_2|\theta_1) = N(0, \theta_1)$. It is easy to verify that $\mathcal{L}(X_i|\theta_1) = N(0, (\alpha_i^2+1)/\alpha_i^2)$, that

$$\frac{1}{n} \log q^n(\underset{\sim}{x}^n|t) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \theta_1 - \frac{1}{2n} \sum_{i=1}^n \log \frac{\alpha_i^2 + 1}{\alpha_i^2} - \frac{1}{2n\theta_1} \sum_{i=1}^n \frac{\alpha_i^2}{1 + \alpha_i^2} X_i^2,$$

and that consequently $\hat{\theta}_q = (1/n) \sum_1^n \alpha_i^2 X_i^2/(1+\alpha_i^2)$. Let $\overline{\alpha}_n = (1/n) \sum_1^n \alpha_i^2/(1+\alpha_i^2)$ and $\overline{\beta}_n = 1 - \overline{\alpha}_n$; then $E\left[\hat{\theta}_q \big| \theta_1, \theta_2\right] = \overline{\alpha}_n \theta_1 + \overline{\beta}_n \theta_2^2$. If $\lim_{n\to\infty} |\alpha_i| = \infty$, then $\hat{\theta}_q \overset{p^n}{\to} \theta_1$ and hence $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$. It is also easy to verify that

$$D_n(\theta_1, t) = \frac{1}{n} \log \frac{q^n(\underset{\sim}{x}^n|\theta_1)}{q^n(\underset{\sim}{x}^n|t)} = \frac{1}{2} \log \frac{t}{\theta_1} + \frac{1}{2}(\frac{1}{t} - \frac{1}{\theta_1})\hat{\theta}_q \approx \frac{1}{2} \log \frac{t}{\theta_1} + \frac{1}{2}(\frac{\theta_1}{t} - 1)$$

as $n \to \infty$. Analysis of the function $f(u) = \log u + (1/u - 1)$ shows that the assumptions C1′, C2′, C3′ and the uniformity assumptions in Corollary 3.1 are satisfied. Hence, Proposition 3.1 and Corollary 3.1 hold also.

For the asymptotic distribution of $\hat{\theta}_q$, we may calculate in Assumption AN1′ that

$$\sigma_n(\theta_1, \theta_2) = [\text{Var}\,(\overline{L}_n'(\theta_1)|\theta_1, \theta_2)]^{1/2} = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\alpha_i^2}{1+\alpha_i^2}\right) \left(\frac{1}{2\theta_1^2} + \frac{4\theta_1\theta_2^2}{4\theta_1^4\alpha_i^2}\right)\right]^{1/2},$$

which tends to $c(\theta_1) = 1/\sqrt{2}\theta_1$ as $n \to \infty$. To obtain an asymptotic normality result following Proposition 3.2 for $\hat{\theta}_q$, we also need to identify a function $b(\theta_1)$ such that $\sqrt{n}(E[\overline{L}_n'(\theta_1)|\theta_1, \theta_2] - b(\theta_1))/\sigma_n(\theta_1, \theta_2) \to 0$. Such a function $b(\theta_1)$ can be identified in this case only by assuming the stronger rate $\sqrt{n}\,\overline{\beta}_n \to 0$.

Turning to the asymptotic posterior distribution, note that

$$\begin{aligned} \overline{I}(\theta_1; \theta_2) = E\left[-\frac{1}{n}L_n''(\theta_1)\bigg|\theta_1, \theta_2\right] &= \frac{1}{2\theta_1^2} - \frac{1}{\theta_1^3}E\left[\hat{\theta}_q \big| \theta_1, \theta_2\right] \\ &= \frac{1}{\theta_1^2}(\overline{\alpha}_n - \frac{1}{2}) - \frac{\overline{\beta}_n\theta_2^2}{\theta_1^3} \\ &\to \frac{1}{2\theta_1^2}, \end{aligned}$$

as $n \to \infty$, and

$$\frac{1}{n}L_n''(\theta_1) + \overline{I}_n(\theta_1; \theta_2) = \frac{1}{\theta_1^3}[\hat{\theta}_q - (\overline{\alpha}_n\theta_1 + \overline{\beta}_n\theta_2^2)] \overset{p^n}{\to} 0,$$

as $n \to \infty$. From these it is easy to verify Assumptions PN1$'$ and PN2$'$. For Assumption PN3$'$, we note that

$$E\left[\overline{M}_n(\delta,\theta_1)\Big|\theta_1,\theta_2\right] = E\left[\sup_{t\in B_\delta(\theta_1)}\left|\frac{t-2}{2t^3}\right|\hat{\theta}_q\Big|\theta_1,\theta_2\right] = (\overline{\alpha}_n\theta_1 + \overline{\beta}_n\theta_2{}^2)\sup_{t\in B_\delta(\theta_1)}\left|\frac{t-2}{2t^3}\right|,$$

from which Assumption PN3$'$ follows. In this case the scale of the asymptotic posterior distribution is the same as for the asymptotic distribution of $\hat{\theta}_q$ (however, the rate $\sqrt{n}\,\overline{\beta}_n \to 0$ is not required). $\square$

