Supplement to "Large Sample Justifications for the Bayesian Empirical Likelihood"

Naoya Sueishi*
Kobe University

October 20, 2022

1 Convolution theorem

This section gives the proof of Theorem 4.1.

Let $\dot{\ell}_{\theta_0,\eta_0}(x) = -\mathbb{E}[\nabla m_{\theta_0}(X)]\mathbb{E}[m_{\theta_0}(X)m_{\theta_0}(X)']^{-1}m_{\theta_0}(x)$. Our tangent set is given by $\dot{\mathcal{P}}_P = \text{lin } \dot{\ell}_{\theta_0,\eta_0} + {}_\eta\dot{\mathcal{P}}_P$, where

$$ {}_\eta\dot{\mathcal{P}}_P = \left\{ \dot{l} \in L_2(P) : \mathbb{E}[\dot{l}(X)] = 0 \text{ and } \mathbb{E}[\dot{l}(X)m_{\theta_0}(X)] = 0 \right\}. $$

It is clear that $\dot{\mathcal{P}}_P$ is a linear space.

For $g \in \dot{\mathcal{P}}_P$, $P_{t,g}$ denotes the one-dimensional submodel whose score is g. A sequence of estimators $\{T_n\}$ is regular with respect to $\dot{\mathcal{P}}_P$ if there exists a fixed probability measure L such that

$$ \sqrt{n}(T_n - \psi(P_{1/\sqrt{n},g})) \overset{P_{1/\sqrt{n},g}}{\rightsquigarrow} L $$

for all $g \in \dot{\mathcal{P}}_P$, where $\overset{P_{1/\sqrt{n},g}}{\rightsquigarrow}$ denotes weak convergence under $P_{1/\sqrt{n},g}$.

We utilize the following lemma for the proof.

Lemma 4.1 *Suppose that Assumption 4.1 holds. Then, for any $h \in \mathbb{R}^p$ and $\dot{l} \in {}_\eta\dot{\mathcal{P}}_P$, there exists a path $t \mapsto \eta_t$ such that*

$$ \int \left(\frac{\sqrt{p_{\theta_0+th,\eta_t}} - \sqrt{p_{\theta_0,\eta_0}}}{t} - \frac{1}{2}(h'\dot{\ell}_{\theta_0,\eta_0} + \dot{l})\sqrt{p_{\theta_0,\eta_0}} \right)^2 d\xi \to 0 $$

as $t \to 0$.

The following proof closely follows that of Theorem 25.20 in van der Vaart (1998) and Theorem 3.11.2 in van der Vaart and Wellner (1996).

*Graduate School of Economics, 2-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan. Email: sueishi@econ.kobe-u.ac.jp. Phone: +81-78-803-6827.

Proof of Theorem 4.1 Let $g_P = (g_1, \dots, g_m)'$ be an orthonormal basis of a subspace of $\dot{\mathcal{P}}_P$. Then, by Lemma 4.1 and Lemma of 25.14 of van der Vaart (1998), it follows that

$$\Lambda_{n,h} \equiv \log \prod_{i=1}^{n} \frac{dP_{1/\sqrt{n},h'g_P}}{dP}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h' g_P(X_i) - \frac{1}{2} h'h + o_P(1) \tag{1}$$

for any $h \in \mathbb{R}^m$. Lemma 4.1 also implies that

$$\sqrt{n}(\psi(P_{1/\sqrt{n},h'g_P}) - \psi(P)) \to Ah, \tag{2}$$

where $A = \mathbb{E}[\tilde{\psi}_P(X) g_P(X)']$ and $\tilde{\psi}_P = I^{-1}_{\theta_0,\eta_0} \dot{\ell}_{\theta_0,\eta_0}$ (see Lemma 25.25 of van der Vaart 1998).

