EXISTENCE OF THE MLE AND PROPRIETY OF POSTERIORS FOR A GENERAL MULTINOMIAL CHOICE MODEL

Paul L. Speckman, Jaeyong Lee and Dongchu Sun

The University of Missouri-Columbia, Seoul National University and The University of Missouri-Columbia

Supplementary Material

This supplement contains proofs of lemmas in Appendix A and some supplemental large sample theory results in Appendix B. All section and equation numbers refer to the main article.

Appendix A. Proofs

PROOF OF LEMMA 1. We show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

(i) $\Rightarrow$ (ii): The dual cone of $coni(\mathcal{A})$ is $coni(\mathcal{A})^* = \{\boldsymbol{b} : \boldsymbol{b}^t\boldsymbol{z}_i \le 0 \text{ for all } \boldsymbol{z}_i \in \mathcal{A}\}$. By the Duality Theorem for finite cones (e.g. Panik, 1993, Theorem 4.2.1), $coni(\mathcal{A})^{**} = coni(\mathcal{A})$. Suppose that (ii) is false, i.e. $coni(\mathcal{A}) \ne \mathbb{R}^m$. Then there is a nonzero $\boldsymbol{b} \in coni(\mathcal{A})^*$. But then $-\boldsymbol{b}^t\boldsymbol{z}_i \ge 0$ for all $\boldsymbol{z}_i \in \mathcal{A}$, and there is a quasi-complete separation of the sample $(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)})$. This contradicts the assumption of overlap, hence (ii) holds.

(ii) $\Rightarrow$ (iii): Assume that $coni(\mathcal{A}) = \mathbb{R}^m$, and let $\boldsymbol{b} \in \mathbb{R}^m$ be arbitrary. Since (ii) holds, $-\boldsymbol{b} \in coni(\mathcal{A})$ and

$$-\boldsymbol{b} = \sum_{j=1}^{i} \lambda_j \boldsymbol{z}_j,\ \ \boldsymbol{z}_j \in \mathcal{A},$$

for some constants $\lambda_j \ge 0$. Hence

$$0 \le \boldsymbol{b}^t\boldsymbol{b} = \sum_{j=1}^{i} \lambda_j(-\boldsymbol{b}^t\boldsymbol{z}_j) \le \sum_{j=1}^{i} \lambda_j \|\boldsymbol{b}\|_{\mathcal{A}} = \|\boldsymbol{b}\|_{\mathcal{A}} \sum_{j=1}^{i} \lambda_j.$$

Suppose $\boldsymbol{b} \ne 0$ but $\|\boldsymbol{b}\|_{\mathcal{A}} < 0$. Then $\lambda_j > 0$ for some j, and the right side of (5.1) is strictly negative, a contradiction. Therefore (2.6) holds. Moreover, if $\|\boldsymbol{b}\|_{\mathcal{A}} = 0$, then $\boldsymbol{b}^t\boldsymbol{b} = 0$ by (5.1) again, and (2.7) holds.

(iii) $\Rightarrow$ (i): If (2.7) holds and $\boldsymbol{0} \ne \boldsymbol{b} \in \mathbb{R}^m$, $\|\boldsymbol{b}\|_{\mathcal{A}} = -\min_{\boldsymbol{z}\in\mathcal{A}} \boldsymbol{z}^t\boldsymbol{b} > 0$. Thus there is a $\boldsymbol{z} \in \mathcal{A}$ such that $\boldsymbol{z}^t\boldsymbol{b} < 0$. □

