Received: 14 November 2018 | Revised: 22 August 2019 | Accepted: 23 November 2019

DOI: 10.1111/sjos.12436

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Asymptotic normality of generalized maximum spacing estimators for multivariate observations

Kristi Kuljus[1] | **Bo Ranneby**[2]

[1]Institute of Mathematics and Statistics, University of Tartu

[2]Department of Forest Economics, Swedish University of Agricultural Sciences

Correspondence
Kristi Kuljus, Institute of Mathematics and Statistics, J. Liivi 2, 50409, Tartu, Estonia.
Email: kristi.kuljus@ut.ee

Funding information
Estonian Institutional Research Funding, IUT34-5

Abstract

In this paper, the maximum spacing method is considered for multivariate observations. Nearest neighbor balls are used as a multidimensional analogue to univariate spacings. A class of information-type measures is used to generalize the concept of maximum spacing estimators of model parameters. Asymptotic normality of these generalized maximum spacing estimators is proved when the assigned model class is correct, that is, the true density is a member of the model class.

KEYWORDS

asymptotic normality, consistency, divergence measures, maximum spacing estimation, nearest neighbor balls

1 | GENERALIZED MAXIMUM SPACING ESTIMATORS

1.1 | Introduction

For independent and identically distributed univariate observations, a new estimation method, the maximum spacing (MSP) method, was defined in Ranneby (1984) and independently by Cheng and Amin (1983). In Ranneby, Jammalamadaka, and Teterukovskiy (2005), the MSP method was extended to multivariate observations for the Kullback–Leibler information measure. In Kuljus and Ranneby (2015), the multivariate MSP estimation method based on nearest neighbor balls was considered for a broader class of information-type measures. Weak and strong consistency of these generalized maximum spacing (GMSP) estimators of model parameters based on a random sample from a multivariate parametric distribution was proved under general conditions. In the univariate case, such GMSP estimators based on different metrics were studied in Ranneby and Ekström (1997), Ekström (2001) and Ghosh and Jammalamadaka (2001); in the last work also, asymptotic normality of GMSP estimates was proved. Consistency and asymptotic normality of GMSP estimates in the univariate case was also considered in Luong (2018). As exemplified already in Ranneby (1984), an advantage of the MSP method compared to the maximum likelihood method is the possibility of checking the validity of the assigned model class at the same time with solving the estimation problem. In Kuljus and Ranneby (2015), it was demonstrated that combining information from spacing functions under different divergence measures can provide further insight in the model validation context. In the present paper, we study asymptotic normality of GMSP estimators for information-type measures considered in Kuljus and Ranneby (2015).

1.2 | Notation and definitions

Let $\xi_1, \dots, \xi_n$ be a sequence of independent and identically distributed d-dimensional random vectors with distribution P_0 that is absolutely continuous with respect to Lebesgue measure. Let the corresponding density function be $g(x)$. Define the nearest neighbor distance to the random variable ξ_i as

$$R_n(i) = \min_{j \neq i} |\xi_i - \xi_j|, \qquad i = 1, \dots, n.$$

Let $B(x, r) = \{y : |x - y| \leq r\}$ denote the ball of radius r and center x. Let NN_i denote the nearest neighbor of ξ_i and let $B_n(\xi_i)$ denote its nearest neighbor ball, that is, this is a ball with center ξ_i and radius $R_n(i)$. Suppose we assign a model with density functions $\{f_\theta(x),\ \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^q$, and assume that the true density $g(x)$ belongs to the family with the parameter vector given by θ_0. Define random variables $z_{i,n}(\theta)$ as

$$z_{i,n}(\theta) = nP_\theta(B_n(\xi_i)) = n \int_{B_n(\xi_i)} f_\theta(y)dy, \quad i = 1, \dots, n.$$

In Kuljus and Ranneby (2015) the MSP method was generalized to multivariate observations for strictly concave functions $h : (0, \infty) \to (-\infty, 0]$ with maximum at $x = 1$. The generalized MSP function $S_n(\theta)$ was defined as

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(z_{i,n}(\theta)).$$

Definition 1. The parameter value that maximizes $S_n(\theta)$ is called the GMSP estimate of θ and denoted by $\hat{\theta}_n$. If $\sup_\theta S_n(\theta)$ is not attained for any θ in the admissible set Θ, then the GMSP estimate $\hat{\theta}_n$ is defined as any point of Θ that satisfies

$$S_n(\hat{\theta}_n) \geq -c_n + \sup_{\theta \in \Theta} S_n(\theta),$$

where $c_n > 0$ is a sequence of constants such that $c_n \to 0$ as $n \to \infty$.

To prove asymptotic normality of $\hat{\theta}_n$, we will work with the partial derivatives of $h(z_{i,n}(\theta))$, the vector of partial derivatives is denoted by $\psi_n(\xi_i, \theta)$. Let $v(z_{i,n}(\theta)) = h'(z_{i,n}(\theta))z_{i,n}(\theta)$. Then

$$\psi_n(\xi_i, \theta) = v(z_{i,n}(\theta))\frac{1}{P_\theta(B_n(\xi_i))}\int_{B_n(\xi_i)} \nabla f_\theta(y)dy.$$

Define $\tilde{\psi}_n(\xi_i, \theta)$ and $\psi(\xi_i, \theta)$ as follows:

$$\tilde{\psi}_n(\xi_i, \theta) = v\left(z_{i,n}(\theta_0)\frac{f_\theta(\xi_i)}{g(\xi_i)}\right)\frac{\nabla f_\theta(\xi_i)}{f_\theta(\xi_i)}, \qquad \psi(\xi_i, \theta) = v\left(Z\frac{f_\theta(\xi_i)}{g(\xi_i)}\right)\frac{\nabla f_\theta(\xi_i)}{f_\theta(\xi_i)},$$

where $Z \sim Exp(1)$ and Z is independent of ξ_i. Observe that $z_{i,n}(\theta)$ can be written as

$$z_{i,n}(\theta) = z_{i,n}(\theta_0)\frac{1}{P_{\theta_0}(B_n(\xi_i))}\int_{B_n(\xi_i)} f_\theta(y)dy.$$

Since

$$\frac{1}{P_{\theta_0}(B_n(\xi_i))}\int_{B_n(\xi_i)} f_\theta(y)dy \overset{a.s.}{\to} \frac{f_\theta(\xi_i)}{g(\xi_i)}, \qquad \frac{1}{P_\theta(B_n(\xi_i))}\int_{B_n(\xi_i)} \nabla f_\theta(y)dy \overset{a.s.}{\to} \frac{\nabla f_\theta(\xi_i)}{f_\theta(\xi_i)},$$

it follows that $\tilde{\psi}_n(\xi_i, \theta)$ is obtained from $\psi_n(\xi_i, \theta)$ by substituting the integral quantities above with their almost sure limits. Let $\lambda_n(\theta) = E[\psi_n(\xi_i, \theta)]$ and $\lambda(\theta) = E[\psi(\xi_i, \theta)]$.

Examples of strictly concave functions h with maximum at $x = 1$ are:

$$h_1(x) = \ln\ x - x + 1, \quad h_2(x) = (1 - x)\ \ln\ x, \quad h_3(x) = -|1 - x^{1/p}|^p,$$
$$h_4(x) = -|1 - x|^p, \quad h_5(x) = \text{sgn}(1 - \alpha)(x^\alpha - \alpha x + \alpha - 1),$$

where $\alpha > 0$, $\alpha \neq 1$, and $p \geq 1$. Here h_2 corresponds to Jeffreys' divergence measure, h_3 to the Hellinger distance, h_4 to Vajda's measure of information and h_5 to Rényi's divergence measure. In this article, we will consider only $p = 2$ for function families h_3 and h_4. For h_5, we restrict α to $(0, 1) \cup (1, 2]$. Observe that $\alpha = 1/2$ corresponds to h_3 with $p = 2$ and $\alpha = 2$ corresponds to h_4 with $p = 2$. Thus, h_3 and h_4 with $p = 2$ will be covered by h_5.