Let $Z_{n,h} = \sqrt{n}(T_n - \psi(P_{1/\sqrt{n},h'g_P}))$ and $\Delta_{n,h} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h' g_P(X_i)$. Then, the marginals of the sequence $(Z_{n,0}, \Delta_{n,h})$ converge in distribution under P. Thus, by Prohorov's theorem, there exists a subsequence $\{n\}$ such that

$$(Z_{n,0}, \Delta_{n,h}) \overset{P}{\rightsquigarrow} (Z, h'\Delta)$$

jointly. By (1) and (2), we also have

$$(Z_{n,h}, \Lambda_{n,h}) \overset{P}{\rightsquigarrow} \left(Z - Ah, h'\Delta - \frac{1}{2} h'h \right).$$

Notice that $\Lambda_{n,h}$ converges to $N(-\frac{1}{2}h'h, h'h)$. Thus, by applying the Le Cam's third lemma, the limit law of $Z_{n,h}$ under $P_{1/\sqrt{n},h'g_P}$ is obtained by

$$L_h(B) = \mathbb{E} 1_B (Z - Ah) e^{h'\Delta - \frac{1}{2}h'h}.$$

Because of the regularity of T_n, L_h must coincide with L for any h. Thus, taking the averaging of both sides over h with respect to $N(0, \lambda^{-1} I)$, we obtain

$$L(B) = \int \mathbb{E} 1_B \left(Z - \frac{A\Delta}{1+\lambda} - \frac{Ah}{(1+\lambda)^{1/2}} \right) c_\lambda(\Delta) dN(0, I)(h),$$

where $c_\lambda(\Delta) = (1 + \lambda^{-1})^{-m/2} \exp(\frac{1}{2}(1+\lambda)^{-1} \Delta'\Delta)$. This representation shows that L can be written as the law of the sum of two independent random elements $-G_\lambda$ and W_λ, where $G_\lambda \sim N(0, AA'/(1+\lambda))$ and

$$P(W_\lambda \in B) = \mathbb{E} 1_B \left(Z - \frac{A\Delta}{1+\lambda} \right) c_\lambda(\Delta).$$

By letting $\lambda \to 0$, we have $(G_\lambda, W_\lambda) \rightsquigarrow (G + W)$, where G and W are independent, $G \sim N(0, AA')$, and $G + W \sim L$.

The proof completes if the difference between AA' and $\mathbb{E}[\tilde{\psi}_P(X) \tilde{\psi}_P(X)']$ can be arbitrarily small. Here, since g_P is orthonormal, Ag_P is the orthogonal projection of $\tilde{\psi}_P$ onto lin g_P. Because $\tilde{\psi}_P$ is contained in the closed linear span of $\dot{\mathcal{P}}_P$, we can choose g_P so that Ag_P is arbitrarily close to $\tilde{\psi}_P$. $\square$

2 Bernstein-von Mises theorem

This section gives proofs of Theorems 5.1 and 6.1.

Throughout this section, $\|P-Q\|$ denotes the total variation distance between two probability measures P and Q. For a probability measure P, we define $P^K(B) = P(B \cap K)/P(K)$.

Let Φ_n denote the normal distribution with mean Δ_n and variance matrix I_{θ_0,η_0}^{-1}. Let $\pi_n(h)$ be the Lebesgue prior density of the local parameter $h = \sqrt{n}(\theta - \theta_0)$. Moreover, let Π_n^{EL} and Π_n denote the BEL posterior and the semiparametric Bayesian posterior of the local parameter, respectively. Our goal is to show that

$$\|\Pi_n^{EL} - \Phi_n\| \xrightarrow{P} 0 \quad \text{and} \quad \|\Pi_n - \Phi_n\| \xrightarrow{P} 0.$$

To show the theorems, we utilize the following lemmas.

Lemma A.1 *Suppose that Assumption 5.1 holds. Then, we have*

$$\sup_{h \in K} \left| \log \prod_{i=1}^{n} \frac{dP_{\theta_0 + h/\sqrt{n}, \mathbb{P}_n}}{dP_{\theta_0, \mathbb{P}_n}}(X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h' \dot{\ell}_{\theta_0,\eta_0}(X_i) + \frac{1}{2} h' I_{\theta_0,\eta_0} h \right| = o_P(1)$$

for any compact $K \subset \mathbb{R}^p$.