PROOF OF COROLLARY 1. The proof, which is essentially the same as the proof of equivalence of norms in a finite-dimensional normed space (see e.g. Schecter, p. 83), is included here for completeness. Suppose that the conclusion of the corollary does not hold. Then there is a sequence $\boldsymbol{b}_n = (b_{n1}, \ldots, b_{nm})^t$, $n = 1, 2, \ldots$, such that $\|\boldsymbol{b}_n\|_{\mathcal{A}}/\|\boldsymbol{b}_n\| \to 0$ as $n \to \infty$. Since (2.8) holds for positive α, without loss of generality assume $\|\boldsymbol{b}_n\| = 1$ for all n. The unit sphere in $\mathbb{R}^m$ is compact, so there is a convergent subsequence $\boldsymbol{b}_{n(k)} \to \boldsymbol{b}$

with $\|\boldsymbol{b}\| = 1$. Clearly $\|\cdot\|_{\mathcal{A}}$ is continuous. Then this implies that $\|\boldsymbol{b}\|_{\mathcal{A}} = 0$ for some $\boldsymbol{b} \neq 0$, contradicting (iii) of Lemma 1, and the proof is complete. □

PROOF OF LEMMA 2. Without loss of generality, assume that $y_i = 0$ for $i = 1, \ldots, r$ and $y_i = 1$ for $i = r+1, \ldots, n$. We show that (S1) and (S2) are equivalent to (ii) of Lemma 1.

(ii) ⇒ (S1) and (S2):

For the binary case, suppose that $coni(\mathcal{A}) = \mathbb{R}^m$. Then from (2.9), the set $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ has full rank m. Suppose neither $\mathcal{S}$ nor $\mathcal{F}$ is all of $\mathbb{R}^m$. Since $\mathbf{0}$ is in $\mathbb{R}^m$, it follows from Lemma 3 in Appendix A that there are $\lambda_i > 0$ so that $\mathbf{0} = \sum_{i=1}^r \lambda_i \boldsymbol{x}_i + \sum_{i=r+1}^n \lambda_i(-\boldsymbol{x}_i)$, or $\sum_{i=1}^r \lambda_i \boldsymbol{x}_i = \sum_{i=r+1}^n \lambda_i \boldsymbol{x}_i$. Therefore $\mathcal{S} \cap \mathcal{F} \neq \emptyset$.

(S1) and (S2) ⇒ (ii):

If either $\mathcal{S}$ or $\mathcal{F}$ is $\mathbb{R}^m$, then clearly $coni(\mathcal{A}) = \mathbb{R}^m$. Suppose $\mathcal{S} \cap \mathcal{F} \neq \emptyset$. Let $\boldsymbol{z} \in \mathbb{R}^m$. Since the rank of $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is m, there are constants $c_1, \ldots, c_n$ such that

$$\boldsymbol{z} \;=\; \sum_{i=1}^{r} c_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} c_i \boldsymbol{x}_i \;=\; \sum_{i=1}^{r} c_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} (-c_i)(-\boldsymbol{x}_i). \tag{5.1}$$

Since $\mathcal{S} \cap \mathcal{F} \neq \emptyset$, there is a $\boldsymbol{u} \in \mathcal{S} \cap \mathcal{F}$. Thus $\boldsymbol{u} = \sum_{i=1}^r a_i \boldsymbol{x}_i = \sum_{i=r+1}^n b_i \boldsymbol{x}_i$, where $a_i, b_i > 0$, and

$$\mathbf{0} = \sum_{i=1}^{r} a_i \boldsymbol{x}_i - \sum_{i=r+1}^{n} b_i \boldsymbol{x}_i. \tag{5.2}$$

Combining (5.1) and (5.2), we know for any constant M,

$$\boldsymbol{z} \;=\; \sum_{i=1}^{r} c_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} (-c_i)(-\boldsymbol{x}_i) + \sum_{i=1}^{r} M a_i \boldsymbol{x}_i + \sum_{i=r+1}^{n} M b_i (-\boldsymbol{x}_i).$$

Choose $M > 0$ large enough so that

$$\begin{aligned} M a_i + c_i &\geq 0, \quad i = 1, \ldots, r; \\ M b_i - c_i &\geq 0, \quad i = r+1, \ldots, n. \end{aligned}$$

The result then follows. □

Lemma 3 *Suppose that $coni(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = \mathbb{R}^m$.*
(a) There are positive constants $\lambda_1 > 0, \ldots, \lambda_k > 0$, such that $\sum_{i=1}^k \lambda_i \boldsymbol{z}_i = 0$;
(b) For any $\boldsymbol{z} \in \mathbb{R}^m$, there exist constants $\lambda_1 > 0, \ldots, \lambda_k > 0$, such that $\boldsymbol{z} = \sum_{i=1}^k \lambda_i \boldsymbol{z}_i$.

