EMPIRICAL LIKELIHOOD FOR IRREGULARLY LOCATED SPATIAL DATA

Matthew Van Hala, Daniel J. Nordman, and Zhengyuan Zhu

Department of Statistics, Iowa State University

Supplementary Material

This supplementary appendix provides additional numerical summaries for the spatial blockwise empirical likelihood (SBEL) method as well as proofs of some distributional results from the main manuscript. In the following, Section S.1.1 contains a further table and figure supporting the simulation studies of Section 5 of the main manuscript, while Section S.1.2 has additional material regarding block selection for the data example given in Section 6 of the main manuscript. Proofs remaining from the manuscript are given in Section S.2 to follow. Section S.2.1 provides a proof of Lemma 2(d) from the appendix of the main manuscript, used in part to establish Theorem 1 there (regarding the chi-square limit of the log-SBEL ratio). In Section S.2.2, we establish Theorem 2 involving distributional results for the SBEL method based maximum EL estimation. Equation numbers (1)-(4) and (A.1)-(A.3) refer to the main manuscript, and any further equations are then subsequently enumerated in the following, as needed. A reference section at the end of this supplementary appendix contains citations appearing here.

S.1 Supporting Numerical Material

S.1.1 Additional Simulation Studies

Table 1 summarizes empirical coverage probabilities for additional simulation work, as described in Section 5.1 of the main manuscript, regarding confidence intervals for the SBEL and other interval methods. Figure 1 provides additional empirical power curves for SBEL goodness-of-fit tests of normality, as described in Section 5.2 of the main manuscript, for 24×24 sampling regions in the simulation design.

S.1.2 Additional Material for Data Example

This section describes the block b selection approach used for the SBEL method in Section 6 of the main manuscript for fitting a spatial regression model. That is, to choose a block size b, we used the "minimal volatility" technique of Politis, Romano and Wolf (1999, Sec. 9.3.2). While heuristic, this block selection

Table 1: Empirical coverage of 90% intervals for the mean over spatial dependence values $r = 1/3, 3$ and various methods: SBEL with chi-square calibration (ELC), SBEL with bootstrap-based calibration (ELB), normal approximation (Nor), and block bootstrap (Boot)

				Uniform Sites				Non-Uniform Sites			
				Method				Method			
	Points	Grid Size	b	ELC	ELB	Nor	Boot	ELC	ELB	Nor	Boot
$r=3$	$n=100$	12×12	2	43.4	50.3	38.0	39.5	31.6	38.4	31.7	33.6
			4	57.9	70.5	47.7	47.6	40.5	48.5	31.5	32.6
			6	64.8	81.3	46.8	46.8	41.0	55.0	23.0	23.5
		24×24	4	60.4	68.2	57.8	58.1	47.9	55.2	49.3	44.3
			6	72.0	79.4	64.8	65.3	52.7	62.2	45.0	45.6
			8	76.7	84.2	64.1	64.1	53.7	60.3	44.3	41.5
	$n=900$	12×12	2	34.4	41.6	32.3	33.4	29.3	35.7	26.0	26.1
			4	56.6	69.5	46.2	48.3	36.6	47.7	29.1	29.4
			6	61.6	79.2	42.7	42.2	38.9	54.7	24.3	20.8
		24×24	4	54.6	66.6	52.1	52.3	43.5	54.1	41.6	42.3
			6	65.3	78.5	59.1	59.5	48.8	59.6	43.2	38.9
			8	71.6	84.1	62.7	61.8	50.6	59.6	39.0	42.0
$r=\frac{1}{3}$	$n=100$	12×12	2	87.9	89.0	84.0	83.4	80.5	81.6	83.0	76.8
			4	91.7	91.2	81.6	83.5	78.3	76.6	70.2	67.4
			6	90.7	91.4	77.7	72.2	69.6	70.3	50.7	44.9
		24×24	4	91.2	91.8	89.7	85.8	87.5	86.0	83.8	84.0
			6	92.3	92.3	84.9	85.9	87.8	85.1	81.0	78.2
			8	92.8	91.0	85.9	82.8	85.3	79.6	75.0	69.7
	$n=900$	12×12	2	82.9	85.0	83.4	79.6	75.3	79.8	73.2	71.4
			4	89.6	91.2	80.4	83.3	74.2	74.9	65.4	60.7
			6	88.6	91.5	75.6	73.1	63.5	65.2	42.8	47.6
		24×24	4	90.3	91.3	87.5	86.2	84.1	84.2	79.8	79.6
			6	91.1	89.9	87.3	86.5	82.5	81.7	78.3	73.1
			8	94.2	93.4	83.6	84.0	81.7	77.3	67.4	64.3