1.3 | Assumptions

We will work with first and second order derivatives of different functions with respect to the components θ_j, $j = 1, \dots, q$, of the parameter vector θ. The notations $\partial f_\theta(\xi)$ and $\partial^2 f_\theta(\xi)$ will be used instead of $\frac{\partial f_\theta(\xi)}{\partial \theta_j}$ and $\frac{\partial^2 f_\theta(\xi)}{\partial \theta_j \partial \theta_l}$, respectively, when the computations are analogous or a certain condition has to hold independently of $j, l = 1, \dots, q$. Let Θ_0 denote a neighborhood of θ_0. Let $\theta_0^{(n)}$ denote the point maximizing the expectation function $E[h(z_{i,n}(\theta))]$, thus it satisfies $\lambda_n(\theta_0^{(n)}) = 0$. Recall that θ_0 satisfies $\lambda(\theta_0) = 0$. Asymptotic normality of $\hat{\theta}_n$ will be proved under different combinations (depending on the function h) of the following assumptions:

- A1 $\{x \in \mathbb{R}^d\ :\ g(x) > 0\}$ is an open set in $\mathbb{R}^d$, g is uniformly bounded and continuous on $\{g > 0\}$.
- A2 Assume that θ_0 is an interior point of Θ. Suppose $f_\theta(x)$ and its first and second order derivatives with respect to the components of θ are continuous in $\theta \in \Theta_0$ and in x.

A3

$$E\left[\sup_{\Theta_0}\left|\frac{\partial f_\theta(\xi)}{g(\xi)}\right|^2\right]<\infty.$$

A4 The following random variables are uniformly integrable:

$$\sup_{\Theta_0}\frac{1}{P_\theta(B_n(\xi))}\int_{B_n(\xi)}|\partial f_\theta(y)|dy.$$

A5 The following random variables are uniformly integrable:

$$\left(\frac{1}{P_{\theta_0^{(n)}}(B_n(\xi))}\int_{B_n(\xi)}|\partial f_{\theta_0^{(n)}}(y)|dy\right)^2.$$

A6 The following random variables are uniformly integrable:

$$\sup_{\Theta_0}\left(\frac{1}{P_\theta(B_n(\xi))}\int_{B_n(\xi)}|\partial f_\theta(y)|dy\right)^2.$$

A7

$$E\left[\sup_{\Theta_0}\left|\frac{\partial f_\theta(\xi)}{g(\xi)}\right|^4\right]<\infty,\quad E\left[\sup_{\Theta_0}\left|\frac{f_\theta(\xi)}{g(\xi)}\right|^4\right]<\infty.$$

A8 The following random variables are uniformly integrable:

$$\sup_{\Theta_0}\frac{1}{P_\theta(B_n(\xi))}\int_{B_n(\xi)}|\partial^2 f_\theta(y)|dy.$$

A9 For some ε, $0<\varepsilon<1$,

$$E\left[\sup_{\Theta_0}\left|\frac{\partial^2 f_\theta(\xi)}{g(\xi)}\right|^{1+\varepsilon}\right]<\infty.$$

A10

$$E\left[\sup_{\Theta_0}\left|\frac{\partial^2 f_\theta(\xi)}{g(\xi)}\right|^2\right]<\infty.$$

Let $q(x)\le\inf_{\Theta_0} f_\theta(x)$ be the Radon–Nikodym derivative of a finite measure Q with respect to the Lebesgue measure on $\mathbb{R}^d$. The following two remarks give two examples of conditions when A4 and A5 are satisfied.

Remark 1. Suppose

$$E_q\left(\sup_{\Theta_0}\frac{|\partial f_\theta(\xi)|}{q(\xi)}\right)^2 < \infty, \qquad E_g\left(\frac{g(\xi)}{q(\xi)}\right) < \infty,$$

then A4 holds. If we instead assume $E_q\left(\sup_{\Theta_0}\frac{|\partial f_\theta(\xi)|}{q(\xi)}\right)^4 < \infty$, then A6 holds.

Remark 2. If for some constants c_1, c_2 and $\forall x \in \mathbb{R}^d$, $\forall \theta \in \Theta_0$,

$$c_1 g(x) \leq f_\theta(x) \leq c_2 g(x),$$

then A4 and A5 follow from A3.

For most distributions, it is not difficult to check assumptions A3, A7, A9, and A10. Remark 1 is useful for checking assumptions A4, A6, and A8. It is not difficult to verify that our assumptions are satisfied for the class of multivariate normal distributions and for finite mixtures of normal distributions.

1.4 | Main result and idea for proving asymptotic normality of $\hat{\theta}_n$

In this section we will state the main theorem with necessary assumptions and we will describe the main steps needed for proving asymptotic normality of $\hat{\theta}_n$. The proof of the theorem will be given in Section 4. Let $I(\theta_0)$ denote the Fisher information matrix at $\theta = \theta_0$, that is, $I(\theta_0)$ is the covariance matrix of $(\nabla f_\theta(\xi_i))_{\theta=\theta_0}/g(\xi_i)$.

Theorem 1. *Let $\hat{\theta}_n \xrightarrow{p} \theta_0$ hold. Suppose assumption A1 is satisfied and $I(\theta_0)$ is positive definite. Suppose assumptions*

(i) A2, A3, A4, A5, A6, A8, A9 hold and consider function h_1 or h_5 with $\alpha \in (0, 1)$;
(ii) A2, A3, A4, A5, A6, A7, A8, A10 hold and consider h_2;
(iii) A2, A3, A7, A10 hold and consider h_5 with $\alpha \in (1, 2]$.

Then for (i), (ii), and (iii),

$$\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)}) \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma_q^2}{b_h^2} I(\theta_0)^{-1}\right),$$

where $b_h = E[h''(Z)Z^2]$, and where σ_q^2 will be calculated as in (5) with $q(t) = h'(t)t$.

Consistency of $\hat{\theta}_n$ implies that a sequence $\{\delta_n\}_1^\infty$ can be chosen so that $\delta_n \to 0$ slowly enough to ensure

$$P(|\hat{\theta}_n - \theta_0| \geq \delta_n) \to 0 \quad \text{as} \quad n \to \infty.$$

We will show that $\lambda_n(\theta)$ converges uniformly to $\lambda(\theta)$ in a neighborhood of θ_0, thus $\theta_0^{(n)} \to \theta_0$ as $n \to \infty$. Therefore, we will consider shrinking neighborhoods $\Theta_n = \{\theta : |\theta - \theta_0| \leq \alpha_n\}$, where

$\delta_n < \alpha_n$ and $\alpha_n \to 0$ is such, that $\theta_0^{(n)} \in \Theta_n$ for every n. The key steps for proving asymptotic normality of $\hat{\theta}_n$ are the following.

Step 1. First we will show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_n(\xi_i, \theta_0),$$

is asymptotically normally distributed. To prove asymptotic normality of this quantity, we will interpret it as a function of a weighted empirical process which converges weakly to a Gaussian process.

Step 2. We will prove that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_n(\xi_i, \theta_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_n(\xi_i, \theta_0^{(n)}) \xrightarrow{p} 0.$$

Thus, $\frac{1}{\sqrt{n}} \sum_i \psi_n(\xi_i, \theta_0^{(n)})$ has the same asymptotic distribution as $\frac{1}{\sqrt{n}} \sum_i \tilde{\psi}_n(\xi_i, \theta_0)$.