PROOF. Since $coni(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k) = \mathbb{R}^m$, there are constants $C_{ij} \geq 0$, so that

$$-\boldsymbol{z}_i = \sum_{i=1}^{k} C_{ij} \boldsymbol{z}_j, \text{ for } i = 1, \ldots, k.$$

Then,

$$\begin{aligned} 0 \;&=\; \sum_{i=1}^{k} \boldsymbol{z}_i - \sum_{i=1}^{k} \boldsymbol{z}_i \;=\; \sum_{j=1}^{k} \boldsymbol{z}_j + \sum_{i=1}^{k}\sum_{j=1}^{k} C_{ij} \boldsymbol{z}_j \\ &=\; \sum_{j=1}^{k} (1 + \sum_{i=1}^{k} C_{ij}) \boldsymbol{z}_j \;=\; \sum_{j=1}^{k} \lambda_j \boldsymbol{z}_j, \end{aligned}$$

where $\lambda_j = 1 + \sum_{i=1}^k C_{ij} \geq 1 > 0$. Part (a) holds. For Part (b), for any $\boldsymbol{z}$, there are $d_i \geq 0$ so that $\boldsymbol{z} = \sum_{i=1}^k d_i \boldsymbol{z}_i$. From part (a), there are $c_i > 0$ so that $\sum_{i=1}^k c_i \boldsymbol{z}_i = 0$. Thus

$$\boldsymbol{z} = \sum_{i=1}^k c_i \boldsymbol{z}_i + \sum_{i=1}^k d_i \boldsymbol{z}_i = \sum_{i=1}^k (c_i + d_i) \boldsymbol{z}_i,$$

and $c_i + d_i > 0$ for all i. Part (b) follows. □

Appendix B. Large Sample Properties of MLE and the Posterior

C.1. Notation and Assumptions. We consider the general multinomial choice model given by (1.1) and (1.4) in Section . Assume that for $i = 1, \ldots, n$, the $m \times k$ matrices $\boldsymbol{X}_i$'s are i.i.d. with distribution function F. Given $\boldsymbol{X}_i$, y_i is multinomial$(1, \boldsymbol{p}(\boldsymbol{X}_i, \boldsymbol{\beta}))$, where $\boldsymbol{p}(\boldsymbol{X}_i, \boldsymbol{\beta}) = (p_1(\boldsymbol{X}_i, \boldsymbol{\beta}), \ldots, p_k(\boldsymbol{X}_i, \boldsymbol{\beta}))$, and $p_j(\boldsymbol{X}_i, \boldsymbol{\beta})$ is defined by (1.4). Let $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{ik}), i = 1, \ldots, n$, be i.i.d. random vectors with a permutation invariant distribution. Let $G(u_1, \ldots, u_{k-1})$ be the $(k-1)$-dimensional common distribution function of $(\xi_{i1} - \xi_{ik}, \ldots, \xi_{i,k-1} - \xi_{ik})$, and define

$$\boldsymbol{A}(\boldsymbol{X}_i, j) = ((\boldsymbol{x}_{ij} - \boldsymbol{x}_{i1}), \ldots, (\boldsymbol{x}_{ij} - \boldsymbol{x}_{i,j-1}), (\boldsymbol{x}_{ij} - \boldsymbol{x}_{i,j+1}), \ldots, (\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik}))^t.$$