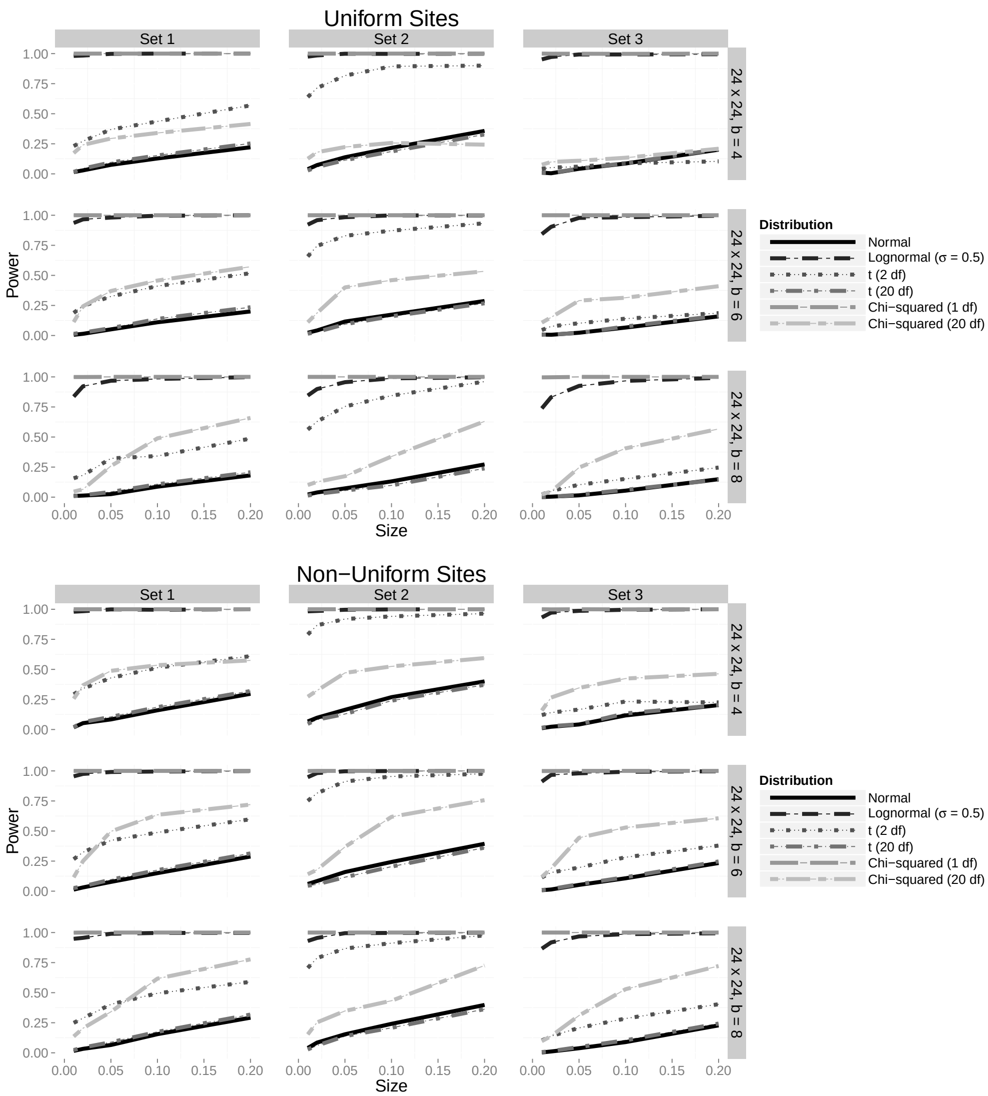

Figure 1: Empirical power functions for SBEL goodness-of-fit tests of normality using three sets of estimating functions and block sizes $b = 4$, 6, and 8 on a 24×24 region, sample size $n = 100$, and uniform and non-uniform locations; data are marginally normal, log-normal, t_2, t_{20}, χ^2_1, and χ^2_{20}.

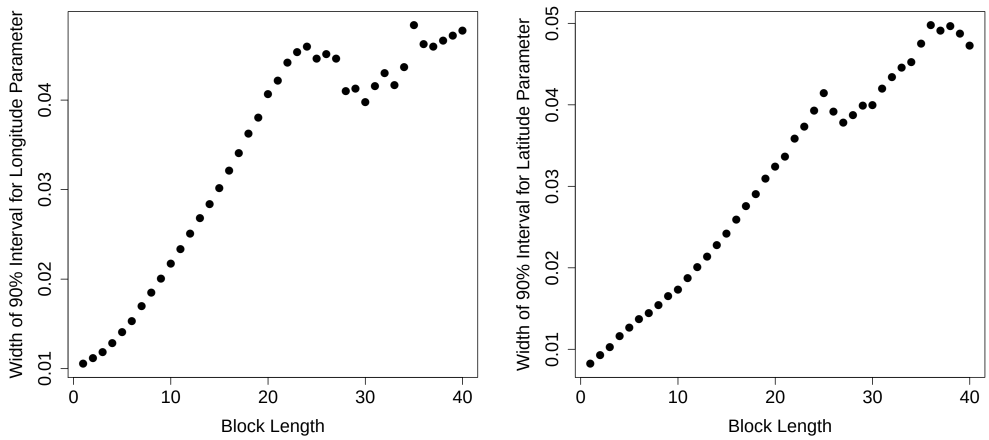

Figure 2: Lengths of 90% CIs by block size for longitude (left) and latitude (right) regression parameters

approach is motivated by the principle that approximately correct block sizes for inference may be characterized by confidence regions/intervals with stable behavior as a function of b. By creating SBEL intervals over a range of b, an adequate block length can be chosen by visual inspection. To illustrate, Figure 2 displays lengths of 90% confidence intervals (CIs) for the longitude and latitude regression parameters as a function of b, where CIs for individual parameters are found by profiling the SBEL log-ratio statistic as with parametric likelihood. These plots suggest a block size b of about 27 where CI lengths exhibit stability.

S.2 Supplementary Proofs

In the following, we use notation developed in the appendix of the main manuscript and the proof of Theorem 1 there.

S.2.1 Proof of Lemma 2(d)

We need to establish that, under Assumptions A1-A4, $P_{\cdot|\boldsymbol{X}}(R_n(\theta_0) > 0) \to 1$ as $n \to$ w.p.1 ($P_{\boldsymbol{X}}$). Note that $R_n(\theta_0) > 0$ if 0_r is interior to the convex hull of $\{A_n(\boldsymbol{i};\theta_0) : \boldsymbol{i} \in \mathcal{I}_n\}$ (cf. (A.1)), so that it suffices to show that the $P_{\cdot|\boldsymbol{X}}$-probability of this latter event converges to 1 (a.s. $P_{\boldsymbol{X}}$). For a given integer $\ell \geq 1$, consider an arbitrary integer vector $\boldsymbol{k} \in \mathbb{Z}^d$ such that $\|\boldsymbol{k}\|_1 \leq \ell$. By Theorem 3.2 of Lahiri (2003), under the mixing/moments assumptions and by the continuity/positivity of the probability density $f(\cdot)$ of $\boldsymbol{X}_1$, it holds that $b\boldsymbol{k} \in \mathcal{I}_n$ eventually and $(\lambda_n/b)^{d/2}A_n(b\boldsymbol{k};\theta_0) \xrightarrow{d} \tilde{Z} \sim N(0_r, \tilde{\Sigma}_\infty)$ a.s. $(P_{\boldsymbol{X}})$, where $\tilde{\Sigma}_\infty \equiv c\sigma(\mathbf{0})/f(\mathbf{0}) + \Sigma_0$ is positive definite. Hence, for a given $\ell \geq 1$,

$$\Delta_{n,\ell} \equiv \max_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} \sup_{a\in\mathcal{S}} \sup_{y\in\mathbb{R}} \left| P_{\cdot|\boldsymbol{X}}\left((\lambda_n/b)^{d/2}a'A_n(b\boldsymbol{k};\theta_0) \leq y\right) - P(a'\tilde{Z} \leq y)\right| \to 0$$

holds a.s. $(P_{\boldsymbol{X}})$ by Poyla's theorem on half-planes, where $\mathcal{S} = \{a \in \mathbb{R}^r : \|a\| = 1\}$ is the $\mathbb{R}^d$ unit sphere. By the above, for a given $\epsilon > 0$, one may choose $\epsilon_1 > 0$ to make the averages