Example 3.2 Let $\mathcal{L}(X_i|\theta_1,\theta_2)$ be i.i.d. $N(\theta_1,\theta_2^{-2})$, so that the common marginal density under p^n is $p(x|\theta_1,\theta_2) = (\theta_2/\sqrt{2\pi})\exp\{-\theta_2^2(x-\theta_1)^2/2\}$ for $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}^+$, and assume that $\omega(\theta_2|\theta_1) = c\exp\{-\theta_2^2/2\}1_{\{\theta_2\geq 1\}}$, where c is the normalizing constant. Then $q(x|\theta_1) = e^{-\frac{1}{2}[(x-\theta_1)^2+1]}/\pi[1+(x-\theta_1)^2]$. For future reference, we note that

$$\frac{\partial}{\partial\theta_1}\log q(x|\theta_1) = (x-\theta_1)\frac{(x-\theta_1)^2+3}{(x-\theta_1)^2+1}, \quad \text{and}$$

$$\frac{\partial^2}{\partial\theta_1{}^2}\log q(x|\theta_1) = -\frac{(x-\theta_1)^4+3}{[(x-\theta_1)^2+1]^2}.$$

Since $E[X_i|\theta_1,\theta_2] = \theta_1$ and a LLN holds under p^n we see that $\overline{X}_n$ is consistent for θ_1 under p^n, and hence also under q^n by integrating out θ_2. Using the Taylor expansion

$$(\hat{\theta}_q - \theta_1) = -\frac{\frac{1}{n}L_n'(\theta_1)}{\frac{1}{n}L_n''(\xi)} = -\frac{\frac{1}{n}\sum_1^n(X_i-\theta_1)\frac{(X_i-\theta_1)^2+3}{(X_i-\theta_1)^2+1}}{\frac{1}{n}\sum_1^n\frac{(X_i-\xi)^4+3}{[(X_i-\xi)^2+1]^2}}$$

where ξ is between $\hat{\theta}_q$ and θ_1, we see that the denominator is bounded away from zero, and the numerator is sandwiched between $\overline{X}_n - \theta_1$ and $3(\overline{X}_n - \theta_1)$; hence by consistency of $\overline{X}_n$, $\hat{\theta}_q$ is also consistent for θ_1.

Instead of verifying Assumptions C1$'$ through C3$'$, we will establish the conclusion of Proposition 3.1 directly. Note that

$$\begin{aligned}D_n(\theta_1,\theta) &= \frac{1}{n}\sum_{i=1}^n\{\log[(X_i-\theta)^2+1] - \log[(X_i-\theta_1)^2+1]\} \\ &\quad + \frac{1}{2n}\sum_{i=1}^n\{[(X_i-\theta)^2+1] - [(X_i-\theta_1)^2+1]\}.\end{aligned}$$

By the LLN under p^n, the second sum converges to $\frac{1}{2}(\theta_1 - \theta)^2$. Similarly, the first sum converges to its expected value under p^n, which achieves the minimum value of zero when

$\theta = \theta_1$. The convexity condition of Proposition 3.1 now follows by bounding in probability arguments. Note that by Proposition 3.2, the asymptotic distribution of $\hat{\theta}_q$ under ν^n must be a variance mixture of normals.

Using the boundedness and uniform continuity of $(t^4 + 3)/(t^2 + 1)^2$, it is easy to verify Assumptions AN1′ through AN3′, and hence under Assumption PN4′ it follows that ω_q will be asymptotically normal, centered at $\hat{\theta}_q$ and scaled by $-L_n''(\hat{\theta}_q)$. □

4   Discussion

When assumptions cannot safely be made about the dependence structure of a model a natural approach is to regard the data as coming from a distribution $\nu^n \left(\underset{\sim}{x}^n \, | \, \theta_1 \right)$ conditioned only on the parameter of interest, but otherwise unspecified. Because ν^n may be difficult to work with, practitioners may be led to use the product of marginals $q^n \left(\underset{\sim}{x}^n \, | \, \theta_1 \right)$. This is an approximation, and is the best possible in two senses discussed in Appendix A.

We have identified two physically distinct categories of problems in which asymptotic inference based on the product of marginals can proceed: Section 2 treats the case in which laws of large numbers (LLN's) are imposed upon ν^n; Section 3 treats the case in which ν^n can be embedded, as a mixture over nuisance parameters $\underset{\sim}{\theta}_2^d$, in a larger conditional independence model $p^n \left(\underset{\sim}{x}^n \, | \, \underset{\sim}{\theta}_1^d \right) = \prod_{i=1}^{n} p_i \left(x_i \, | \, \underset{\sim}{\theta}_1^d \right)$. In this second case, the LLN's which hold naturally under p^n are enough. In both settings we have obtained the consistency of the Bayesian oriented q^n-based posterior distribution $\omega_q(\theta_1 | \underset{\sim}{x}^n)$, and of the frequentist oriented q^n-based MLE $\hat{\theta}_q$, both under the true density $\nu^n \left(\underset{\sim}{x}^n \, | \, \theta_1 \right)$.

We have also examined the limiting distribution of the q^n based posterior and of the q^n based MLE. The q^n posterior ω_q is centered at $\hat{\theta}_q$ and scaled by σ_n where σ_n^{-2} is the q^n based empirical Fisher information. By contrast, the asymptotic distribution of the q^n-based MLE cannot be determined without further assumptions on ν^n or, if it is assumed to exist, p^n.