Then

$$p_j(\boldsymbol{X}_i, \boldsymbol{\beta}) \quad = \quad G((\boldsymbol{x}_{ij} - \boldsymbol{x}_{i1})^t \boldsymbol{\beta}, \ldots, (\boldsymbol{x}_{ij} - \boldsymbol{x}_{ik})^t \boldsymbol{\beta}) \;=\; G(\boldsymbol{A}(\boldsymbol{X}_i, j)\boldsymbol{\beta}).$$

The likelihood of $\boldsymbol{\beta}$ based on $(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)})$ is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta}).$$

Denote $\mathcal{A}$ defined in (2.4) as $\mathcal{A}^{(n)}$, and define

$$C(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}) \quad = \quad \sup\{C > 0 : \min_{\boldsymbol{z} \in \mathcal{A}^{(n)}} \boldsymbol{z}^t \boldsymbol{\beta} \leq -C\|\boldsymbol{\beta}\| \text{ for all } \boldsymbol{\beta} \in \mathbb{R}^m\}.$$

Let $\nabla_i f$, $\nabla^2_{ij} f$, and $\nabla^3_{ijk} f$ be the ith first order partial derivative, the (i,j)th second order partial derivative, and the (i,j,k)th third order partial derivative of f respectively. Also, let $(\boldsymbol{B})_{ij}$ be the (i,j)th coordinate of a matrix $\boldsymbol{B}$. Let $\boldsymbol{\beta}_0$ be the true parameter value of $\boldsymbol{\beta}$.

We need the following conditions.

C1: G is a continuous distribution with support $\mathbb{R}^{k-1}$ and $0 < G(\boldsymbol{u}) < 1$ for all $\boldsymbol{u} \in \mathbb{R}^{k-1}$.

C2: $E_{\boldsymbol{\beta}_0} |\log G(\boldsymbol{A}(\boldsymbol{X}_1, y_1)\boldsymbol{\beta})|$ is finite and is a continuous function of $\boldsymbol{\beta} \in \mathbb{R}^{m-1}$.

C3: There exists an integer n_* such that

$$P_{\boldsymbol{\beta}_0}((\boldsymbol{X}^{(n_*)}, \boldsymbol{y}^{(n_*)}) : coni(\mathcal{A}^{(n_*)}) = \mathbb{R}^m) > 0.$$

C4: The prior of $\boldsymbol{\beta}$ is proper and has a density π with respect to Lebesgue measure; π is continuous and positive at $\boldsymbol{\beta}_0$.

C5: Each coordinate of $\boldsymbol{X}_i$ has finite expectation under $\boldsymbol{\beta}_0$.

C6: $\nabla_i G$ and $\nabla^2_{ij} G$ are bounded for all i and j.

C7: $E_{\boldsymbol{\beta}_0} |\nabla^3_{ijk} G(\boldsymbol{A}(\boldsymbol{X}_1, y_1)\boldsymbol{\beta})(\boldsymbol{A}(\boldsymbol{X}_1, y_1))_{l,h}| < \infty$ for all i, j, k, l, h.

C8: The Fisher information, $I(\boldsymbol{\beta})$, based on $(\boldsymbol{X}_1, y_1)$, is nonsingular for all $\boldsymbol{\beta} \in \mathbb{R}^m$.

For any positive integer j and $D \subset \mathbb{R}^m$, define

$$Z_j(D) = \inf_{\boldsymbol{\beta} \in D} \log \prod_{i=1}^{j} \frac{G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta}_0)}{G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta})}.$$

Lemma 4 *Under Conditions C1 and C2, for any $\boldsymbol{\beta}$, there is a neighborhood $N(\boldsymbol{\beta})$ such that $E_{\boldsymbol{\beta}_0} Z_1(N(\boldsymbol{\beta})) > -\infty$.*

PROOF. Let $N(\boldsymbol{\beta}, 1/k) = \{\boldsymbol{\theta} \in \mathbb{R}^m : \|\boldsymbol{\theta} - \boldsymbol{\beta}\| < 1/k\}$. Since G is continuous,

$$\inf_{\boldsymbol{\theta} \in N(\boldsymbol{\beta}, 1/k)} (-\log G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\theta})) \uparrow -\log G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta}).$$