$$\Delta_{1n,\ell} \equiv \sup_{a\in\mathcal{S}} \frac{1}{L} \sum_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} P_{\cdot|\boldsymbol{X}}\left((\lambda_n/b)^{d/2}|a'A_n(b\boldsymbol{k};\theta_0)| < \epsilon_1\right) < \epsilon$$

and

$$\Delta_{2n,\ell} \equiv \sup_{a\in\mathcal{S}} \frac{1}{L} \sum_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} P_{\cdot|\boldsymbol{X}}\left((\lambda_n/b)^{d/2}\|A_n(b\boldsymbol{k};\theta_0)\| \geq \epsilon_1^{-1}\right) < \epsilon$$

eventually (a.s. $P_{\boldsymbol{X}}$), where $L \equiv (2\ell+1)^d = |\{\boldsymbol{k} \in \mathbb{Z}^d : \|\boldsymbol{k}\|_1 \leq \ell\}|$. Next define an empirical distribution $\hat{F}_{n,\ell}(a) = L^{-1}\sum_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} \mathbb{I}[(\lambda_n/b)^{d/2}a'A_n(b\boldsymbol{k};\theta_0) < 0]$ for $a \in \mathcal{S}$. As $\mathcal{S}$ is compact, this can be covered with a finite collection of open balls of radius ϵ_1^2 around points $a_1, \ldots, a_t \in \mathcal{S}$ (where t depends on ϵ_1). For $a \in \mathcal{S}$, there exists a_i such that $\|a_i - a\| < \epsilon_1^2$ so that

$$|\hat{F}_{n,\ell}(a) - 1/2| \leq |\hat{F}_{n,\ell}(a) - \hat{F}_{n,\ell}(a_i)| + |\hat{F}_{n,\ell}(a_i) - 1/2| \leq T_{1n,\ell}^{(i)} + T_{2n,\ell} + |\hat{F}_{n,\ell}(a_i) - 1/2|$$

using bounds on indicator functions, where

$$\begin{aligned} T_{1n,\ell}^{(i)} &= L^{-1} \sum_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} \mathbb{I}[(\lambda_n/b)^{d/2}|a_i'A_n(b\boldsymbol{k};\theta_0)| < \epsilon_1],\ i = 1, \ldots, t, \\ T_{2n,\ell} &= L^{-1} \sum_{\boldsymbol{k}\in\mathbb{Z}^d, \|\boldsymbol{k}\|_1\leq\ell} \mathbb{I}[(\lambda_n/b)^{d/2}\|A_n(b\boldsymbol{k};\theta_0)\| \geq \epsilon_1^{-1}]. \end{aligned}$$

Hence, for $T_{3n,\ell} = \sum_{i=1}^t |\hat{F}_{n,\ell}(a_i) - \mathrm{E}_{\cdot|\boldsymbol{X}}\{\hat{F}_{n,\ell}(a_i)\}|$ and $T_{4n,\ell} = \sum_{i=1}^t |T_{1n,\ell}^{(i)} - \mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{1n,\ell}^{(i)}\}|$, we have

$$\mathrm{E}_{\cdot|\boldsymbol{X}}\left\{\sup_{a\in\mathcal{S}} |\hat{F}_{n,\ell}(a) - 1/2|\right\} \leq \sum_{j=2}^4 \mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{jn,\ell}\} + \max_{1\leq i\leq m} \mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{1n,\ell}^{(i)}\} + \Delta_{n,\ell}.$$

Note that $\max_{1\leq i\leq m} \mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{1n,\ell}^{(i)}\} \leq \Delta_{1n,\ell} < \epsilon$, $\mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{2n,\ell}\} \leq \Delta_{2n,\ell} < \epsilon$ and $\Delta_{n,\ell} < \epsilon$ eventually for a given choice of ℓ (a.s. $P_{\boldsymbol{X}}$). Also, it holds by Jensen's inequality that

$$\begin{aligned}\mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{3n,\ell}\} &\leq \sum_{i=1}^{t}\left[\mathrm{E}_{\cdot|\boldsymbol{X}}|\{\hat{F}_{n,\ell}(a_i)\} - \mathrm{E}_{\cdot|\boldsymbol{X}}\{\hat{F}_{n,\ell}(a_i)\}|^2\right]^{1/2} \\ &\leq \left(1+4\sum_{k=1}^{\infty}k^{-\tau_1}\right)^{1/2} L^{-1/2}t\end{aligned}$$

using the standard covariance bound

$$\begin{aligned}&|\mathrm{Cov}_{\cdot|\boldsymbol{X}}\{\mathbb{I}[(\lambda_n/b)^{d/2}a_i' A_n(b\boldsymbol{k}_1;\theta_0)<0], \mathbb{I}[(\lambda_n/b)^{d/2}a_i' A_n(b\boldsymbol{k}_2;\theta_0)<0]\}| \\ \leq\ & 4\alpha(d[b,\boldsymbol{k}_1,\boldsymbol{k}_2],b^d), \quad d[b,\boldsymbol{k}_1,\boldsymbol{k}_2]\equiv \inf\{\|\boldsymbol{x}_1-\boldsymbol{x}_2\| : \boldsymbol{x}_i\in B_n(b\boldsymbol{k}_i), i=1,2\}\end{aligned}$$

for bounded random variables (cf. Athreya and Lahiri (2006, Corollary 16.2.4(ii))) to show

$$\mathrm{E}_{\cdot|\boldsymbol{X}}|\{\hat{F}_{n,\ell}(a_i)\} - \mathrm{E}_{\cdot|\boldsymbol{X}}\{\hat{F}_{n,\ell}(a_i)\}|^2 \leq \frac{1}{L}\left(1+\sum_{k=1}^{\ell}4\alpha(kb;b)\right) \leq \frac{1}{L}\left(1+4\sum_{k=1}^{\infty}k^{-\tau_1}\right)$$