Lemma A.2 *Suppose that Assumptions 5.1 and 5.2 hold. Then, for any sequence of origin centered balls $\{K_n\}$ with radii $M_n \to \infty$, we have*

$$\Pi_n^{EL}(h \in K_n | X_1, \ldots, X_n) \xrightarrow{P} 1.$$

Lemma A.5 *Suppose that Assumptions 6.1 and 6.2 hold. Then, for any bounded random sequence $\{h_n\}$, we have*

$$\log \frac{s_n(h_n)}{s_n(0)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_n' \dot{\ell}_{\theta_0,\eta_0}(X_i) - \frac{1}{2} h_n' I_{\theta_0,\eta_0} h_n + o_P(1),$$

where

$$s_n(h) = \int \prod_{i=1}^{n} \frac{p_{\theta_0 + h/\sqrt{n}, \eta}}{p_{\theta_0,\eta_0}}(X_i) d\Pi_H(\eta). \tag{3}$$

Lemma A.6 *Suppose that Assumptions 6.1–6.3 hold. Then, for any sequence of origin centered balls $\{K_n\}$ with radii $M_n \to \infty$, we have*

$$\Pi_n(h \in K_n | X_1, \ldots, X_n) \xrightarrow{P} 1.$$

The following proof closely follows that of Theorem of 2.1 in Kleijn and van der Vaart (2012).

Proof of Theorem 5.1 Let $K \subset \mathbb{R}^p$ be a compact set centered at the origin. Also, let ϕ_n be the Lebesgue density of Φ_n. Then, the function $f_n : K \times K \to \mathbb{R}$ given by

$$f_n(g,h) = \left(1 - \frac{\phi_n(h)}{\phi_n(g)}\frac{s_n(g)}{s_n(h)}\frac{\pi_n(g)}{\pi_n(h)}\right)_+$$

with

$$s_n(h) = \prod_{i=1}^{n} \frac{dP_{\theta_0+h/\sqrt{n},\mathbb{P}_n}}{dP_{\theta_0,\mathbb{P}_n}}(X_i)$$

is well-defined for large n.

For any random sequences $\{h_n\},\{g_n\} \subset K$, we have $\pi_n(g_n)/\pi_n(h_n) \to 1$. Therefore, it follows from Lemma A.1 that

$$\log \frac{\phi_n(h_n)}{\phi_n(g_n)}\frac{s_n(g_n)}{s_n(h_n)}\frac{\pi_n(g_n)}{\pi_n(h_n)} = o_P(1).$$

Since f_n is continuous with respect to its two arguments, the above equation also implies that

$$\sup_{g,h\in K} f_n(g,h) \xrightarrow{P} 0.$$

Let $\Pi_n^{EL,K}$ be the conditional version of Π_n^{EL}. Similarly, let Φ_n^K and ϕ_n^K denote the conditional versions of Φ_n and ϕ_n, respectively. Since K contains a neighborhood of 0, $\Phi_n(K) > 0$ is guaranteed. Let Ξ_n be a sequence of events such that $\left\{\Pi_n^{EL}(K) > 0\right\}$. For a given $\eta > 0$, we define

$$\Omega_n = \left\{\sup_{g,h\in K} f_n(g,h) \le \eta\right\}.$$

Because the total variation distance can be written as $\|P-Q\| = 2\int(1-\frac{dQ}{dP})_+ dP$ for any probability measures P and Q, we have

$$\begin{aligned}
\frac{1}{2}\|\Pi_n^{EL,K} - \Phi_n^K\|1_{\Omega_n\cap\Xi_n} &= \int_K \left(1 - \frac{\phi_n^K(h)\int_K s_n(g)\pi_n(g)dg}{s_n(h)\pi_n(h)}\right)_+ d\Pi_n^{EL,K}(h)1_{\Omega_n\cap\Xi_n} \\
&= \int_K \left(1 - \int_K \frac{s_n(g)\pi_n(g)\phi_n(h)}{s_n(h)\pi_n(h)\phi_n(g)} d\Phi_n^K(g)\right)_+ d\Pi_n^{EL,K}(h)1_{\Omega_n\cap\Xi_n},
\end{aligned}$$