When ν^n is represented as the mixture over nuisance parameters of p^n, imposing an LLN on ν^n has the effect of forcing the dependence of $p^n \left(\cdot \, | \, \underset{\sim}{\theta}_1^d \right)$ on the nuisance parameters $\underset{\sim}{\theta}_2^d$ to attenuate as $n \to \infty$. Example 3.1 illustrates this. The attenuation is not necessary for useful asymptotic results; see Example 3.2. However, without sufficiently fast attenuation of

the nuisance parameters, the asymptotic distribution of $\hat{\theta}_q$ is a mixture of normals which is difficult to analyze.

In analyzing the asymptotic behavior of the posterior distribution (in any model) we may distinguish between concentration of the posterior around a centering value and convergence of the centering value to some limit, as n increases. This distinction might be ignored under conventional assumptions, where both convergences occur at the rate $1/\sqrt{n}$. In our situation this distinction cannot be ignored: the dependency in the data is not reflected in the inference model q^n, so that convergence of $\hat{\theta}_q$ to θ_1 may be slower than the rate of concentration of the posterior about $\hat{\theta}_q$. This may not matter to a Bayesian, for whom the notion of a "true value" of θ_1 is not meaningful. But to a frequentist, for whom the posterior distribution is another way to generate inferences about the true value of θ_1, this distinction is important since concentration around $\hat{\theta}_q$ is no longer as informative about the reliability of inference for θ_1. Even the Bayesian must be careful, since the usual interpretation of the posterior, in the sense of updating belief, cannot be used since we do not assume the data comes from $q^n\left(\underset{\sim}{x}^n \mid \theta_1\right)$.

The disparity between MLE-based and posterior-based asymptotic inference can be illustrated in a practical setting. In Example 2.2 we give a situation in educational testing which involves explicit dependence among the response variables. In this setting, asymptotic $\hat{\theta}_q$-based confidence intervals can be calculated: They are wider than the highest posterior density intervals based on the asymptotic normality of ω_q. An important aspect of this example is that the inflation of the standard error of the MLE can be estimated by the non-parametric technique outlined there. Since the width of the posterior intervals are in general the same in the dependent case and in the independent case the limit of ω_q has nothing to do with the true dependence structure. As a result, if the dependence figures at all in the asymptotic behavior of the posterior it will be seen in the rate of convergence to normality, rather than the rate of concentration about $\hat{\theta}_q$.

The fact is that practitioners, Bayesian or frequentist, typically reduce to the independent if not i.i.d. case, for both the derivation of estimators and the evaluation of their performance. In this context, Bayesian and frequentist methods usually give results which are equivalent in

practice. Here, like practitioners, we have based estimation on a product of marginals. Then we have evaluated the performance of the estimators in terms of the density reflecting the true dependent structure of the data. The discrepancy between the Bayes and frequentist results is important when the model for inference is different from the model of data generation.

In a hypothetical Bayesian-Frequentist exchange, the Bayesian would argue that failing to model the dependence omits important features of the problem and he wouldn't do it. Of course, neither would the frequentist, if it could be avoided. The frequentist would go on to assert that if you are going to omit important features of the problem, it's better to be a frequentist: You lose less since sensitivity to the omitted features (dependence) is retained.

Finally, we suggest that the greater sensitivity of $\hat{\theta}_q$ to dependence in data may be useful, even to a committed Bayesian, as a good starting place for diagnostic checks of conditional independence: for example, inflated standard errors would indicate positive dependence in ν^n that might make it worthwhile to model ν^n directly; see Junker (1991) in this regard.

Appendix A: The Best Independence Model

A.1 An estimation interpretation

As discussed in Section 1, we have based our inference for θ_1 on an independent likelihood $r^n \left(\underset{\sim}{x}^n \,|\, \theta_1 \right) = \prod_{i=1}^n r_i \left(x_i \,|\, \theta_1 \right)$, even though $\nu^n \left(\underset{\sim}{x}^n \,|\, \theta_1 \right)$ is the correct likelihood. When p^n is assumed to exist, we would like r^n to be as close as possible to p^n. When p^n cannot be assumed to exist, we would like r^n to be as close as possible to ν^n.

Recall the Kullback-Leibler distance $D\left(f \,\|\, g \right) = E[\log(f(\underset{\sim}{X}^n)/g(\underset{\sim}{X}^n))]$, where $\underset{\sim}{X}^n$ has density $f(\cdot)$. See, for example, Section 4 of Bahadur (1971) for basic properties of $D\left(\cdot \,\|\, \cdot \right)$. Fixed parameters will appear as subscripts.

Proposition A.1 $D_{\theta_1} \left(\nu^n \,\|\, r^n \right)$ *is minimized over r^n by taking $r^n \equiv q^n$.*

Indeed, following Aitchison (1975), we note that

$$D_{\theta_1} \left(\nu^n \,\|\, r^n \right) = D_{\theta_1} \left(\nu^n \,\|\, q^n \right) + \sum_{i=1}^n D_{\theta_1} \left(q_i \,\|\, r_i \right),$$

which is minimized by taking $r_i \equiv q_i$ in each term of the summation at right.