a.s. ($P_{\boldsymbol{X}}$) under Assumption A1; likewise, $\mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{4n,\ell}\} \leq (1+4\sum_{k=1}^{\infty}k^{-\tau_1})^{1/2} L^{-1/2}t$. Since t depends on ϵ_1 and ℓ is arbitrary, we may choose ℓ (i.e. $L=(2\ell+1)^d$) so that $\mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{3n,\ell}\} + \mathrm{E}_{\cdot|\boldsymbol{X}}\{T_{4n,\ell}\} < \epsilon$ (a.s. ($P_{\boldsymbol{X}}$)). Hence, we may pick ℓ and $\epsilon \in (0,1/16)$ so that $P_{\cdot|\boldsymbol{X}}(\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a)-1/2|>1/4) \leq 16\epsilon$ holds for all large n (a.s. ($P_{\boldsymbol{X}}$)). Note that the event $\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a)-1/2|\leq 1/4$ implies $\inf_{a\in\mathcal{S}}\hat{F}_{n,\ell}(a)\geq 1/4$ which further implies that 0_r lies in the interior of the convex hull of $\{A_n(\boldsymbol{i};\theta_0) : \boldsymbol{i}\in\mathcal{I}_n\}$; this last event implies $R_n(\theta_0)>0$ and hence $P_{\cdot|\boldsymbol{X}}(R_n(\theta_0)>0)\geq 1-16\epsilon$ holds for any arbitrary $\epsilon>0$ (a.s. ($P_{\boldsymbol{X}}$)). (Note if 0_r is not in the interior as claimed, then there exists $a_*\in\mathcal{S}$ such that $a_*' A_n(\boldsymbol{i};\theta_0)\geq 0$ holds for all $\boldsymbol{i}\in\mathcal{I}_n$ by the separating/supporting hyperplane theorem; however, $\hat{F}_{n,\ell}(a_*)\geq 1/4$ entails that there exists a $\boldsymbol{k}\in\mathbb{Z}^d$, $\|\boldsymbol{k}\|_1\leq\ell$, such that $a_*' A_n(b\boldsymbol{k};\theta_0)<0$ and $b\boldsymbol{k}\in\mathcal{I}_n$, which is a contradiction.) □

S.2.2 Proof of Theorem 2 on Maximum EL Estimation

To establish Theorem 2, we first require a preliminary result in Lemma 3. Define $D_n(\theta_0)=\sum_{\boldsymbol{i}\in\mathcal{I}_n}\lambda_n^{-d/2}\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta$ for

$$\lambda_n^{-d/2}\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta = n^{-1}b^{-d}\sum_{j=1}^{n}\partial g(Z(\boldsymbol{s}_j);\theta_0)/\partial\theta\mathbb{I}(\boldsymbol{s}_j\in B_n(\boldsymbol{i})), \quad \boldsymbol{i}\in\mathcal{I}_n,$$

and let $D(\theta_0)=\mathrm{E}\{\partial g(Z(\boldsymbol{0});\theta_0)/\partial\theta\}$.

Lemma 3 *Under the assumptions of Theorem 2, the following hold a.s.* ($P_{\boldsymbol{X}}$)*:*

(a) $D_n(\theta_0) \to D(\theta_0)$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability;*

(b) $(\log n)^2 b^d \lambda_n^{-d/2} \max_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\| \to 0$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability;*

(c) $\lambda_n^{-d/2}\mathrm{E}_{\cdot|\boldsymbol{X}}\left\{\sum_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|\right\} = O(1)$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability.*

(d) $\max_{\boldsymbol{i}\in\mathcal{I}_n} |\{\boldsymbol{s}_j : 1 \le j \le n, \boldsymbol{s}_j \in B_n(\boldsymbol{i})\}| = O(b^d m_n)$ *in* $P_{\cdot|\boldsymbol{X}}$*-probability.*

Proof of Lemma 3. Lemma 3(a) follows by showing

$$\mathrm{E}_{\cdot|\boldsymbol{X}}\{\|D_n(\theta_0) - D(\theta_0)\|\} \to 0 \quad \text{a.s. } (P_{\boldsymbol{X}}),$$

Note that by Lemma 1(i) with $k = 1$ (under the mixing/moment assumptions) and Jensen's inequality

$$\begin{aligned}
& \mathrm{E}_{\cdot|\boldsymbol{X}}\{\|D_n(\theta_0) - \mathrm{E}_{\cdot|\boldsymbol{X}} D_n(\theta_0)\|\} \\
\le & \left(|\mathcal{I}_n| \sum_{\boldsymbol{i}\in\mathcal{I}_n} \lambda_n^{-d} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta - \mathrm{E}_{\cdot|\boldsymbol{X}}\{\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|^2\}\right)^{1/2} \\
= & \ O(\lambda_n^{d/2} n^{-1} b^{-d/2} m_n) = o(1)
\end{aligned}$$

(a.s. ($P_{\boldsymbol{X}}$)) and (similarly to the proof of Lemma 2(a)) $\|\mathrm{E}_{\cdot|\boldsymbol{X}}\{D_n(\theta_0)\} - D(\theta_0)\| = O(b/\lambda_n) = o(1)$; this establishes the result.

Lemma 3(d) follows from the fact that

$$P_{\boldsymbol{X}}\left(\max_{\boldsymbol{j}\in\mathbb{Z}^d, \boldsymbol{i}\in\mathcal{I}_n} \sum_{i=1}^{n} \mathbb{I}(\lambda_n \boldsymbol{X}_i \in \{\boldsymbol{j} + (0,1]^d\} \cap B_n(\boldsymbol{i})) > C m_n \text{ infinitely often}\right) = 0 \tag{S.1}$$

holds for some $C > 0$ (cf. Lahiri and Zhu (2006, p. 1809)).

To show Lemma 3(b), we bound

$$\mathrm{E}_{\cdot|\boldsymbol{X}}\left\{\max_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|\right\} \le e_{1n} + e_{2n}$$

where $e_{1n} \equiv \max_{\boldsymbol{i}\in\mathcal{I}_n} \|\mathrm{E}_{\cdot|\boldsymbol{X}}\{\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\}\| = O(\lambda_n^{d/2} b^{-d} n^{-1} b^d m_n)$ a.s. ($P_{\boldsymbol{X}}$) by Lemma 3(d) and, by Jensen's inequality and Lemma 1(i) (with $k = 1$),

$$\begin{aligned}
e_{2n} \ \equiv & \left(\sum_{\boldsymbol{i}\in\mathcal{I}_n} \mathrm{E}_{\cdot|\boldsymbol{X}}\left\{\|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta - \mathrm{E}_{\cdot|\boldsymbol{X}}\{\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\}\|^2\right\}\right)^{1/2} \\
= & \ O\left[\lambda_n^{d/2} n^{-1} b^{-d} (\lambda_n^d m_n^2 b^d)^{1/2}\right]
\end{aligned}$$

a.s. $(P_{\boldsymbol{X}})$. Hence, $(\log n)^2 b^d \lambda_n^{-d/2} \mathrm{E}_{\cdot|\boldsymbol{X}}\{\max_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|\} = O((\log n)^2 \lambda_n^{d/2} b^{d/2} n^{-1} m_n) = o(1)$ a.s. $(P_{\boldsymbol{X}})$.