By the Monotone Convergence Theorem, we have

$$\lim_{k \to \infty} E_{\boldsymbol{\beta}_0} \inf_{\boldsymbol{\theta} \in N(\boldsymbol{\beta}, 1/k)} (-\log G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\theta})) = E_{\boldsymbol{\beta}_0}(-\log G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta})) < \infty.$$

Hence, for some k,

$$E_{\boldsymbol{\beta}_0} \inf_{\boldsymbol{\theta} \in N(\boldsymbol{\beta}, 1/k)} (-\log G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\theta})) < \infty.$$

The conclusion follows. □

Lemma 5 *Under C3, there exists a compact set $D \subset \mathbb{R}^m$ such that $E_{\boldsymbol{\beta}_0} Z_{n_*}(D^c) > 0$.*

PROOF. We will show there is a large positive number M with $\|\boldsymbol{\beta}_0\| < M < \infty$ and

$$E_{\boldsymbol{\beta}_0} \inf_{\|\beta\| > M} \sum_{i=1}^{n_*} \log \frac{G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta}_0)}{G(\boldsymbol{A}(\boldsymbol{X}_i, y_i)\boldsymbol{\beta})} > 0.$$

Let

$$\begin{aligned} B &= \{(\boldsymbol{X}^{(n_*)}, \boldsymbol{y}^{(n_*)}) : coni(\mathcal{A}^{(n_*)}) = \mathbb{R}^m\}; \\ B_\epsilon &= B \cap \{(\boldsymbol{X}^{(n_*)}, \boldsymbol{y}^{(n_*)}) : C(\boldsymbol{X}^{(n_*)}, \boldsymbol{y}^{(n_*)}) > \epsilon\}, \quad \epsilon > 0. \end{aligned}$$

Since $C(\boldsymbol{X}^{(n_*)},\boldsymbol{y}^{(n_*)})>0$ for all $(\boldsymbol{X}^{(n_*)},\boldsymbol{y}^{(n_*)})\in B$ and $P_{\boldsymbol{\beta}_0}(B)>0$ under C3, there exists $\epsilon>0$ with $P_{\boldsymbol{\beta}_0}(B_\epsilon)>0$. For all $(\boldsymbol{X}^{(n_*)},\boldsymbol{y}^{(n_*)})\in B_\epsilon$, by an argument similar to the one proving (3.2),

$$\begin{aligned}\sum_{i=1}^{n_*}\log G(A(\boldsymbol{X}_i,y_i)\boldsymbol{\beta}) &\le \log H(-C(\boldsymbol{X}^{(n_*)},\boldsymbol{y}^{(n_*)})\|\boldsymbol{\beta}\|)\\ &\le \log H(-\epsilon\|\boldsymbol{\beta}\|),\end{aligned}$$

where H is the cumulative distribution function of $\min_{l\neq j=1,\ldots,k}(\xi_{1l}-\xi_{1j})$. Then for any $M>0$,

$$\begin{aligned}&E_{\boldsymbol{\beta}_0}Z_{n_*}\{\|\boldsymbol{\beta}\|>M\}\\ &= E_{\boldsymbol{\beta}_0}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta}_0)-E_{\boldsymbol{\beta}_0}\sup_{\|\boldsymbol{\beta}\|>M}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta})\\ &\ge E_{\boldsymbol{\beta}_0}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta}_0)-E_{\boldsymbol{\beta}_0}I_{B_\epsilon}\sup_{\|\boldsymbol{\beta}\|>M}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta})\\ &\ge E_{\boldsymbol{\beta}_0}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta}_0)-E_{\boldsymbol{\beta}_0}I_{B_\epsilon}\sup_{\|\boldsymbol{\beta}\|>M}\log H(-\epsilon\|\boldsymbol{\beta}\|)\\ &\ge E_{\boldsymbol{\beta}_0}\sum_{i=1}^{n_*}\log G(\boldsymbol{A}(\boldsymbol{X}_i,y_i)\boldsymbol{\beta}_0)-P_{\boldsymbol{\beta}_0}(B_\epsilon)\log H(-\epsilon M).\end{aligned}$$