Step 3. To finalize proving asymptotic normality of $\hat{\theta}_n$, we will follow the approach by Huber (1967) and show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_n(\xi_i, \theta_0^{(n)}) + \sqrt{n}\lambda_n(\hat{\theta}_n) \xrightarrow{p} 0. \tag{1}$$

Expanding $\lambda_n(\hat{\theta}_n)$ around $\theta_0^{(n)}$ then gives that $\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})$ is asymptotically normally distributed. Crucial for proving Equation (1) is lemma 3 in Huber (1967) stating that

$$\sup_{\Theta_n} \frac{\left| \frac{1}{\sqrt{n}} \sum_i (\psi_n(\xi_i, \theta) - \psi_n(\xi_i, \theta_0^{(n)})) - \sqrt{n}\lambda_n(\theta) \right|}{1 + \sqrt{n}|\lambda_n(\theta)|} \xrightarrow{p} 0. \tag{2}$$

Since our assumptions imply that $\lambda_n(\theta)$ is continuously differentiable in a neighborhood of θ_0 with a negative definite derivative matrix $V_n(\theta)$, there exists $C > 0$ such that $|\lambda_n(\theta)| \geq C|\theta - \theta_0^{(n)}|$ when n is large enough. Thus, the convergence in Equation (2) follows if

$$\sup_{\Theta_n} \frac{\left| \frac{1}{\sqrt{n}} \sum_i (\psi_n(\xi_i, \theta) - \psi_n(\xi_i, \theta_0^{(n)})) - \sqrt{n}\lambda_n(\theta) \right|}{1 + \sqrt{n}|\theta - \theta_0^{(n)}|} \xrightarrow{p} 0$$

holds. Therefore, we will work with expressions having $1 + \sqrt{n}|\theta - \theta_0^{(n)}|$ in the denominator.

Convergence of the weighted empirical process in Step 1 is used to prove asymptotic normality of general functions of the process. Both convergences will be proved in Section 2. In Section 3, this result will be applied for proving asymptotic normality of the approximation of our function of interest, that is, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_n(\xi_i, \theta_0)$. Step 2 will also be proved in Section 3. Section 4 deals with

Step 3: bracketing technique and a stochastic differentiability condition will be used to prove asymptotic normality of $\hat{\theta}_n$.

2 | ASYMPTOTIC NORMALITY OF WEIGHTED EMPIRICAL PROCESSES

In this section we will modify the results of Schilling (1983) and prove that a weighted empirical process of $z_{1,n}(\theta_0), \dots, z_{n,n}(\theta_0)$ converges to a Gaussian process. Using a suitable transformation we then obtain asymptotic normality of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(z_{i,n}(\theta_0))u(\xi_i),$$

where $q(\cdot)$ is a function of bounded variation on $[0, A]$ for any $A > 0$, and $u(\cdot)$ is a continuous weight function with the properties $Eu(\xi_i) = 0$, $Eu^2(\xi_i) < \infty$. To prove asymptotic normality of the sum above, we use results from Bickel and Breiman (1983), Schilling (1983) and Zhou and Jammalamadaka (1993). Let

$$W_{i,n} = ng(\xi_i)V(R_n(i)), \quad i = 1, \dots, n,$$

where $V(r)$ represents the volume of a d-dimensional sphere of radius r. In Bickel and Breiman (1983), it is shown that the normalized (centered and scaled) empirical distribution function of $e^{-W_{1,n}}, \dots, e^{-W_{n,n}}$ converges under the true distribution weakly to a Gaussian process with mean zero and covariance function independent of the true underlying density. In Schilling (1983), the same result is proved for a weighted empirical process with a bounded continuous weight function. To be able to use theorem 2.2 from Schilling (1983), we will study truncated weight functions $u_N(\cdot)$ defined as follows. Since

$$Eu(\xi) = \int u^+(x)dP_0(x) - \int u^-(x)dP_0(x) = 0,$$

we can find for every $N > 0$ a constant $N^* > 0$ such that

$$\int_{\{u^+(x) \le N\}} u^+(x)dP_0(x) = \int_{\{u^-(x) \le N^*\}} u^-(x)dP_0(x).$$

Define a bounded weight function $u_N(x)$ as follows:

$$u_N(x) = \begin{cases} u(x), & -N^* \le u(x) \le N, \\ 0, & otherwise. \end{cases} \tag{3}$$

Take $N_1 = \max\{N, N^*\}$, then $Eu_N(\xi) = 0$ and $|u_N(\xi)| \le N_1$.

The general idea for proving the convergence

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(z_{i,n}(\theta_0))u(\xi_i) \xrightarrow{D} \mathcal{N}(0, \sigma_q^2 \tau^2),$$

where $\tau^2 = E[u^2(\xi_i)]$ and where σ_q^2 will be defined in Equation (5), will be as follows. We consider bounded weight functions $u_N(x)$ defined as in Equation (3) and define the weighted empirical processes $\{Y_n(t) \,:\, 0 \le t \le \infty\}$ and $\{Z_n(t) \,:\, 0 \le t \le \infty\}$:

$$Z_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I(z_{i,n}(\theta_0) > t) u_N(\xi_i), \tag{4}$$

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ I(W_{i,n} > t) u_N(\xi_i) - E\left[I(W_{i,n} > t) u_N(\xi_i) \right] \right\}.$$

From Schilling (1983) it follows that $\{Y_n(t)\}$ converges weakly to a Gaussian process. For large n we have $ng(\xi_i)V(R_n(i)) = W_{i,n} \approx nP_0(B_n(\xi_i)) = z_{i,n}(\theta_0)$. Thus, if we can show that $\{Z_n(t)\}$ is tight and $\mathrm{Var}(Z_n(t) - Y_n(t)) \to 0$ for every t, then $\{Z_n(t)\}$ converges to the same Gaussian process. Therefore, using the results from Bickel and Breiman (1983), Schilling (1983) and Zhou and Jammalamadaka (1993), we can show that

$$Z_n(t) \xrightarrow{D} Z(t),$$

where $\{Z(t) \,:\, 0 \le t \le \infty\}$ is a Gaussian process with mean zero and with a certain covariance function. We then apply the integral transform

$$s(x) = q(0)x(0) + \int_0^A x(t)dq(t)$$

to $Z_n(t)$ and obtain via $s(Z_n(t)) \xrightarrow{D} s(Z(t))$ the desired result.

Proposition 1. *Suppose A1 and the following conditions hold:*

$$|u_N(\xi_i)| \le N_1 \text{ for some } N_1 > 0, \quad Eu_N(\xi_i) = 0, \quad Eu_N^2(\xi_i) = \tau_N^2.$$

Then $\{Z_n(t) \,:\, 0 \le t \le \infty\}$ defined in Equation (4) converges weakly to a Gaussian process $\{Z(t) \,:\, 0 \le t \le \infty\}$ with mean zero and covariance function $\tau_N^2 k(s,t)$, where $k(s,t)$ is given by

$$k(s,t) = e^{-t} - te^{-s-t} + e^{-s-t} \int_{W(s,t)} (e^{\beta(s,t,x)} - 1)dx, \quad 0 \le s \le t \le \infty,$$

where

$$W(s,t) = \{x \in \mathbb{R}^d \,:\, r_1 \le |x| \le r_1 + r_2\}, \quad \beta(s,t,x) = \int_{B(0,r_1)\cap B(x,r_2)} dz,$$

with r_1 and r_2 corresponding to the volumes t and s of the balls $B(0,r_1)$ and $B(0,r_2)$, respectively.

Proof. From Schilling (1983) it follows directly that the centered empirical process $\{Y_n(t)\}$ converges weakly to the Gaussian process defined above. To conclude that $\{Z_n(t)\}$ converges to the same limit, we prove that for every t, $\mathrm{Var}(Z_n(t) - Y_n(t)) \to 0$, and that $\{Z_n(t)\}$ is tight.