For Lemma 3(c), note $\lambda_n^{d/2} \sum_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\| \le n^{-1} \sum_{i=1}^n \|\partial g(Z(\boldsymbol{s}_i);\theta_0)/\partial\theta\|$ so that $\lambda_n^{d/2} \mathrm{E}_{\cdot|\boldsymbol{X}}\{\| \sum_{\boldsymbol{i}\in\mathcal{I}_n} \partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|\} \le \mathrm{E}_{\cdot|\boldsymbol{X}}\{\|\partial g(Z(\boldsymbol{0});\theta_0)/\partial\theta\|\}$ a.s. $(P_{\boldsymbol{X}})$. □

Proof of Theorem 2. As in the proof of Theorem 1, there exists $A \in \mathcal{F}$ with $P(A) = 1$, on the common probability space $(\Omega, \mathcal{F}, P)$, such that all events in Lemma 2(a)-(d) and Lemma 3 hold simultaneously conditioned on $\boldsymbol{X}_1 \equiv \boldsymbol{X}_1(\omega), \boldsymbol{X}_2 \equiv \boldsymbol{X}_2(\omega), \ldots$ for any $\omega \in A$. For simplicity, we again fix $\omega \in A$ throughout the proof and consider distributional convergence conditioned on a given sequence $\{\boldsymbol{X}_n(\omega)\}$; then $P_{\cdot|\boldsymbol{X}}$ is the only probability measure needed in the proof and we let $o_p(\cdot)$ and $O_p(\cdot)$ denote probabilistic order notation as usual in $P_{\cdot|\boldsymbol{X}}$-probability.

Set $\Theta_n = \{\theta \in \Theta : \lambda_n^{d/2}\|\theta - \theta_0\| \le \log n\}$, $\partial\Theta_n = \{\theta \in \Theta : \lambda_n^{d/2}\|\theta - \theta_0\| = \log n\}$, and $\nu_\theta = \max\{1, \lambda_n\|\theta - \theta_0\|\}$. For $A_n(\boldsymbol{i};\theta)$, $\boldsymbol{i} \in \mathcal{I}_n$, in (A.1), recall $A_n(\theta) = \sum_{\boldsymbol{i}\in\mathcal{I}_n} A_n(\boldsymbol{i};\theta)$ and define $\hat{\Sigma}_n(\theta) = b^d \sum_{\boldsymbol{i}\in\mathcal{I}_n} A_n(\boldsymbol{i};\theta)A_n(\boldsymbol{i};\theta)'$ and $Z_n(\theta) = \max_{\boldsymbol{i}\in\mathcal{I}_n} \|A_n(\boldsymbol{i};\theta)\|$ for $\theta \in \Theta$. We collect some preliminary results in (S.2), (S.3) and (S.4) below. Note that $\nu_\theta \le \log n$ and $\|\theta - \theta_0\| \le (\log n)\lambda_n^{-d/2}$ for $\theta \in \Theta_n$. Hence,

$$\sup_{\theta\in\Theta_n} \nu_\theta b^d Z_n(\theta) \le \sup_{\theta\in\Theta_n} (\log n) b^d Z_n(\theta) = o_p(1) \tag{S.2}$$

follows using the differentiability of $g(\cdot;\theta)$ in θ along with the Lipschitz condition (with parameter $\gamma > 0$) on $\partial g(\cdot;\theta)/\partial\theta$ in θ to show

$$\begin{aligned}\sup_{\theta\in\Theta_n} (\log n) b^d Z_n(\theta) &\le (\log n) b^d Z_n(\theta_0) + (\log n)^2 b^d \lambda_n^{-d/2} \max_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\| \\ &\quad + O((\log n) b^d \lambda_n^{d/2} n^{-1} b^{-d} b^d m_n (\log n \lambda_n^{-d/2})^{1+\gamma}) = o_p(1)\end{aligned}$$

using Lemma 2(b), Lemma 3(b) and (d) with the Lipschitz condition. Also, for $J_n(\theta_0) = \lambda_n^{-d/2} \sum_{\boldsymbol{i}\in\mathcal{I}_n} \|\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta\|$, it holds that $J_n(\theta_0) = O_p(1)$ by Lemma 3(c) so that

$$\begin{aligned}\sup_{\theta\in\Theta_n} \|A_n(\theta)\|/\nu_\theta &\le \|A_n(\theta_0)\| + J_n(\theta_0) \sup_{\theta\in\Theta_n} \nu_\theta^{-1} \lambda_n^{d/2} \|\theta - \theta_0\| \\ &\quad + O(|\mathcal{I}_n| \lambda_n^{d/2} b^{-d} n^{-1} b^d m_n [\lambda_n^{-d/2} \log n]^{1+\gamma}) \\ &= O_p(1) \end{aligned} \tag{S.3}$$

by Lemma 2(a) and Lemma 3(d). Finally, note that

$$\sup_{\theta\in\Theta_n}\|\hat{\Sigma}_n(\theta)-\Sigma_\infty\| \le \sup_{\theta\in\Theta_n}\|\hat{\Sigma}_n(\theta_0)-\Sigma_\infty\|+\sup_{\theta\in\Theta_n}\|\hat{\Sigma}_n(\theta)-\hat{\Sigma}_n(\theta_0)\| = o_p(1) \quad \text{(S.4)}$$

by Lemma 2(c) along with

$$\begin{aligned}
&\sup_{\theta\in\Theta_n}\|\hat{\Sigma}_n(\theta)-\hat{\Sigma}_n(\theta_0)\| \\
\le\ & b^d \sup_{\theta\in\Theta_n} Z_n(\theta) \sup_{\theta\in\Theta_n} \sum_{\boldsymbol{i}\in\mathcal{I}_n} \|A_n(\boldsymbol{i};\theta)-A_n(\boldsymbol{i};\theta_0)\| \\
\le\ & o_p((\log n)^{-1}) \sup_{\theta\in\Theta_n}\left[J_n(\theta_0)\lambda_n^{d/2}\|\theta-\theta_0\| + \lambda_n^{d/2}n^{-1}b^{-d}\lambda_n^d b^d m_n\|\theta-\theta_0\|^{1+\gamma}\right] \\
=\ & o_p((\log n)^{-1})[O_p(\log n)+o(1)] = o_p(1)
\end{aligned}$$

by (S.2), $J_n(\theta_0)=O_p(1)$ and Lemma 3(d) with the Lipschitz condition.