Since $\lim_{M\to\infty}H(-\epsilon M)=0$, we can choose a large enough M so that the above quantity is positive. □

C.2. Asymptotic Properties of the MLE. Let $\hat{\boldsymbol{\beta}}_n^M$ be the maximum likelihood estimator of $\boldsymbol{\beta}$ based on the observations $(\boldsymbol{X}^{(n)},\boldsymbol{y}^{(n)})$.

Theorem 5 *Under Conditions C1–C3, we have*

$$\hat{\boldsymbol{\beta}}_n^M\to\boldsymbol{\beta}_0 \text{ as } n\to\infty \text{ a.s. } P_{\boldsymbol{\beta}_0}.$$

Proof. Under C1–C3, the conclusions of Lemmas 4 and 5 hold. These, in turn, satisfy the assumptions of Lemma 7.54, Lemma 7.83 and Theorem 7.49 of Schervish (1995). Hence the MLE of $\boldsymbol{\beta}$ is consistent. □

Lemma 6 *Under C5 and C6,*

$$\begin{aligned}E_{\boldsymbol{\beta}}\Big(\frac{\partial}{\partial\beta_i}\log G(\boldsymbol{A}(\boldsymbol{X}_1,y_1)\boldsymbol{\beta})\Big) &= 0;\\ E_{\boldsymbol{\beta}}\Big[\Big(\frac{\partial^2}{\partial\beta_i\partial\beta_j}\log G(\boldsymbol{A}(\boldsymbol{X}_1,y_1)\boldsymbol{\beta})\Big)_{i,j=1,\ldots,m}\Big] &= -I(\boldsymbol{\beta}).\end{aligned}$$

Proof. C5 and C6 justify the interchange of differentiation and integration, and we obtain the conclusions. □

Lemma 7 *Under C7, there exists a function $H_r(\boldsymbol{X}_1,y_1,\boldsymbol{\beta})$ such that*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le r}\left|\frac{\partial^2}{\partial\beta_i\partial\beta_j}\log G(\boldsymbol{A}(\boldsymbol{X}_1,y_1)\boldsymbol{\beta}_0)-\frac{\partial^2}{\partial\beta_i\partial\beta_j}\log G(\boldsymbol{A}(\boldsymbol{X}_1,y_1)\boldsymbol{\beta})\right|\le H_r(\boldsymbol{X}_1,y_1,\boldsymbol{\beta}_0)$$

for all i and j, and

$$\lim_{r\to 0} E_{\boldsymbol{\beta}_0} H_r(\boldsymbol{X}_1, y_1, \boldsymbol{\beta}_0) = 0.$$

PROOF. Use the one term Taylor expansion. □

Theorem 6 *(Asymptotic Normality of the MLE). Under C1–C3 and C5–C8,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^M - \boldsymbol{\beta}_0) \to N(0, I(\boldsymbol{\beta}_0)^{-1}) \text{ in distribution as } n \to \infty.$$

PROOF. Theorem 5 guarantees that the MLE is consistent, and Lemmas 6 and 7 give the assumptions of Theorem 7.63 of Schervish (1995). □

L.3. Asymptotics of Posteriors under a Proper Prior. In this subsection, we assume a proper prior of $\boldsymbol{\beta}$ is used.

Theorem 7 *Under C1–C3, the posterior distribution of $\boldsymbol{\beta}$, $\pi_0(\cdot|\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)})$, satisfies*

$$\pi_0(U|\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)}) \to 1 \text{ a.s. } P_{\boldsymbol{\beta}_0}$$

for any open neighbor U of $\boldsymbol{\beta}_0$.