The quantity $R_{\theta_1} \left(p^n, r^n \right) = E\left[D_{(\theta_1, \underset{\sim}{\Theta}_2^d)} \left(p^n \,\|\, r^n \right) \Big| \theta_1 \right]$, can be interpreted as a 'pointwise' Bayes risk in estimating $p^n \left(\cdot \,|\, \underset{\sim}{\theta}_1^d \right)$ by $r^n \left(\cdot \,|\, \theta_1 \right)$. We have the following.

Proposition A.2 *The Bayes risk $R_{\theta_1} \left(p^n, r^n \right)$ is also minimized over r^n by taking $r^n \equiv q^n$.*

To see this, note that

$$R_{\theta_1} \left(p^n, r^n \right) \;\; = \;\; \sum_{i=1}^n R_{\theta_1} \left(p_i, r_i \right).$$

Decomposing the summands gives two nonnegative terms from which it can be seen that the sum is minimized by taking $r_i \equiv q_i$.

A.2 A Stein's Lemma interpretation

The basic data analyzed with (1) and (2) consists of i.i.d. vectors $(\Theta_{11}, \underset{\sim}{X}_1^n), \ldots, (\Theta_{1m}, \underset{\sim}{X}_m^n)$, from m individuals, where the subvectors $\underset{\sim}{X}_j^n$ are actually observed, one for each individual,

and the Θ_{1j} are latent variables. Let Γ be a collection of densities of the form $\tau(\theta_1)r_{\theta_1}^n$ where $r_{\theta_1}^n \equiv r^n(\underset{\sim}{x}^n | \theta_1)$ is any fixed independence density, and τ is any marginal density for θ_1. Consider the simple versus composite hypothesis test

$$H_0 : \omega(\theta_1)\nu_{\theta_1}^n \text{ versus } K : \tau(\theta_1)r_{\theta_1}^n \in \Gamma.$$

Let φ be the indicator function for a rejection region for H_0, which we denote by A_φ^c.

Stein's test for H_0 vs. an element of H_1 has the acceptance region

$$A_{\text{Stein } \tau r_{\theta_1}^n, \epsilon} = \left\{ \left| \frac{1}{m} \sum_{j=1}^m \log \frac{\nu^n(\underset{\sim}{X}_j^n | \theta_{1j})\omega(\theta_{1j})}{r^n(\underset{\sim}{X}_j^n | \theta_{1j})\tau(\theta_j)} - D' \right| < \epsilon \right\},$$

for H_0, where $D' = D(\omega||\tau) + \int D(\nu_{\theta_1}^n || r_{\theta_1}^n)\omega(\theta_1)d\theta_1$, see Chernoff (1956). It is well known that Stein's test has probability of type II error satisfying

$$[1 - o(\underline{1})]e^{-m(D'+\epsilon)} \leq P_{\tau r_{\theta_1}^n}\left(A_{\text{Stein } \tau r_{\theta_1}^n, \epsilon}\right) \leq e^{-m(D'-\epsilon)}.$$

As a result, if we choose ϵ so that Stein's test is level α, and we let $A_{\varphi \tau r_{\theta_1}^n}$ be the acceptance region for some other level α test φ, then from the proof of Proposition 3.C in Clarke and Barron (1990) we can deduce that

$$P_{\tau r_{\theta_1}^n}\left(A_{\varphi \tau r_{\theta_1}^n}\right) \geq (1 - 2\alpha)e^{-2m\epsilon}P_{\tau r_{\theta_1}^n}\left(A_{\text{Stein } \tau r_{\theta_1}^n, \epsilon}\right). \tag{23}$$

Now replace the $1 - 2\alpha$ by $1 - 2\alpha - \eta$, where η is small enough that (23) remains nontrivial.

Proposition A.3 *Assume there is an $\eta > 0$ so that (23) holds uniformly over Γ then the Stein test based on $A_{\text{Stein } \omega q_{\theta_1}^n, \epsilon}$ is near asymptotically minimax in the sense that*

$$\lim_{\epsilon \to 0^+} \lim_{m \to \infty} \frac{1}{m}\left(\log \min_\varphi \max_{\tau r_{\theta_1}^n} P_{\tau r_{\theta_1}^n}(A_\varphi) - \log P_{H_0}(A_{\text{Stein } \omega q_{\theta_1}^n, \epsilon})\right) = 0. \tag{24}$$

Thus, for some choices of Γ, the Stein test for H_0 versus H_1 with $\tau = \omega$ and $r_{\theta_1}^n = q_{\theta_1}^n$, is near asymptotically minimax.

Recall that a minimax test achieves $\max_\varphi \min_{\tau r_{\theta_1}^n \in \Gamma} P_{\tau r_{\theta_1}^n}(A_\varphi^c) = 1 - \min_\varphi \max_{\tau r_{\theta_1}^n \in \Gamma} P_{\tau r_{\theta_1}^n}(A_\varphi)$; typically such tests exist (e.g. Lehmann, 1959, p. 341). Let $D = \int D(\nu_{\theta_1}^n || q_{\theta_1}^n)\omega(\theta_1)d\theta_1$; evidently $-D = \lim_{\epsilon \to 0^+} \log P_{H_0}(A_{\text{Stein } \omega q_{\theta_1}^n, \epsilon})$. To establish (24), it is enough to examine

$$\frac{1}{m} \log \min_\varphi \max_{\tau r_{\theta_1}^n \in \Gamma} P_{\tau r_{\theta_1}^n}(A_\varphi).$$

Upper and lower bounds of the form $-D + \epsilon \pm O(\frac{1}{n})$ follow by straightforward manipulations.