We next show the log-EL ratio $\ell_n(\theta) = -2b^{-d}\log R_n(\theta)$ exists finitely on Θ_n and is continuously differentiable. This implies a sequence of minimums $\hat{\theta}_n$ of $\ell_n(\theta)$ exists on Θ_n (i.e., $\hat{\theta}_n$ is a maximizer of $R_n(\theta)$) and we show additionally that $\hat{\theta}_n \notin \partial\Theta_n$ and $\partial\ell_n(\theta)/\partial\theta = 0_p$ at $\theta=\hat{\theta}_n$. Define functions

$$Q_{1n}(\theta,t) = \sum_{\boldsymbol{i}\in\mathcal{I}_n}\frac{A_n(\boldsymbol{i};\theta)}{1+t'A_n(\boldsymbol{i};\theta)}, \qquad Q_{2n}(\theta,t) = b^{-d}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\frac{\left(\partial A_n(\boldsymbol{i};\theta)/\partial\theta\right)'t}{1+t'A_n(\boldsymbol{i};\theta)}, \quad \text{(S.5)}$$

on $\Theta\times\mathbb{R}^r$. It can be shown that

$$P_{\cdot|\boldsymbol{X}}\left(R_n(\theta)>0 \text{ holds for any } \theta\in\Theta_n\right)\to 1. \quad \text{(S.6)}$$

(To see this, we modify the proof of Lemma 2(d). For a given integer $\ell\ge 1$, define $\hat{F}_{n,\ell}(a,\theta) \equiv L^{-1}\sum_{\boldsymbol{k}\in\mathbb{Z}^d,\|\boldsymbol{k}\|_1\le\ell}\mathbb{I}[(\lambda_n/b)^{d/2}a'A_n(b\boldsymbol{k};\theta)<0]$ for $a\in\mathcal{S}=\{u\in\mathbb{R}^r:\|u\|=1\}$, $\theta\in\Theta$ and $L=(2\ell+1)^d$. In the notation of the proof of Lemma 2(d), note that $\hat{F}_{n,\ell}(a)=\hat{F}_{n,\ell}(a,\theta_0)$ there and, for a given $\epsilon>0$, we may pick ℓ so that $P_{\cdot|\boldsymbol{X}}(\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a,\theta_0)-1/2|>1/12)<\epsilon$ for all large n. Additionally, by the same argument, we can choose an $\epsilon_2>0$ and large ℓ so that, for $W_{n,\ell,\epsilon_2}\equiv\sup_{a\in\mathcal{S}}L^{-1}\sum_{\boldsymbol{k}\in\mathbb{Z}^d,\|\boldsymbol{k}\|_1\le\ell}\mathbb{I}[(\lambda_n/b)^{d/2}|a'A_n(b\boldsymbol{k};\theta)|<\epsilon_2]$, it holds that $P_{\cdot|\boldsymbol{X}}(W_{n,\ell,\epsilon_2}>1/12)<\epsilon$ for all large n. Finally, it holds by the differentiability of the estimating function in θ that, for a given ℓ,

$$\begin{aligned}
Z_{n,\ell} &\equiv \sup_{\theta\in\Theta_n}(\lambda_n/b)^{1/2}\max\{\|A_n(b\boldsymbol{k};\theta)-A_n(b\boldsymbol{k};\theta_0)\|:\boldsymbol{k}\in\mathbb{Z}^d,\|\boldsymbol{k}\|_1\le\ell\} \\
&= O_p(\lambda_n^{d/2}b^{-d/2}\lambda_n^{d/2}n^{-1}b^{-d}[b^d m_n\lambda_n^{-d/2}\log n]) = o_p(1)
\end{aligned}$$

so that $P_{\cdot|\boldsymbol{X}}(Z_{n,\ell} > \epsilon_2/12) < \epsilon$ for all large n. Then, bounding the difference of summed indicator functions as

$$\sup_{\theta\in\Theta_n}\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a,\theta)-1/2| \le \sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a,\theta_0)-1/2| + W_{n,\ell,\epsilon_2} + \epsilon_2^{-1}Z_{n,\ell},$$

we can pick ℓ so that $P_{\cdot|\boldsymbol{X}}(\sup_{\theta\in\Theta_n}\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a,\theta)-1/2| > 1/4) < 3\epsilon$ for all large n. When $\sup_{\theta\in\Theta_n}\sup_{a\in\mathcal{S}}|\hat{F}_{n,\ell}(a,\theta)-1/2| \le 1/4$ holds, then it follows that $R_n(\theta) > 0$ for each $\theta\in\Theta_n$ by the supporting/separating hyperplane theorem as in the proof of Lemma 2(d). Hence, $P_{\cdot|\boldsymbol{X}}(R_n(\theta) > 0, \text{ any } \theta\in\Theta_n) \ge 1-3\epsilon$ for any arbitrary ϵ.)

As in the proof of Lemma 2(d), when the event in the probably statement (S.6) holds, then for any $\theta\in\Theta_n$, we may write $R_n(\theta) = \prod_{\boldsymbol{i}\in\mathcal{I}_n}(1+\gamma_{\theta,\boldsymbol{i}})^{-1} > 0$ where $\gamma_{\boldsymbol{i},\theta} = t'_{n,\theta}A_n(\boldsymbol{i};\theta)$ for a Lagrange multiplier $\tilde{t}_{n,\theta}\in\mathbb{R}^d$ such that $Q_{1n}(\theta,t_{n,\theta}) = 0_r$; the relationship between $\tilde{t}_{n,\theta}$ and the Lagrange multiplier $t_{n,\theta}$ defining (3.3) is $\tilde{t}_{n,\theta} = \lambda_n^{-d/2}nt_{n,\theta}$. Hence, by the positive definiteness of Σ_∞ under Assumption A4 and (S.4), $\hat{\Sigma}_n(\theta)$ is positive definite and $\hat{\Sigma}_n(\theta)^{-1}$ exists uniformly in $\theta\in\Theta_n$; this also implies that for each fixed $\theta\in\Theta_n$, $\partial Q_{1n}(\theta,t)/\partial t$ is negative definitive for $t\in\{u\in\mathbb{R}^r : 1+u'A_n(\boldsymbol{i};\theta)\ge 1/|\mathcal{I}_n|, \boldsymbol{i}\in\mathcal{I}_n\}$ so that, by implicit function theorem using $Q_{1n}(\theta,\tilde{t}_{n,\theta}) = 0_r$, $\tilde{t}_{n,\theta}$ is a continuously differentiable function of θ on Θ_n and the function $\ell_n(\theta) = -2b^{-d}\log R_n(\theta)$ is as well (e.g., Qin and Lawless (1994, p. 304-305)). Hence, with large probability as $n\to\infty$, the minimizer of $\ell_n(\theta)$ exists on Θ_n.