(a) That $\mathrm{Var}(Z_n(t) - Y_n(t)) \to 0$ as $n \to \infty$, follows with minor modifications from Zhou and Jammalamadaka (1993). Since

$$\mathrm{Var}[Z_n(t) - Y_n(t)] \leq E\big[\big(I(z_{1,n}(\theta_0) > t) - I(W_{1,n} > t)\big)\, u_N(\xi_1)\big]^2$$
$$+\, n\left|\mathrm{Cov}\left[\big(I(z_{1,n}(\theta_0) > t) - I(W_{1,n} > t)\big)\, u_N(\xi_1), \big(I(z_{2,n}(\theta_0) > t) - I(W_{2,n} > t)\big)\, u_N(\xi_2)\right]\right|,$$

we need to show that

$$\lim_{n\to\infty} E\big[I(z_{1,n}(\theta_0) > t) - I(W_{1,n} > t)\big]^2 = 0,$$
$$\lim_{n\to\infty} n\left|\mathrm{Cov}\left[\big(I(z_{1,n}(\theta_0) > t) - I(W_{1,n} > t)\big)\, u_N(\xi_1), \big(I(z_{2,n}(\theta_0) > t) - I(W_{2,n} > t)\big)\, u_N(\xi_2)\right]\right| = 0.$$

The convergence of both terms follows as in the proof of Proposition 1 of Zhou and Jammalamadaka (1993). For the proof of the convergence of the covariance term, lemma 2.11 in Bickel and Breiman (1983) is fundamental.

(b) Tightness of $\{Z_n(t)\}$ can be proved similarly to Schilling (1983) and Bickel and Breiman (1983). As in Schilling (1983), we can split $Z_n(t)$ as follows: $Z_n(t) = Z_n^+(t) - Z_n^-(t)$, where

$$Z_n^{\pm}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ I(z_{i,n}(\theta_0) > t)u_N^{\pm}(\xi_i) - E\left[I(z_{i,n}(\theta_0) > t)u_N^{\pm}(\xi_i)\right] \right\}.$$

It is enough to show that $\{Z_n^+(t)\}$ is tight. Let $l_i = I(a \leq z_{i,n}(\theta_0) \leq b)$, then

$$El_1 = E\left[I\left(a \leq nP_0(B_n(\xi_1)) \leq b\right)\right] = \frac{n-1}{n}\int_a^b \left(1 - \frac{w}{n}\right)^{n-2} dw$$
$$< e^2 \int_a^b e^{-w} dw = e^2(e^{-a} - e^{-b}),$$
$$E\left[l_1 P_0^2(B_n(\xi_1))\right] = \frac{1}{n^2}\int_a^b \frac{n-1}{n} w^2 \left(1 - \frac{w}{n}\right)^{n-2} dw$$
$$< \frac{e^2}{n^2}\left(e^{-a}(a^2 + 2a + 2) - e^{-b}(b^2 + 2b + 2)\right).$$

Thus, applying theorem 2.1 in Bickel and Breiman (1983) gives that for some constant $M > 0$,

$$E\left[\sum_{i=1}^{n}(l_i - El_i)\right]^4 < M\left[n^2(Q(b) - Q(a))^2 + n\right],$$

where $Q(t)$ is a continuous distribution function defined as

$$Q(t) = 1 - \frac{1}{3}e^{-t}(t^2 + 2t + 3).$$

The rest of the proof goes according to Schilling (1983) and Bickel and Breiman (1983). ∎

Proposition 2. *Suppose the assumptions of Proposition* 1 *hold and* $q(t)$ *is of bounded variation on* $[0, A]$ *for each* $A > 0$. *Let* $|\int_1^\infty (te^{-t})^{1/2} dq(t)| < \infty$. *Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(z_{i,n}(\theta_0)) u_N(\xi_i) \xrightarrow{D} \mathcal{N}(0, \tau_N^2 \sigma_q^2),$$

where

$$\sigma_q^2 = q^2(0) + \int_0^\infty \int_0^\infty k(s,t) dq(s) dq(t) + 2q(0) \int_0^\infty k(0,t) dq(t), \tag{5}$$

with $k(0,t) = e^{-t} - te^{-t}$.

Proof. Recall the definition of the empirical process $Z_n(t)$ in Equation (4). In the proof of proposition 2 in Zhou and Jammalamadaka (1993) it is shown that for some $M > 0$, $\mathrm{Var}[Z_n(t)] \leq Mte^{-t}$, $t \geq 1$. Therefore it follows according to our assumption that $\int_0^\infty (\mathrm{Var}[Z_n(t)])^{1/2} dq(t) < \infty$, which implies (see e.g., Cramér and Leadbetter (1967), pp. 90–91) that for every n,

$$\int_0^A Z_n(t) dq(t) \stackrel{A\to\infty}{\to} \int_0^\infty Z_n(t) dq(t).$$

Analogously,

$$\int_0^A Z(t) dq(t) \stackrel{A\to\infty}{\to} \int_0^\infty Z(t) dq(t).$$

Therefore, $\int_0^\infty Z_n(t) dq(t)$ is well defined and it holds with probability 1 that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(z_{i,n}(\theta_0)) u_N(\xi_i) = -\int_0^\infty q(t) dZ_n(t) = q(0) Z_n(0) + \int_0^\infty Z_n(t) dq(t).$$

Since $q(t)$ is of bounded variation on $[0, A]$ for every $A > 0$, it follows from Proposition 1 that

$$-\int_0^A q(t) dZ_n(t) \xrightarrow{D} q(0) Z(0) + \int_0^A Z(t) dq(t).$$

Because

$$\lim_{A\to\infty} \lim_{n\to\infty} P\left(\left| \int_0^\infty Z_n(t) dq(t) - \int_0^A Z_n(t) dq(t) \right| > \varepsilon \right)$$
$$\leq \lim_{A\to\infty} \frac{1}{\varepsilon^2} \mathrm{Var}\left(\int_A^\infty Z_n(t) dq(t) \right) \leq \lim_{A\to\infty} \frac{1}{\varepsilon^2} M \left(\int_A^\infty (te^{-t})^{1/2} dq(t) \right)^2 = 0,$$

it follows according to theorem 4.2 in Billingsley (1968) that

$$q(0) Z_n(0) + \int_0^\infty Z_n(t) dq(t) \xrightarrow{D} q(0) Z(0) + \int_0^\infty Z(t) dq(t). \tag{6}$$

The variance of the limiting distribution can now be calculated using the covariance function of $Z(t)$. That the random variable on the right hand side of Equation (6) is normally distributed, follows since it is an integral of a normal process. ∎

Proposition 3. *Assume the assumptions of Proposition 2 are valid. Substitute the truncated function $u_N(\xi_i)$ with $u(\xi_i)$ and suppose $\tau^2 = Eu^2(\xi_i) < \infty$. Then*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} q(z_{i,n}(\theta_0))u(\xi_i) \xrightarrow{D} \mathcal{N}(0, \tau^2\sigma_q^2). \tag{7}$$

Proof. Let $\Delta_N(\xi_i) = u(\xi_i) - u_N(\xi_i)$ and $z_i = z_{i,n}(\theta_0)$. Then

$$E\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(z_i)\Delta_N(\xi_i)\right]^2 = E[q^2(z_1)\Delta_N^2(\xi_1)] + (n-1)E\left[q(z_1)\Delta_N(\xi_1)q(z_2)\Delta_N(\xi_2)\right].$$

To prove that the covariance term converges to 0 as $N \to \infty$, we use the conditional approach of Schilling (1986). Let $\{\mathrm{NN}_1 = \xi_2\}$ denote the event that the nearest neighbor of ξ_1 is ξ_2. Consider the following five mutually exclusive sets for various nearest neighbor geometries of ξ_1 and ξ_2:

$$D_1 = \{\mathrm{NN}_1 = \xi_2,\ \mathrm{NN}_2 = \xi_1\},\quad D_2 = \{\mathrm{NN}_1 = \mathrm{NN}_2\},\quad D_3 = \{\mathrm{NN}_1 = \xi_2,\ \mathrm{NN}_2 \neq \xi_1\},$$
$$D_4 = \{\mathrm{NN}_1 \neq \xi_2,\ \mathrm{NN}_2 = \xi_1\},\quad D_5 = \{\mathrm{NN}_1 \neq \xi_2,\ \mathrm{NN}_2 \neq \xi_1,\ \mathrm{NN}_1 \neq \mathrm{NN}_2\}.$$

Then,

$$E\left[q(z_1)\Delta_N(\xi_1)q(z_2)\Delta_N(\xi_2)\right] = P(D_5)E[q(z_1)\Delta_N(\xi_1)q(z_2)\Delta_N(\xi_2)|D_5] + \sum_{i=1}^{4} P(D_i)E[q(z_1)\Delta_N(\xi_1)q(z_2)\Delta_N(\xi_2)|D_i].$$