Thus the hardest independence model to test against is the product of marginals.

Appendix B: Proofs of main results

Proof of Proposition 2.1. Let $\Omega_{\Theta_1} = S_\Delta \cup C \cup B_\epsilon(\theta_1)$ where $S_\Delta = \{\theta : |\theta| > \Delta\}$, $C = \Omega_{\Theta_1} \setminus [S_\Delta \cup B_\epsilon(\theta_1)]$, and ϵ is fixed in (8). For $t \neq \theta_1$, take $\gamma(t) = c(t)/2$ from Assumption C1 and take $\xi = \gamma(t)/2$. Then for δ as in Assumption C2,

$$
\begin{aligned}
\lim_{n \to \infty} P\left[\inf_{\theta \in B_\delta(t)} D_n(\theta_1, \theta) > \gamma(t) \,\middle|\, \theta_1 \right] &= \lim_{n \to \infty} P\left[D_n(\theta_1, t) + \inf_{\theta \in B_\delta(t)} D_n(t, \theta) > \gamma(t) \,\middle|\, \theta_1 \right] \\
&\geq \lim_{n \to \infty} P\left[D_n(\theta_1, t) + (-2) \cdot \gamma(t)/2 > \gamma(t) \,\middle|\, \theta_1 \right] \\
&= 1. \tag{25}
\end{aligned}
$$

For fixed Δ, C is a compact set and so can be covered by finitely many balls $S_1 = B_{\delta_1}(t_1)$, ..., $S_m = B_{\delta_m}(t_m)$, such that (25) holds for each: $\lim_{n \to \infty} P\left[\inf_{\theta \in S_j} D_n(\theta_1, \theta) > \gamma_j \,\middle|\, \theta_1 \right] = 1$, $j = 1, \ldots, m$. Then, letting $\gamma = \min\{\gamma_1, \ldots, \gamma_m, c_\Delta\}$, we have

$$
\begin{aligned}
\liminf_{n \to \infty} P\left[\inf_{\theta \notin B_\epsilon(\theta_1)} D_n(\theta_1, \theta) \geq \gamma \,\middle|\, \theta_1 \right] &\geq \liminf_{n \to \infty} P\left[\bigcap_{j=1,\ldots,m,\Delta} \{ \inf_{\theta \in S_j} D_n(\theta_1, \theta) \geq \gamma \} \,\middle|\, \theta_1 \right] \\
&\geq 1 - \delta,
\end{aligned}
$$

using Assumption C3 for S_Δ and (25) for each S_j, $j = 1, \ldots, m$. This is (8).

Proof of Theorem 2.1. Before proving Theorem 2.1, we require a preliminary proposition which allows us to approximate $\overline{I}_n(\theta_1)$ with $-(1/n)L_n''(\hat{\theta}_q)$ in the usual way.

Proposition B.4 *Suppose $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$, Assumptions PN1 through PN3 and (6) hold.*

(a) *Let $\theta_q^* = \hat{\theta}_q + r(\theta_1 - \hat{\theta}_q)$, where $r \in [0,1]$, and let $B_\epsilon(\theta_1)$ be as in Assumption PN3. Then for all $\xi > 0$ there exists ϵ sufficiently small that*

$$
\lim_{n \to \infty} P\left[\sup_{\{r : \theta_q^* \in B_\epsilon(\theta_1)\}} \left| \frac{1}{n} L_n''(\theta_q^*) + \overline{I}_n(\theta_1) \right| < \xi \,\middle|\, \theta_1 \right] = 1.
$$

(b) *In particular, $(1/n)L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1) \overset{\nu^n}{\to} 0$ as $n \to \infty$.*

Proof. By Assumption PN2, $\bar{I}_n(\theta_1) = -(1/n)E\left[L_n''(\theta_1)|\theta_1\right]$; hence it suffices to show each of the following, for all $\xi > 0$:

$$P\left[\frac{1}{n}|L_n''(\theta_1) - E\left[L_n''(\theta_1)|\theta_1\right]| < \xi \,\middle|\, \theta_1\right] \quad \to \quad 1; \tag{26}$$

$$P\left[\sup_{\theta_q^* \in B_\epsilon(\theta_1)} \frac{1}{n}|L_n''(\theta_q^*) - L_n''(\theta_1)| < \xi \,\middle|\, \theta_1\right] \quad \to \quad 1. \tag{27}$$