Now expanding $Q_{1n}(\theta,\tilde{t}_{n,\theta}) = 0_r$ for $\theta\in\Theta_n$, we can repeat the same essential argument in the proof of Theorem 1 based on (A.2) to find

$$\nu_\theta^{-1}\|A_n(\theta)\| \ge \frac{\|\nu_\theta^{-1}b^{-d}\tilde{t}_{n,\theta}\|v'_{n,\theta}\hat{\Sigma}_n(\theta)v_{n,\theta}}{1+[\nu_\theta b_n^d Z_n(\theta)]\|\nu_\theta^{-1}b^{-d}\tilde{t}_{n,\theta}\|},$$

for $\tilde{t}_{n,\theta} = \|\tilde{t}_{n,\theta}\|v_{n,\theta}$, $v_{n,\theta}\in\mathbb{R}^r$, $\|v_{n,\theta}\| = 1$; from this and (S.2)-(S.4), we have $\sup_{\theta\in\Theta_n}\nu_\theta^{-1}b^{-d}\|\tilde{t}_{n,\theta}\| = O_p(1)$. Then, analogously to the proof of Theorem 1 again, we may expand $Q_{1n}(\theta,\tilde{t}_{n,\theta}) = 0_r$ to yield $b^{-d}\tilde{t}_{n,\theta} = \hat{\Sigma}_n(\theta)^{-1}[A_n(\theta)+\beta_n(\theta)]$ for $\theta\in\Theta_n$ where

$$\sup_{\theta\in\Theta_n}\nu_\theta^{-1}\|\beta_n(\theta)\| \le \sup_{\theta\in\Theta_n}\nu_\theta^{-1}Z_n(\theta)\|\tilde{t}_{n,\theta}\|^2 b^{-d}\text{trace}[\hat{\Sigma}_n(\theta)]/(1-\|\tilde{t}_{n,\theta}\|Z_n(\theta)) = o_p(1).$$

Using now these orders of $\|\beta_n(\theta)\|, \|\tilde{t}_{n,\theta}\|$ and $Z_n(\theta)$ with arguments as in the proof of Theorem 1, we may then expand $\ell_n(\theta)$ uniformly in $\theta\in\Theta_n$ as

$$\begin{aligned}
&\sup_{\theta\in\Theta_n}\nu_\theta^{-2}|\ell_n(\theta) - A_n(\theta)'\hat{\Sigma}_n(\theta)^{-1}A_n(\theta)| \\
\le\ & O_p\left(\sup_{\theta\in\Theta_n}\nu_\theta^{-2}\left[\beta_n(\theta)'\hat{\Sigma}_n(\theta)\beta_n(\theta) + \frac{2b^{-2d}Z_n(\theta)\|\tilde{t}_{n,\theta}\|^3\text{trace}[\hat{\Sigma}_n(\theta)]}{(1-Z_\theta\|\tilde{t}_\theta\|)^3}\right]\right) = o_p(1),
\end{aligned}$$

so that, using (S.3)-(S.4),

$$\sup_{\theta\in\Theta_n} \nu_\theta^{-2}|\ell_n(\theta) - A_n(\theta)'\Sigma_\infty^{-1}A_n(\theta)| = o_p(1)$$

follows. For each $\theta \in \Theta_n$, we write $A_n(\theta) = A_n(\theta_0) + D_n(\theta_0)\lambda_n^{d/2}(\theta-\theta_0) + E_n(\theta)$ for $D_n(\theta_0) = \sum_{\boldsymbol{i}\in\mathcal{I}_n}\lambda_n^{-d/2}\partial A_n(\boldsymbol{i};\theta_0)/\partial\theta$ and a remainder $E_n(\theta)$ satisfying $\sup_{\theta\in\Theta_n}\|E_n(\theta)\| = O_p(\lambda_n^{d/2}b^{-d}n^{-1}\lambda_n^d b^d m_n(\log\lambda_n^{-d/2})^{1+\gamma}) = o_p(1)$. By Lemma 3(a), $D_n(\theta_0) = D(\theta_0) + o_p(1)$ so that, by (S.3), it now follows that

$$\begin{aligned}
&\sup_{\theta\in\Theta_n} \nu_\theta^{-2}\Big|\ell_n(\theta) - \\
&\qquad\big[A_n(\theta_0) + D(\theta_0)\lambda_n^{d/2}(\theta-\theta_0)\big]'\Sigma_\infty^{-1}\big[A_n(\theta_0) + D(\theta_0)\lambda_n^{d/2}(\theta-\theta_0)\big]\Big| \\
= \;& o_p(1). \qquad\qquad (S.7)
\end{aligned}$$

For $\theta = u\lambda_n^{-d/2}\log n + \theta_0 \in \partial\Theta_n$, with some $u\in\mathbb{R}^r$, $\|u\|=1$, we have $\nu_\theta = \log n$ so that from (S.7) we find that $\ell_n(\theta) \geq (\sigma^*/2)(\log n)^2$ holds uniformly in $\theta\in\partial\Theta_n$ with arbitrarily large probability when n is large, where σ^* denotes the smallest eigenvalue of positive definite $D(\theta_0)'\Sigma_\infty^{-1}D(\theta_0)$. At the same time, by Theorem 1, we have $\ell_n(\theta_0) = O_p(1)$. Hence, with probability approaching 1, the minimum $\hat\theta_n$ of $\ell_n(\theta)$ on Θ_n cannot be an element of $\partial\Theta_n$. Hence, $\hat\theta_n$ must satisfy $\hat\theta_n \in \Theta_n\backslash\partial\Theta_n$ and $0_r = Q_{1n}(\hat\theta_n, \tilde t_{n,\hat\theta_n})$ in addition to

$$0_p = 2^{-1}\partial\ell_n(\theta)/\partial\theta|_{\theta=\hat\theta_n} = Q_{2n}(\hat\theta_n, \tilde t_{n,\hat\theta_n})$$

by the differentiability of $\ell_n(\theta)$, for $Q_{1n}(\cdot,\cdot), Q_{2n}(\cdot,\cdot)$ from (S.5).