Given D_5, we have independence, therefore the covariance is 0. Since for $i = 1, \dots, 4$, $P(D_i) = \mathcal{O}(1/n)$, it is sufficient to show that the conditional expectations tend to 0 as $N \to \infty$, $i = 1, \dots, 4$. We have

$$|E[q(z_1)\Delta_N(\xi_1)q(z_2)\Delta_N(\xi_2)|D_i]| \le E[q^2(z_1)\Delta_N^2(\xi_1)|D_i]$$
$$= E[q^2(z_1)\Delta_N^2(\xi_1)] = E[q^2(z_1)]E[\Delta_N^2(\xi_1)] \to 0 \quad \text{as} \quad N \to \infty.$$

Thus, theorem 4.2 in Billingsley (1968) implies Equation (7). ∎

3 | ASYMPTOTIC NORMALITY OF THE DERIVATIVE OF THE GMSP FUNCTION

In Proposition 3, we proved asymptotic normality for a general function $u(\xi_i)$ satisfying $Eu(\xi_i) = 0$ and $Eu^2(\xi_i) < \infty$. Since any linear combination of such functions has also expectation 0 and a finite second moment, we can use Proposition 3 for proving asymptotic normality of our random vector of interest.

Proposition 4. *Suppose that $q(t) = h'(t)t$ satisfies the conditions of Proposition 2. Assume that $E[(\partial f_\theta(\xi_i))_{\theta=\theta_0}/g(\xi_i)]^2 < \infty$ holds for all the partial derivatives and that the covariance matrix $I(\theta_0)$ is positive definite. Then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\psi}_n(\xi_i,\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h'(z_{i,n}(\theta_0))z_{i,n}(\theta_0)\frac{(\nabla f_\theta(\xi_i))_{\theta=\theta_0}}{g(\xi_i)},$$

converges in distribution to a normal distribution with mean zero and with covariance matrix $\sigma_q^2 I(\theta_0)$, where σ_q^2 is calculated as in Equation (5). Observe that σ_q^2 depends on function h since $q(t) = h'(t)t$.

Proof. Let $q(z_{i,n}(\theta_0)) = h'(z_{i,n}(\theta_0))z_{i,n}(\theta_0)$. Define $u(\xi_i)$ as

$$u(\xi_i) = (t_1, \dots, t_q)(\nabla f_\theta(\xi_i))_{\theta=\theta_0}/g(\xi_i).$$

The assertion then follows from Proposition 3 by using the Cramér–Wold device. ∎

Proposition 5. *Consider a random vector $(z_{i,n}(\theta_0), X_n^T)^T$ such that $X_n \xrightarrow{p} X$ and the components of X are continuous functions of only ξ_i. Then*

$$\begin{pmatrix} z_{i,n}(\theta_0) \\ X_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} Z \\ X \end{pmatrix} \quad \text{with} \quad Z \sim Exp(1).$$

Proof. We have to show that for any Z-continuity set A_1 and any X-continuity set A_2,

$$P(z_{i,n}(\theta_0) \in A_1, X_n \in A_2) \to P(Z \in A_1, X \in A_2) = P(Z \in A_1)P(X \in A_2), \tag{8}$$

where the last equality holds since Z and ξ_i are independent. Since $X_n \xrightarrow{p} X$ and $z_{i,n}(\theta_0)$ is also independent of ξ_i, theorem 4.3 in Billingsley (1968) implies that Equation (8) is the same as

$$P(z_{i,n}(\theta_0) \in A_1, X \in A_2) = P(z_{i,n}(\theta_0) \in A_1)P(X \in A_2) \to P(Z \in A_1)P(X \in A_2).$$

∎

Lemma 1. *Suppose assumptions*

(i) A2, A3, A4 hold and consider function h_1, h_2, or h_5 with $\alpha \in (0,1)$;
(ii) A2, A3 hold and consider h_5 with $\alpha \in (1,2]$.

Then both for (i) and (ii), $\lambda_n(\theta) \to \lambda(\theta)$ uniformly for $\theta \in \Theta_0$ as $n \to \infty$.

The proof of Lemma 1 is given in the Appendix. Suppose that $-|\lambda(\theta)|$ has a unique maximum at θ_0 in Θ_0. This holds for example under the following weak identifiability condition:

$$\mu\{x : f_\theta(x) \neq f_{\theta_0}(x)\} > 0 \quad \text{for} \quad \theta \in \Theta_0, \tag{9}$$

where μ is Lebesgue measure. Then it follows from Lemma 1 that $\theta_0^{(n)} \to \theta_0$. In the following we assume that Equation (9) is fulfilled.

Proposition 6. *Suppose assumptions*

(i) A2, A3, A4, A5 are fulfilled and consider h_1 or h_5 with $\alpha \in (0,1)$;

(ii) A2, A3, A4, A5, A7 are fulfilled and consider h_2;
(iii) A2, A3, A7 are fulfilled and consider h_5 with $\alpha \in (1, 2]$.

Then for (i), (ii), and (iii),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_n(\xi_i, \theta_0^{(n)}) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\psi}_n(\xi_i, \theta_0) \xrightarrow{p} 0.$$

Proof. Since we are considering convergence in probability, there is no restriction to assume that the studied parameter is one-dimensional (corresponds to looking at the components separately). Let $A_{i,n} = \psi_n(\xi_i, \theta_0^{(n)}) - \tilde{\psi}_n(\xi_i, \theta_0)$. Observe that $E(A_{i,n}) = 0$ and

$$E\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} A_{i,n}\right)^2 = E(A_{1,n}^2) + (n-1)E(A_{1,n}A_{2,n}).$$

But $E(A_{1,n}A_{2,n}) = \sum_{i=1}^{5} P(D_i)E(A_{1,n}A_{2,n}|D_i)$, compare the proof of Proposition 3. Given D_5, the variables $A_{1,n}$ and $A_{2,n}$ are independent, thus $E(A_{1,n}A_{2,n}|D_5) = 0$. For $i = 1, \ldots, 4$, $P(D_i) = \mathcal{O}(1/n)$ and $E(A_{1,n}A_{2,n}|D_i) \leq E(A_{1,n}^2|D_i) = E(A_{1,n}^2)$. Therefore, the assertion follows if $E(A_{1,n}^2) \to 0$. Write $A_{1,n}$ as

$$A_{1,n} = h'(z_{1,n}(\theta_0^{(n)}))z_{1,n}(\theta_0)\frac{1}{P_{\theta_0}(B_n(\xi_1)}\int_{B_n(\xi_1)} \partial f_{\theta_0^{(n)}}(y)dy - h'(z_{1,n}(\theta_0))z_{1,n}(\theta_0)\frac{\partial f_{\theta_0}(\xi_1)}{g(\xi_1)},$$

and recall that

$$z_{1,n}(\theta_0^{(n)}) = z_{1,n}(\theta_0)\frac{1}{P_{\theta_0}(B_n(\xi_1))}\int_{B_n(\xi_1)} f_{\theta_0^{(n)}}(y)dy.$$

Proposition 5 together with Lemma 1 imply that $A_{1,n} \xrightarrow{D} 0$ and $A_{1,n}^2 \xrightarrow{D} 0$. Thus, $E(A_{1,n})^2 \to 0$ follows because under our assumptions the random variables $\psi_n^2(\xi_1, \theta_0^{(n)})$ and $\tilde{\psi}_n^2(\xi_1, \theta_0)$ are uniformly integrable. ■

4 | ASYMPTOTIC NORMALITY OF GMSP ESTIMATE VIA STOCHASTIC DIFFERENTIABILITY

To prove asymptotic normality of $\hat{\theta}_n$, we need to use a stochastic differentiability condition similar to Pollard (1985) and Huber (1967). We will prove that

$$\sup_{\Theta_n} \frac{\left|\frac{1}{\sqrt{n}} \sum_i (\psi_n(\xi_i, \theta) - \psi_n(\xi_i, \theta_0^{(n)})) - \sqrt{n}\lambda_n(\theta)\right|}{1 + \sqrt{n}|\theta - \theta_0{}^{(n)}|} \xrightarrow{p} 0, \tag{10}$$

where Θ_n is a compact set shrinking to θ_0 as $n \to \infty$.