The limit (26) follows from Assumptions (6) and PN3. For (27), let $\epsilon > 0$ be small enough that Assumption PN3 holds, with $E\left[\overline{M}_n(\epsilon,\theta_1)\middle|\theta_1\right] < \xi/2$, for all large n. By assumption, both $\hat{\theta}_q \xrightarrow{\nu^n} \theta_1$ and $\theta_q^* \xrightarrow{\nu^n} \theta_1$; hence

$$\begin{aligned}
\lim_{n\to\infty} P\left[\frac{1}{n}|L_n''(\theta_q^*) - L_n''(\theta_1)| < \xi \,\middle|\, \theta_1\right] &= \lim_{n\to\infty} P\left[\theta_q^* \in B_\epsilon(\theta_1), \, \frac{1}{n}|L_n''(\theta_q^*) - L_n''(\theta_1)| < \xi \,\middle|\, \theta_1\right] \\
&\geq \lim_{n\to\infty} P\left[\overline{M}_n(\epsilon,\theta_1) < \xi \,\middle|\, \theta_1\right] \tag{28} \\
&\geq \lim_{n\to\infty} P\{|\overline{M}_n(\epsilon,\theta_1) - E\left[\overline{M}_n(\epsilon,\theta_1)\middle|\theta_1\right]| < \xi/2 \,|\, \theta_1\} \\
&= 1,
\end{aligned}$$

by (6) and Assumption PN3. Note that the bound in (28) is uniform on $B_\epsilon(\theta_1)$, giving the uniformity in (27). $\square$

For the proof of Theorem 2.1, break up the integral in (11) as follows:

$$\begin{aligned}
\int_{\hat{\theta}_q + a\sigma_n}^{\hat{\theta}_q + b\sigma_n} \omega_q(\theta|\underset{\sim}{X}^n)d\theta &= \frac{\int_{\hat{\theta}_q + a\sigma_n}^{\hat{\theta}_q + b\sigma_n} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta}{\int_{-\infty}^{\infty} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta} \\
&= \frac{\int_{\hat{\theta}_q + a\sigma_n}^{\hat{\theta}_q + b\sigma_n} q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta}{\left[\int_{B_\epsilon(\theta_1)} + \int_{B_\epsilon(\theta_1)^c}\right] q^n(\underset{\sim}{X}^n|\theta)\omega(\theta)d\theta} \\
&\equiv \frac{I_3}{I_1 + I_2},
\end{aligned}$$

with ϵ to be determined below. We will examine these three integrals in the order in which they are numbered. The calculations are modeled after Walker (1969).

Claim B.1 *For all $\xi > 0$, there exists ϵ small enough that*

$$\lim_{n\to\infty} P\left[|I_1/\{\sigma_n q^n(\underset{\sim}{X}^n|\hat{\theta}_q)\} - (2\pi)^{1/2}\omega(\theta_1)| < \xi \,\middle|\, \theta_1\right] = 1.$$

Proof. Using a two-term Taylor expansion of $L_n(\theta)$ about $\hat{\theta}_q$,

$$I_1/q^n(\underset{\sim}{X}{}^n|\hat{\theta}_q) = \int_{B_\epsilon(\theta_1)} \exp\left\{-\frac{(\theta-\hat{\theta}_q)^2}{2\sigma_n^2}(-L_n''(\theta_q^*)\sigma_n^2)\right\} \omega(\theta_1)\frac{\omega(\theta)}{\omega(\theta_1)}d\theta,$$

where $\theta_q^* = \hat{\theta}_q + r(\theta_1 - \hat{\theta}_q) \overset{\nu^n}{\to} \theta_1$ with $\hat{\theta}_q$ (by Proposition 2.1). For ξ_1 and ξ_2 to be determined momentarily, fix $\epsilon > 0$ so small that, by Proposition B.4, $P[\sup_{\theta_q^* \in B_\epsilon(\theta_1)} |L_n''(\theta_q^*)/L_n''(\hat{\theta}_q)-1| < \xi_1 \mid \theta_1] \to 1$; and by Assumption PN4, $1-\xi_2 \leq \inf_{\theta \in B_\epsilon(\theta_1)} \omega(\theta)/\omega(\theta_1) \leq \sup_{\theta \in B_\epsilon(\theta_1)} \omega(\theta)/\omega(\theta_1) \leq 1+\xi_2$. Then $I_1/q^n(\underset{\sim}{X}{}^n|\hat{\theta}_q)$ may be bounded above and below in probability by

$$\sigma_n\omega(\theta_1)(1\mp\xi_2)[2\pi/(1\pm\xi_1)]^{1/2}$$
$$\times\left[\Phi\{(1\pm\xi_1)^{1/2}\sigma_n^{-1}[\theta_1+\epsilon-\hat{\theta}_q]\} - \Phi\{(1\pm\xi_1)^{1/2}\sigma_n^{-1}[\theta_1-\epsilon-\hat{\theta}_q]\}\right].$$

The factor involving Φ tends to 1 in probability, since $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$ and $\sigma_n^{-1} \overset{\nu^n}{\to} \infty$. For each fixed ξ, an appropriate choice of ξ_1 and ξ_2 finishes the proof. $\square$

Claim B.2 *For each fixed $\epsilon > 0$,*

$$I_2/\{\sigma_n q^n(\underset{\sim}{X}{}^n|\hat{\theta}_q)\} \overset{\nu^n}{\to} 0.$$

Proof.