We next establish the asymptotic normality of $\hat\theta_n$ in Theorem 2(i). From the above arguments, we may solve $Q_{1n}(\hat\theta_n, \tilde t_{n,\hat\theta_n}) = 0_r$ for $b^{-d}\tilde t_{n,\hat\theta_n} = \hat\Sigma_{\hat\theta_n}^{-1}[A_n(\hat\theta_n) + \beta_n(\hat\theta_n)]$ or

$$\begin{aligned}
b^{-d}\tilde t_{n,\hat\theta_n} &= \hat\Sigma_n(\hat\theta_n)^{-1}[A_n(\hat\theta_n) + \beta_n(\hat\theta_n)] \qquad (S.8)\\
&= \Sigma_\infty^{-1}\big[A_n(\theta_0) + D(\theta_0)\lambda_n^{d/2}(\hat\theta_n - \theta_0)\big] + o_p(\nu_{\hat\theta_n}).
\end{aligned}$$

From $\lambda_n^{-d/2}Q_{2n}(\hat\theta_n, \tilde t_{n,\hat\theta_n}) = 0_p$, we have that

$$\begin{aligned}
0_p &= \lambda_n^{-d/2}b^{-d}\sum_{\boldsymbol{i}\in\mathcal{I}_n}\frac{\big(\partial A_n(\boldsymbol{i};\hat\theta_n)/\partial\theta\big)'\tilde t_{n,\hat\theta_n}}{1+\tilde t'_{n,\hat\theta_n}A_n(\boldsymbol{i};\hat\theta_n)} \qquad (S.9)\\
&= D_n(\theta_0)'b_n^{-d}\tilde t_{n,\hat\theta_n} + O_p(C_n)\\
&= D(\theta_0)'b^{-d}\tilde t_{n,\hat\theta_n} + o_p(\|b^{-d}\tilde t_{n,\hat\theta_n}\|)
\end{aligned}$$

by Lemma 3(a) and

$$\begin{aligned}\|C_n\| &\leq \frac{\|\tilde{t}_{n,\hat{\theta}_n}\|^2 Z_n(\hat{\theta}_n) b^{-d}}{1-\|\tilde{t}_{n,\hat{\theta}_n}\| Z_n(\hat{\theta}_n)}[J_n(\theta_0) + O_p(b^{-d}n^{-1}\lambda_n^d b^d m_n(\lambda_n^{-d/2}\log n)^{1+\gamma})] \\ &= o_p(\|b^{-d}\tilde{t}_{n,\hat{\theta}_n}\|).\end{aligned}$$

Now letting $\delta_n = \|b^{-d}\tilde{t}_{n,\hat{\theta}_n}\| + \nu_{\hat{\theta}_n}$, from (S.8) and (S.9), we may write

$$\begin{bmatrix} \Sigma_\infty & -D(\theta_0) \\ D(\theta_0)' & 0_{p\times p} \end{bmatrix} \begin{pmatrix} b_n^{-d}\tilde{t}_{n,\hat{\theta}_n} \\ \lambda_n^{d/2}(\hat{\theta}_n - \theta_0) \end{pmatrix} = \begin{bmatrix} A_n(\theta_0) + o_p(\delta_n) \\ o_p(\delta_n) \end{bmatrix},$$

$$\begin{bmatrix} \Sigma_\infty & -D(\theta_0) \\ D(\theta_0)' & 0_{p\times p} \end{bmatrix}^{-1} = \begin{bmatrix} U(\theta_0) & \Sigma_\infty^{-1}D(\theta_0)V(\theta_0) \\ -V(\theta_0)D(\theta_0)'\Sigma_\infty^{-1} & V(\theta_0) \end{bmatrix}$$

for $V(\theta_0), U(\theta_0)$ defined in Theorem 2. By Lemma 2(a), $A_n(\theta_0) \xrightarrow{d} N(0_r, \Sigma_\infty)$ holds so it follows that $\delta_n = O_p(1)$ and the limiting distribution of $\hat{\theta}_n$ is given by

$$\begin{aligned}\begin{pmatrix} b_n^{-d}\tilde{t}_{n,\hat{\theta}_n} \\ \lambda_n^{d/2}(\hat{\theta}_n - \theta_0) \end{pmatrix} &= \begin{bmatrix} U(\theta_0) \\ -V(\theta_0)D(\theta_0)'\Sigma_\infty^{-1} \end{bmatrix} A_n(\theta_0) + o_p(1) \qquad \text{(S.10)} \\ &\xrightarrow{d} N\left(\begin{pmatrix} 0_r \\ 0_p \end{pmatrix}, \begin{bmatrix} U(\theta_0) & 0 \\ 0 & V(\theta_0) \end{bmatrix} \right)\end{aligned}$$

The proof of Theorem 2(i) is complete, recalling $\tilde{t}_{n,\hat{\theta}_n} = \lambda_n^{-d/2} n t_{n,\hat{\theta}_n}$.

To prove parts (ii) and (iii) of Theorem 2, let $PP = \Sigma_\infty^{-1/2}D(\theta_0)V(\theta_0)D(\theta_0)'\Sigma_\infty^{-1/2}$ denote the projection matrix corresponding to the columns of $\Sigma_\infty^{-1/2}D(\theta_0)$ and let $I_{r\times r}$ denote the $r\times r$ identity matrix. Using (S.7) along with $\|\hat{\theta}_n - \theta_0\| = O_p(\lambda_n^{-d/2})$ by (S.10) and $\nu_{\theta_0} = 1$ in (S.7), we write

$$\begin{aligned}\ell_n(\hat{\theta}_n) &= [\Sigma_\infty^{-1/2}A_n(\theta_0)]'(I_{r\times r} - PP)[\Sigma_\infty^{-1/2}A_n(\theta_0)] + o_p(1), \\ \ell_n(\theta_0) &= [\Sigma_\infty^{-1/2}A_n(\theta_0)]'[\Sigma_\infty^{-1/2}A_n(\theta_0)] + o_p(1).\end{aligned}$$

The chi-square limit distributions for $\ell(\theta_0) - \ell_n(\hat{\theta}_n)$ and $\ell_n(\hat{\theta}_n)$ in Theorem 2(ii) and (iii), respectively, now follow by Lemma 2(a) as PP, $I_{r\times r} - PP$ are orthogonal idempotent matrices with ranks p, $r-p$. □

References

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory.* Springer, New York.

Lahiri, S. N. (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhya Ser. A* **65**, 356-388.

Lahiri, S. N. and Zhu, J. (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Ann. Statist.* **34**, 1774-1813.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling.* Springer-Verlag, New York.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.