To prove Equation (10), we will consider the numerator of the expression in Equation (10) separately on a compact set $K \subset \mathbb{R}^d$ and its complement K^c, and show that the contribution from

K^c is arbitrarily small when choosing K large enough. Let

$$\psi_n(\xi_i,\theta,K)=\psi_n(\xi_i,\theta)I_K(\xi_i),\quad \psi_n(\xi_i,\theta,K^c)=\psi_n(\xi_i,\theta)I_{K^c}(\xi_i),$$
$$E[\psi_n(\xi_i,\theta,K)]=\lambda_n(\theta,K),\quad E[\psi_n(\xi_i,\theta,K^c)]=\lambda_n(\theta,K^c).$$

Consider the following decomposition of the numerator in Equation (10):

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta)-\psi_n(\xi_i,\theta_0^{(n)}))-\sqrt{n}\lambda_n(\theta)\\&=\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta,K)-\psi_n(\xi_i,\theta_0^{(n)},K))-\sqrt{n}(\lambda_n(\theta,K)-\lambda_n({\theta_0}^{(n)},K))\\&+\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta,K^c)-\psi_n(\xi_i,\theta_0^{(n)},K^c))-\sqrt{n}(\lambda_n(\theta,K^c)-\lambda_n({\theta_0}^{(n)},K^c)).\end{aligned}$$

We are going to show the following:

(1) $\forall\varepsilon>0$, a compact set $K\subset\mathbb{R}^d$ can be chosen so that for large n,

$$P\left(\sup_{\Theta_n}\frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta,K^c)-\psi_n(\xi_i,\theta_0^{(n)},K^c))-\sqrt{n}(\lambda_n(\theta,K^c)-\lambda_n({\theta_0}^{(n)},K^c))\right|}{1+\sqrt{n}|\theta-{\theta_0}^{(n)}|}>\varepsilon\right)<\varepsilon, \tag{11}$$

(2) for any compact set $K\subset\mathbb{R}^d$,

$$\sup_{\Theta_n}\frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta,K)-\psi_n(\xi_i,\theta_0^{(n)},K))-\sqrt{n}(\lambda_n(\theta,K)-\lambda_n({\theta_0}^{(n)},K))\right|}{1+\sqrt{n}|\theta-{\theta_0}^{(n)}|}\xrightarrow{p}0. \tag{12}$$

Therefore, Equations (11) and (12) together imply Equation (10).

Let $V_n(\theta)$ denote the following matrix of partial derivatives:

$$V_n(\theta)=(V_n^{(j,l)}(\theta))=\left(\frac{\partial\lambda_{n,j}(\theta)}{\partial\theta_l}\right),\quad j,l=1,\dots,q,$$

where $\lambda_{n,j}(\theta)$ is the j-th element of the vector $\lambda_n(\theta)$. Recall that $\psi_{n,j}(\xi,\theta)$ and $\psi_j(\xi,\theta)$ denote the j-th component of the vectors $\psi_n(\xi,\theta)$ and $\psi(\xi,\theta)$, respectively. Let $V(\theta)=(V^{(j,l)}(\theta))$ with $V^{(j,l)}(\theta)=E\left[\frac{\partial}{\partial\theta_l}\psi_j(\xi,\theta)\right], j,l=1,\dots,q$.

Lemma 2. *Suppose assumptions*

(i) A6, A8, A9 hold and consider h_1 or h_5 with $\alpha\in(0,1)$;
(ii) A6, A8, A10 hold and consider h_2;
(iii) A3, A10 hold and consider h_5 with $\alpha\in(1,2]$.

Then for (i), (ii), and (iii), the following assertions hold. In a neighborhood of θ_0, $\lambda_n(\theta)$ is continuously differentiable. Furthermore, $V_n(\theta)\to V(\theta)$ uniformly for $\theta\in\Theta_0$ as $n\to\infty$.

The proof of Lemma 2 is given in the Appendix.

Proposition 7. *Suppose the assumptions of Lemma 2 hold. Then* $\forall \varepsilon > 0$ *a compact set* $K \subset \mathbb{R}^d$ *can be chosen so that Equation* (11) *holds for large* n *for* h_1, h_2, *and* h_5 *with* $\alpha \in (0,1) \cup (1,2]$.

Proof. Since ψ_n is a vector and the Euclidean norm of a vector is smaller than the sum of the absolute values of its components, it is equivalent to work with single components of the vector and show that the contribution from each component is small. Applying the mean value theorem we obtain:

$$\frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_{n,j}(\xi_i,\theta,K^c) - \psi_{n,j}(\xi_i,\theta_0^{(n)},K^c)) - \sqrt{n}(\lambda_{n,j}(\theta,K^c) - \lambda_{n,j}({\theta_0}^{(n)},K^c))\right|}{1+\sqrt{n}|\theta - {\theta_0}^{(n)}|}$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{q}\left|\frac{\partial}{\partial\theta_l}\psi_{n,j}(\xi_i,\tilde{\theta},K^c) - V_n^{(j,l)}(\tilde{\theta},K^c)\right|$$
$$\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{q}\sup_{\Theta_n}\left|\frac{\partial}{\partial\theta_l}\psi_{n,j}(\xi_i,\theta,K^c) - V_n^{(j,l)}(\theta,K^c)\right|.$$

Thus,

$$P\left(\sup_{\Theta_n}\frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_{n,j}(\xi_i,\theta,K^c) - \psi_{n,j}(\xi_i,\theta_0^{(n)},K^c)) - \sqrt{n}(\lambda_{n,j}(\theta,K^c) - \lambda_{n,j}({\theta_0}^{(n)},K^c))\right|}{1+\sqrt{n}|\theta - {\theta_0}^{(n)}|} > \varepsilon\right)$$
$$\leq \frac{1}{\varepsilon}\sum_{l=1}^{q}E\left[\sup_{\Theta_n}\left|\frac{\partial}{\partial\theta_l}\psi_{n,j}(\xi_i,\theta,K^c) - V_n^{(j,l)}(\theta,K^c)\right|\right]$$
$$\leq \frac{2}{\varepsilon}\sum_{l=1}^{q}E\left[\sup_{\Theta_n}\left|\frac{\partial}{\partial\theta_l}\psi_{n,j}(\xi_i,\theta,K^c)\right|\right] < \varepsilon,$$

if K is large and if $n > n_0$ for some n_0. ∎

To prove Equation (12), we will use lemma 4 in Pollard (1985), which is based on bracketing technique, see van der Vaart (2000) and Pollard (1985). The bracketing condition enables to divide the parameter set of interest into a finite number of subsets and study the supremum of interest over a finite number of smaller parameter sets. We need also to use the following property of the radii of our nearest neighbor balls: $R_n(i) \overset{a.s.}{\to} 0$ for every i. Therefore, according to Egoroff's theorem there exists for each i a set A_i with $P(A_i) > 1 - \frac{\varepsilon}{2}2^{-i}$ such that $R_n(i) \to 0$ uniformly on A_i. Therefore, we can define a set $A = \cap_{i=1}^{\infty} A_i$, such that $P(A^c) < \varepsilon/2$.