$$I_2/q^n(\underset{\sim}{X}{}^n|\hat{\theta}_q) = \sigma_n \exp\{L_n(\theta_1) - L_n(\hat{\theta}_q)\} \int_{B_\epsilon(\theta_1)^c} \sigma_n^{-1}\exp\{L_n(\theta) - L_n(\theta_1)\}\omega(\theta)d\theta$$
$$\leq \sigma_n \cdot 1 \cdot \int_{B_\epsilon(\theta_1)^c} O_p(\sqrt{n})O_p(\exp[-n\gamma])\omega(\theta)d\theta\,,$$

where the bound 1 follows from the fact that $L_n(\theta) \leq L_n(\hat{\theta}_q)$, $\forall\,\theta$ (by the definition of the MLE); and the O_p bounds, which are uniform in $|\theta-\theta_1| \geq \epsilon$, $\forall\,\epsilon$, follow from Proposition B.4, Assumption PN1, and Proposition 2.1. $\square$

Claim B.3 *For all $\xi > 0$,*

$$\lim_{n\to\infty} P\left[|I_3/\{\sigma_n q^n(\underset{\sim}{X}{}^n|\hat{\theta}_q)\} - (2\pi)^{1/2}\omega(\theta_1)[\Phi(b) - \Phi(a)]| < \xi\Big|\theta_1\right] = 1.$$

Proof. Let $N_n = (\hat{\theta}_q + a\sigma_n, \hat{\theta}_q + b\sigma_n)$, and fix ϵ for ξ_1 and ξ_2 as in Claim B.1. Since $P[N_n \subset B_\epsilon(\theta_1)|\theta_1] \to 1$, the argument for I_3 proceeds just as for I_1, but with N_n replacing

$B_\epsilon(\theta_1)$ throughout. We can apply the same continuity arguments as in Claim B.1 to discover that $I_3/q^n(\underset{\sim}{X}^n|\hat{\theta}_q)$ is bounded above and below, with probability tending to 1 as $n \to \infty$, by integrals of the form

$$\int_{N_n} \exp\left\{-\frac{(\theta - \hat{\theta}_q)^2}{2\sigma_n^2}(1 \pm \xi_1)\right\} \omega(\theta_1)(1 \mp \xi_2)d\theta$$
$$\approx \sigma_n\omega(\theta_1)(1 \mp \xi_2)[2\pi/(1 \pm \xi_1)]^{1/2}\left[\Phi\{(1 \pm \xi_1)^{1/2}b\} - \Phi\{(1 \pm \xi_1)^{1/2}a\}\right].$$

The proof is now completed as for Claim B.1. $\square$

Remarks. The fact that asymptotic normality results should depend little on the true dependence structure of the data was made clear by Chen (1985). Kass, Tierney and Kadane (1990) have identified a class of models in which the Chen/Walker-style argument works well, called the "Laplace regular" models. We note that (i) the continuity and boundedness conditions of Laplace regularity correspond to our Assumption PN3; (ii) the positivity of the Hessian for Laplace regularity corresponds to our Assumption PN1 and Proposition B.4, and (iii) their asymptotic convexity condition is our Proposition 2.1.

Proof of Theorem 3.1.

Proposition B.5 *Suppose $\hat{\theta}_q \overset{\nu^n}{\to} \theta_1$, and assumptions PN1′ through PN3′ hold.*

(a) Let $\theta_q^* = \hat{\theta}_q + r(\theta_1 - \hat{\theta}_q)$, $r \in [0,1]$. Then for all $\xi > 0$, there exists $\epsilon > 0$ such that

$$\lim_{n\to\infty} P\left[\sup_{r:\theta_q^* \in B_\epsilon(\theta_1)} |\frac{1}{n}L_n''(\theta_q^*) + \overline{I}_n(\theta_1; \underset{\sim}{\Theta}_2^d)| < \xi \middle| \underset{\sim}{\theta}_1^d\right] = 1;$$

(b) In particular, $(1/n)L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d) \overset{p^n}{\to} 0$.

Proof. The proof of part (a) proceeds as for Lemma B.4, replacing conditioning on θ_1 with conditioning on $\underset{\sim}{\theta}_1^d$. Note that part (b) would follow immediately as long as $\hat{\theta}_q \overset{p^n}{\to} \theta_1$. But this must be true, for almost all $\underset{\sim}{\theta}_2^d$: Suppose on some measurable $\mathcal{K} \subset \text{supp}\,\omega(\underset{\sim}{\theta}_2^d|\theta_1)$, that $|L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d)| > \xi'$. Then

$$E\left[1_\mathcal{K} P\left[|L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d)| > \xi' \middle| \underset{\sim}{\theta}_1^d\right] \middle| \theta_1\right] \le P\left[|L_n''(\hat{\theta}_q) + \overline{I}_n(\theta_1; \underset{\sim}{\theta}_2^d)| > \xi' \middle| \theta_1\right] \to 0$$

and hence $\omega(\mathcal{K}|\theta_1) = 0$. $\square$

The main idea for the proof of Theorem 3.1 is to substitute Proposition B.5 for Proposition B.4 and modify the Claims B.1 through B.3 to assert convergence under $P[\cdot|\underset{\sim}{\theta}_1^d]$, obtaining

$$\int_{\hat{\theta}_q+a\sigma_n}^{\hat{\theta}_q+b\sigma_n} \omega_q\left(\theta \mid \underset{\sim}{X}^n\right) d\theta \xrightarrow{p^n} \Phi(b) - \Phi(a) \tag{29}$$

pointwise in $\underset{\sim}{\theta}_1^d$, and then integrate over $\underset{\sim}{\theta}_2^d$ to obtain (21).