PROOF. The Kullback-Leibler number $I(\boldsymbol{\beta}_0, \boldsymbol{\beta})$ is continuous due to C2, and the prior puts positive mass on every open neighborhood of $\boldsymbol{\beta}_0$ by C4. These facts together with Lemmas 4 and 5 imply that the assumptions of Theorem 7.80 of Schervish (1995) are satisfied. □

Theorem 8 *Let g_n be the posterior density of $\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_n)$ and ϕ be the k-dimensional standard normal density, where $\boldsymbol{\Sigma}_n$ is the observed Fisher information matrix of $\boldsymbol{\beta}$ based on $(\boldsymbol{X}^{(n)}, \boldsymbol{y}^{(n)})$. Under C1–C8, for each compact set $D \subset \mathbb{R}^m$ and $\epsilon > 0$,*

$$\lim_{n\to\infty} P_{\boldsymbol{\beta}_0}(\sup_{\boldsymbol{u}\in D} |g_n(\boldsymbol{u}) - \phi(\boldsymbol{u})| > \epsilon) = 0.$$

PROOF. This follows from Theorem 7.89 and 7.102 of Schervish (1995). □

C.4. Asymptotics of Posteriors under an Improper Prior. We add the following condition.

- **C4′**: The improper prior, π_0, of $\boldsymbol{\beta}$ has a density with respect to Lebesgue measure and is continuous and positive at $\boldsymbol{\beta}_0$. Furthermore, with $P_{\boldsymbol{\beta}_0}$-probability 1, there exists an integer $N < \infty$, which depends on $(\boldsymbol{X}, \boldsymbol{y})$, such that

$$\int \prod_{i=1}^{N} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta}) < \infty.$$

Theorem 9 *With **C4** replaced by **C4′**, Theorems 7 and 8 still hold.*

PROOF. Define the stopping time

$$N = \inf\{n \geq 1 : \int \prod_{i=1}^{n} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta}) < \infty\}.$$

By **C4′**, $P_{\boldsymbol{\beta}_0}(N < \infty) = 1$. From the Strong Markov Property (cf. Billingsley, 1995), N is independent of the sequence $(\boldsymbol{X}_{N+n}, y_{N+n})_{n\geq 1}$. Clearly the asymptotic behavior of the sequence $(\boldsymbol{X}_n, y_n)_{n\geq 1}$ is the same as that of the sequence $(\boldsymbol{X}_{N+n}, y_{N+n})_{n\geq 1}$. Let

$$\pi_N(d\boldsymbol{\beta}) = \frac{\prod_{i=1}^{N} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta})}{\int \prod_{i=1}^{N} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta})}.$$

Then for any $B \subset \mathbb{R}^m$,

$$\begin{aligned} \pi(\boldsymbol{\beta} \in B|\boldsymbol{X}^{(N+n)}, \boldsymbol{y}^{(N+n)}) &= \frac{\int_B \prod_{i=1}^{N+n} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta})}{\int \prod_{i=1}^{N+n} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_0(d\boldsymbol{\beta})} \\ &= \frac{\int_B \prod_{i=N+1}^{N+n} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_N(d\boldsymbol{\beta})}{\int \prod_{i=N+1}^{N+n} f(\boldsymbol{X}_i, y_i|\boldsymbol{\beta})\pi_N(d\boldsymbol{\beta})}. \end{aligned}$$

Hence the posterior with a sample of size $N+n$ and improper prior π is the same as the posterior with another independent sample of size n and proper prior π_N. For any given $(\boldsymbol{X}^{(N)}, \boldsymbol{y}^{(N)})$, apply Theorem 7 and 8. This completes the proof. □

References

Billingsley, P. (1995). *Probability and Measure.* John Wiley & Sons, New York.

Panik, M. J. (1993). *Fundamentals of Convex Analysis, Duality, Separation Representation, and Resolution.* Kluwer Academic Publishers, Boston.

Schecter, M. (1971). *Principles of Functional Analysis.* Academic Press, New York.

Schervish, M. J. (1995). *Theory of Statistics.* Springer-Verlag, New York.