Bracketing. Lemma 4 in Pollard (1985) will be applied to functions in

$$\mathcal{F}_n = \{[\psi_n(\xi_i,\theta,K) - \psi_n(\xi_i,\theta_0^{(n)},K)]I_A, \quad |\theta - \theta_0| \leq \alpha_n, \ i = 1,\ldots,n\}.$$

Since $[\psi_n(\xi_i,\theta,K) - \psi_n(\xi_i,\theta_0^{(n)},K)]I_A$, $i = 1,\ldots,n$, are identically distributed, we can suppress i in ξ_i, $B_n(\xi_i)$, $z_{i,n}(\theta)$, and $R_n(i)$ right now. That the bracketing condition is fulfilled follows since the functions $[\psi_n(\xi,\theta,K) - \psi_n(\xi,\theta_0^{(n)},K)]I_A$ satisfy a Lipschitz condition

$$|\psi_n(\xi,\theta_s,K) - \psi_n(\xi,\theta_t,K)|I_A \leq H_n(\xi,K)|\theta_s - \theta_t|,$$

where

$$H_n(\xi, K) = q \left[\sup_{\Theta_n} |h''(z_n(\theta)) z_n^2(\theta)| (f_{max}^{(1)}(\xi))^2 + \sup_{\Theta_n} |h'(z_n(\theta)) z_n(\theta)| f_{max}^{(2)}(\xi) \right] I_A I_K(\xi),$$

with

$$f_{max}^{(1)}(\xi) = \max_l \sup_{\Theta_n} \left(\frac{1}{P_\theta(B_n)} \int_{B_n} \left| \frac{\partial f_\theta(y)}{\partial \theta_l} \right| dy \right),$$

$$f_{max}^{(2)}(\xi) = \max_{j,l} \sup_{\Theta_n} \left(\frac{1}{P_\theta(B_n)} \int_{B_n} \left| \frac{\partial^2 f_\theta(y)}{\partial \theta_j \partial \theta_l} \right| dy \right),$$

and where for some constant $b_1(h, K)$, $E[H_n(\xi, K)] < b_1(h, K) < \infty$ when n is large enough.

Proposition 8. *Consider a compact set $K \subset \mathbb{R}^d$. Suppose assumption A2 is fulfilled. Then the family $\mathcal{F}_n$ satisfies the bracketing condition and $EH_n^2(\xi, K) < b_2(h, K) < \infty$ holds for some constant $b_2(h, K)$ and for large n. Therefore, the convergence in Equation (12) holds for h_1, h_2, and h_5 with $\alpha \in (0, 1) \cup (1, 2]$.*

Proof. For the bracketing condition to be fulfilled we need to show that $E[H_n(\xi, K)] < b_1(h, K) < \infty$. Since this follows from $E[H_n(\xi, K)]^2 < b_2(h, K) < \infty$, we are going to prove that

$$E\left[\sup_{\Theta_n} |h''(z_n(\theta)) z_n^2(\theta)| [f_{max}^{(1)}(\xi)]^2 \cdot I_A I_K(\xi) \right]^2 < \infty, \tag{13}$$

$$E\left[\sup_{\Theta_n} |h'(z_n(\theta)) z_n(\theta)| f_{max}^{(2)}(\xi) \cdot I_A I_K(\xi) \right]^2 < \infty. \tag{14}$$

Define the closed δ-neighborhood $K_\delta = \{x \in \mathbb{R}^d : d(x, K) \le \delta\}$, where $d(x, K) = \inf\{d(x, y) : y \in K\}$. When $\xi \in K$ and $\omega \in A$, we have for n large enough that $B_n \subset K_\delta$. Therefore, for large n,

$$\sup_{\Theta_n} \left(\frac{1}{P_\theta(B_n)} \int_{B_n} \frac{|\partial f_\theta(y)|}{f_\theta(y)} dP_\theta(y) \right)^4 I_K(\xi) I_A \le \sup_{\Theta_n} \left(\frac{|\partial f_\theta(\xi)|}{f_\theta(\xi)} + \eta \right)^4 I_K(\xi) I_A \le N_K,$$

where the last inequality holds because $|\partial f_\theta(y)|/f_\theta(y)$ is uniformly continuous on $K \times \Theta_n$. In a similar way we obtain

$$\left(\sup_{\Theta_n} \frac{1}{P_\theta(B_n)} \int_{B_n} |\partial^2 f_\theta(y)| dy \cdot I_K(\xi) I_A \right)^2 \le \sup_{\Theta_n} \left(\frac{|\partial^2 f_\theta(\xi)|}{f_\theta(\xi)} + \eta \right)^2 I_K(\xi) I_A < M_K.$$

Since $z_n(\theta) = z_n(\theta_0) \cdot \frac{1}{P_{\theta_0}(B_n)} \int_{B_n} f_\theta(y) dy$ and $z_n(\theta_0)$ has moments of all orders, Equations (13) and (14) follow for our functions h_1, h_2, and h_5. The finite expectation $EH_n^2(\xi, K) < b_2(h, K)$ implies

$$E\left[\sup_{\Theta_n} |\psi_n(\xi_i, \theta, K) - \psi_n(\xi_i, \theta_0^{(n)}, K)|^2 I_A \right] \le E[H_n(\xi_i, K)]^2 \alpha_n^2 \to 0 \quad \text{as} \quad n \to \infty,$$

and that the bracketing functions have finite variance. Moreover, in the same way as in the proof of Proposition 6 it can be shown that

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_i H_n(\xi_i, K)\right) = \mathcal{O}(\mathrm{Var}(H_n(\xi_1, K))),$$

see also p. 306 and the proof of lemma 4 in Pollard (1985). As $P(A^c) < \varepsilon/2$, Equation (12) follows due to lemma 4 in Pollard (1985). ∎

Proof of Theorem 1. As the last step we will use lemma 3 in Huber (1967) and prove the asymptotic normality of $\hat{\theta}_n$.

Propositions 7 and 8 imply

$$\sup_{\Theta_n} \frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta) - \psi_n(\xi_i,\theta_0^{(n)})) - \sqrt{n}\lambda_n(\theta)\right|}{1+\sqrt{n}|\theta-\theta_0^{(n)}|} \xrightarrow{p} 0.$$

Applying the mean value theorem to $\lambda_{n,j}(\theta)$ we obtain that there exists $\tilde{\theta}_{n,j}$ such that

$$\lambda_{n,j}(\theta) = V_n^{(j)}(\tilde{\theta}_{n,j})(\theta - \theta_0^{(n)}), \quad j = 1, \dots, q,$$

where $V_n^{(j)}(\cdot)$ denotes the j-th row of the matrix $V_n(\cdot)$. Define a matrix $V_n^*(\theta)$, where the rows are given by $V_n^{(j)}(\tilde{\theta}_{n,j})$, $j = 1, \dots, q$. We use the argument θ in V_n^* to indicate that it comes from an application of the mean value theorem to the components of $\lambda_n(\theta)$. As $V_n^*(\theta) \to V(\theta_0) = b_h I(\theta_0)$ uniformly for $\theta \in \Theta_n$, it follows that $V_n^*(\theta)$ is invertible for large n and $\theta \in \Theta_n$. Therefore, for some $C > 0$,

$$|\theta - \theta_0^{(n)}| \le |V_n^*(\theta)^{-1}||\lambda_n(\theta)| \le (|V(\theta_0)^{-1}| + \varepsilon)|\lambda_n(\theta)| = C|\lambda_n(\theta)|.$$