Remarks. The asymptotic distribution theory under ν^n necessarily involves mixing over $\underset{\sim}{\theta}_2^d$ and this means that there is no well-defined Fisher information resulting from p^n—unless it too is free of $\underset{\sim}{\theta}_2^d$. The proof of Proposition B.5 breaks down if one tries to use the expected Fisher information under q^n, $\overline{I}_n(\theta_1) = E\left[-(1/n)L_n''(\theta_1)|\,\theta_1\right]$; and, although Proposition B.5 does allow us to replace our definition of σ_n with $\sigma_n' \equiv \overline{I}_n(\theta_1;\underset{\sim}{\theta}_2^d)^{-1/2}$, the result would be of little practical value since the scaling would depend in an unwieldy fashion upon $\underset{\sim}{\theta}_2^d$. Thus we see that scaling with $\sigma_n = \{-L_n''(\hat{\theta}_q)\}^{-1/2}$ is essentially required to obtain a useful result.

References

[Aitchison (1975)] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547–554.

[Bahadur (1971)] Bahadur, R. R. (1971). *Some limit theorems in statistics*. Philadelphia: Society for Industrial and Applied Mathematics.

[Berk (1966)] Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, **37**, 51–58.

[Chang (1991)] Chang, H.-H. (1991). Asymptotic posterior normality of IRT models. Submitted to *Psychometrika*.

[Chen (1985)] Chen, C.-F. (1985). On asymptotic normality of limiting density functions with Bayesian implications. *Journal of the Royal Statistical Society, Series B*, **47**, 540-546.

[Chernoff (1956)] Chernoff, H. (1956). Large sample theory: parametric case. *Annals of Mathematical Statistics*, **27**, 1–22.

[Chung (1974)] Chung, K. L. (1974). *A Course in Probability Theory*. Second Edition. New York: Academic Press.

[Clarke and Barron (1989)] Clarke, B. S. and Barron, A. R. (1989). Information-theoretic asymptotics of Bayes methods. *IEEE Transactions on Information Theory*, **36**, 453–471.

[Cox and Grimmett (1984)] Cox, J. T. and Grimmett, G. (1984). Central limit theorem for associated random variables and the percolation model. *Annals of Probability*, **12**, 514–528.

[Csiszar (1975b)] Csiszar, I. (1975). Information type measures of difference of probability distributions and direct observations. *Studia Scientiarum Mathematicum Hungarica*, **2**, 299–318.

[Drasgow and Parsons (1983)] Drasgow, F. and Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, **7**, 189–199.

[Dvoretzky (1972)] Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Probability and Statistics*, Volume II, 513–535.

[Harrison (1986)] Harrison, D. A. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics*, **11**, 91–115.

[Iosifescu and Theodorescu (1969)] Iosifescu, M. and Theodorescu, M. (1969). *Random processes and learning*. New York: Springer-Verlag.

[Junker (1988)] Junker, B. W. (1988). *Statistical aspects of a new latent trait model*, unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Department of Statistics.

[Junker (1991)] Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, **56**, 255–278.

[Kass, Tierney and Kadane (1990)] Kass, R. E., Tierney, L. and Kadane, J. B. (1990). The validity of posterior expansions based on Laplace's method. In Geisser, S., Hodges, J. S., Press, S. J, and Zellner, A., ed's. (1990). *Bayesian and likelihood methods in statistics and econometrics: Essays in honor of George A. Barnard* (pp. 473–488). New York: North-Holland.

[Lehmann (1959)] Lehmann, E. L. (1959). *Testing statistical hypotheses.* John Wiley and Sons. New York.

[Newman and Wright (1982)] Newman, C. M. and Wright, A. L. (1982). Associated random variables and martingale inequalities. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **59**, 361–371.

[Reckase (1979)] Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics*, **4**, 207–230.

[Reckase, Ackerman and Carlson (1988)] Reckase, M. D., Ackerman, T. A., and Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement*, **25**, 193–203.

[Reiser (1989)] Reiser, M. (1989). An application of the item-response model to psychiatric epidemiology. *Sociological Methods and Research*, **18**, 66–103.

[Rosenbaum (1988)] Rosenbaum, P. R. (1988). Item bundles. *Psychometrika*, **53**, 349-359.

[Stout (1987)] Stout, W. F. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, **52**, 79–98.

[Stout (1990)] Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, **55**, 293–325.

[Wainer and Lewis (1990)] Wainer, H. and Lewis, C. (1990). Towards a psychometrics for testlets. *Journal of Educational Measurement*, **27**, 1–14.

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

[Wang (1986)] Wang, M.-M. (1986). *Fitting a unidimensional model to multidimensional item response data*. Paper presented at Office of Naval Research Model-Based Measurement Contractors' Meeting, Knoxville, TN, April 28, 1986.

[Wang (1987)] Wang, M.-M. (1987). *Estimation of ability parameters from response data that are precalibrated with a unidimensional model*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC, April 22, 1987.

[Yamada (1976)] Yamada, K. (1976). Asymptotic behavior of distributions for random processes under incorrect models. *Journal of Mathematical Analysis and Applications*, **56**, 294–308.

[Yen (1984)] Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, **8**, 125-145.