where we use $\phi_n^K(h)/\phi_n^K(g) = \phi_n(h)/\phi_n(g)$. Thus, by applying the Jensen's inequality, we obtain

$$\begin{aligned}
\frac{1}{2}\|\Pi_n^{EL,K} - \Phi_n^K\|1_{\Omega_n\cap\Xi_n} &\le \int_K\int_K \left(1 - \frac{s_n(g)\pi_n(g)\phi_n(h)}{s_n(h)\pi_n(h)\phi_n(g)}\right)_+ d\Phi_n^K(g)d\Pi_n^{EL,K}(h)1_{\Omega_n\cap\Xi_n} \\
&\le \int_K\int_K \sup_{g,h\in K} f_n(g,h)1_{\Omega_n\cap\Xi_n} d\Phi_n^K(g)d\Pi_n^{EL,K}(h).
\end{aligned}$$

Because the total variation norm is bounded by 2, it follows that

$$\mathbb{E}[\|\Pi_n^{EL,K} - \Phi_n^K\|1_{\Xi_n}] \le \mathbb{E}[\|\Pi_n^{EL} - \Phi_n^K\|1_{\Omega_n\cap\Xi_n}] + 2P(\Xi_n \setminus \Omega_n) \to 0.$$

Let $\{K_m\}$ be a sequence of balls centered at the origin with radii $M_m \to \infty$ as $m \to \infty$. Then, the above display is true for each K_m. Hence, we can choose a sequence $\{K_n\}$ that satisfies

$\mathbb{E}[\|\Pi_n^{EL,K_n} - \Phi_n^{K_n}\| 1_{\Xi_n}] \to 0$. Moreover, the corresponding events $\Xi_n = \{\Pi_n^{EL}(K_n) > 0\}$ satisfy $P(\Xi_n) \to 1$ by Lemma A.2. Thus, we obtain

$$\mathbb{E}[\|\Pi_n^{EL,K_n} - \Phi_n^{K_n}\|] \to 0,$$

where it is understood that the conditional probabilities are well-defined on sets of probability growing to one. Furthermore, by Lemma A.2 and Lemma 5.2 of Kleijn and van der Vaart (2012), we have

$$\Pi_n^{EL}(\mathbb{R}^p \setminus K_n) \xrightarrow{P} 0 \quad \text{and} \quad \Phi_n(\mathbb{R}^p \setminus K_n) \xrightarrow{P} 0.$$

Thus, by Lemma 5.1 of Kleijn and van der Vaart (2012), we conclude that

$$\|\Pi_n^{EL,K} - \Phi_n^K\| - \|\Pi_n^{EL} - \Phi_n\| \xrightarrow{P} 0.$$

□

Proof of Theorem 6.1 The proof is a simple modification of that of Theorem 5.1. We replace $s_n(h)$ in the previous proof by (3). Then, Lemma A.5 implies that

$$\sup_{g,h \in K} f_n(g,h) \xrightarrow{P} 0.$$

Thus, for any compact set K centered at the origin, we have

$$\mathbb{E}[\|\Pi_n^K - \Phi_n^K\| 1_{\Xi_n}] \to 0$$

where $\Xi_n = \{\Pi_n(K) > 0\}$.

Lemma A.6 implies that we can also find a sequence of sets $\{K_n\}$ such that $\Pi_n(\mathbb{R}^p \setminus K_n) \to 0$ and $\mathbb{E}[\|\Pi_n^{K_n} - \Phi_n^{K_n}\|] \to 0$. Thus, using Lemmas of 5.1 and 5.2 of Kleijn and van der Vaart (2012), we obtain the desired result. □

References

Kleijn, B. J. K. and A. W. van der Vaart (2012). The Bernstein-von-Mises theorem under misspecification. *Electronic Journal of Statistics 6*, 354–381.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.