It follows that

$$\sup_{\Theta_n} \frac{\left|\frac{1}{\sqrt{n}}\sum_i(\psi_n(\xi_i,\theta) - \psi_n(\xi_i,\theta_0^{(n)})) - \sqrt{n}\lambda_n(\theta)\right|}{1+\sqrt{n}|\lambda_n(\theta)|} \xrightarrow{p} 0,$$

which corresponds to lemma 3 in Huber (1967). Applying theorem 3 in Huber (1967) and using the consistency of $\hat{\theta}_n$ gives

$$\frac{1}{\sqrt{n}}\sum_i \psi_n(\xi_i, \theta_0^{(n)}) + \sqrt{n}\lambda_n(\hat{\theta}_n) \xrightarrow{p} 0,$$

where $\frac{1}{\sqrt{n}}\sum_i \psi_n(\xi_i, \theta_0^{(n)})$ is asymptotically normally distributed with mean zero and covariance matrix $\sigma_q^2 I(\theta_0)$. It follows that $\sqrt{n}\lambda_n(\hat{\theta}_n) \xrightarrow{D} \mathcal{N}(0, \sigma_q^2 I(\theta_0))$. Applying the mean value theorem again gives that for some $\tilde{\theta}_{n,1}, \dots, \tilde{\theta}_{n,q}$ depending on $\hat{\theta}_n$, we can define a matrix $V_n^*(\hat{\theta}_n)$ so that $\sqrt{n}\lambda_n(\hat{\theta}_n) = V_n^*(\hat{\theta}_n)\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})$. As $\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)}) = V_n(\theta_0)^{-1}V_n(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})$, we obtain

$$\begin{aligned}|\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})| &\le |V_n(\theta_0)^{-1}|(|V_n^*(\hat{\theta}_n)\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})| + |V_n(\theta_0) - V_n^*(\hat{\theta}_n)||\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})|)\\ &\le |V_n(\theta_0)^{-1}|\left(\mathcal{O}_P(1) + o_P(1)\sqrt{n}|\hat{\theta}_n - \theta_0^{(n)}|\right).\end{aligned}$$

Thus, $|\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)})| \leq \mathcal{O}_P(1)$. It follows that

$$\sqrt{n}\lambda_n(\hat{\theta}_n) = V_n(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)}) + o_P(1) = V(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)}) + o_P(1),$$

and thus,

$$\sqrt{n}(\hat{\theta}_n - \theta_0^{(n)}) \xrightarrow{D} \mathcal{N}(0, V(\theta_0)^{-1}\sigma_q^2 I(\theta_0)[V(\theta_0)^{-1}]^T).$$

Using that $V(\theta_0) = b_h I(\theta_0)$, the assertion follows. ∎

5 | DISCUSSION

For univariate spacings, asymptotic normality of GMSP estimators has been shown in Ghosh and Jammalamadaka (2001). Recently, Luong (2018) also considered consistency and asymptotic normality of univariate GMSP estimates. Since the author has overlooked the local dependence between nearest neighbors, the proof of asymptotic normality in Luong (2018) is not correct and thus also the derived asymptotic variance is incorrect. Ghosh and Jammalamadaka (2001) showed that the smallest variance in the asymptotic distribution was obtained for $h(x) = \ln(x)$ and that this smallest variance coincides with the Cramér–Rao lower bound. We have calculated the constants σ_q^2/b_h^2 in the asymptotic covariance matrix for the h-functions studied in this article, see Table 1. The smallest variance is obtained for $h_1(x) = \ln\ x - x + 1$. For $h_5(x) = \mathrm{sgn}(1-\alpha)(x^\alpha - \alpha x + \alpha - 1)$, the variance increases with increasing values of α and when $\alpha \searrow 0$, the variance tends to the variance of h_1.

In this article, we have proved asymptotic normality of the GMSP estimate $\hat{\theta}_n$ around $\theta_0^{(n)}$, where $\theta_0^{(n)}$ maximizes the expectation function $E[h(z_{i,n}(\theta))]$. For the asymptotic normality to hold around θ_0, it has to be shown that $\sqrt{n}(\theta_0^{(n)} - \theta_0) \to 0$. According to the mean value theorem, for some constant $C > 0$,

$$\sqrt{n}|\theta_0^{(n)} - \theta_0| \leq C\sqrt{n}|\lambda_n(\theta_0)|.$$

Thus, if $\sqrt{n}|\lambda_n(\theta_0)| \to 0$, then $\sqrt{n}(\theta_0^{(n)} - \theta_0) \to 0$ follows. The behavior of $\lambda_n(\theta_0)$ depends on what parameters are considered. In the case of multivariate normal distribution $\mathcal{N}_d(\mu_0, \Sigma_0)$ it follows because of symmetry that $E[h(z_{i,n}(\theta))]$ is maximized by μ_0 regardless of Σ_0. Thus, $\mu_0^{(n)} = \mu_0$. For

TABLE 1 Constants σ_q^2/b_h^2 in the asymptotic covariance matrix for different h-functions

$h(x)$		σ_q^2/b_h^2
$h_1(x) = \ln\ x - x + 1$		1.8434
$h_2(x) = (1-x)\ln\ x$		2.2130
$h_5(x) = \mathrm{sgn}(1-\alpha)(x^\alpha - \alpha x + \alpha - 1)$	$\alpha = 0.1$	1.9265
	$\alpha = 0.5$	2.3421
	$\alpha = 0.9$	2.7493
	$\alpha = 2$	3.6546

TABLE 2 Estimated values of $\sqrt{n}\lambda_n(\theta_0)$ for the component corresponding to σ_1 in the case of bivariate normal distribution with the parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = (1, 2, 1, 1, 0.5)$

$\sqrt{n}$	$h_1(x) = \ln x - x + 1$	$h_2(x) = (1 - x) \ln x$	$h_3(x) = -(1 - \sqrt{x})^2$
10	0.3910	1.1613	0.2769
30	0.3641	1.0580	0.2514
40	0.3399	1.0025	0.2368
60	0.2796	0.8124	0.1924
100	0.2213	0.6450	0.1527
200	0.1576	0.4673	0.1101
500	0.1233	0.3504	0.0836
700	0.0382	0.1193	0.0282

bivariate normal distribution with $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, we have studied the behavior of $\lambda_n(\theta_0)$ in simulation studies for the following parameter vector: $\theta_0 = (1, 2, 1, 1, 0.5)$. We simulated a sample of n observations from this distribution and calculated for a randomly chosen observation in the sample the quantity

$$h'(z_{i,n}(\theta_0))z_{i,n}(\theta_0)\frac{1}{P_{\theta_0}(B_n(\xi_i))}\int_{B_n(\xi_i)}(\nabla f_\theta(y))_{\theta=\theta_0}dy.$$

This procedure was repeated for $m = 10,000$ samples and $\lambda_n(\theta_0)$ was estimated with the average of the 10,000 values. In Table 2, the mean values over 20 repetitions are presented for the component of $\sqrt{n}\lambda_n(\theta_0)$ that corresponds to σ_1. We calculated the values for h_1, h_2, and h_3 with $p = 2$. It can be seen in Table 2 that for all the considered h-functions the estimated values of $\sqrt{n}\lambda_n(\theta_0)$ for the component corresponding to σ_1 decrease when n increases and approach 0 slowly. The same behavior can be observed for σ_2 and the correlation parameter ρ. Thus, the simulation results indicate that $\sqrt{n}\lambda_n(\theta_0) \to 0$ as $n \to \infty$ for the components corresponding to σ_1, σ_2, and ρ, although the convergence is very slow.

ACKNOWLEDGMENTS

This work is supported by Estonian institutional research funding IUT34-5. The authors would like to thank the reviewers for the comments that helped to improve the presentation of the article.

ORCID

Kristi Kuljus https://orcid.org/0000-0002-8690-5032

REFERENCES

Bickel, P. J., & Breiman, L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *Annals of Probability*, *11*(1), 185–214.

Billingsley, P. (1968). *Convergence of probability measures*. New York, NY: Wiley.

Cheng, R. C., & Amin, N. A. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *45*(3), 394–403.

Cramér, H., & Leadbetter, M. R. (1967). *Stationary and related stochastic processes*. New York, NY: Wiley.

Ekström, M. (2001). Consistency of generalized maximum spacing estimates. *Scandinavian Journal of Statistics*, *28*, 343–354.

Ghosh, K., & Jammalamadaka, S. R. (2001). A general estimation method using spacings. *Journal of Statistical Planning and Inference*, *93*(1–2), 71–82.

Huber, P. J. (1967). *The behaviour of maximum likelihood estimates under nonstandard conditions*. In *Fifth Berkeley symposium on mathematical statistics and probability* (pp. 221–233). Berkeley, CA: University of California.

Kuljus, K., & Ranneby, B. (2015). Generalized maximum spacing estimation for multivariate observations. *Scandinavian Journal of Statistics*, *42*, 1092–1108.

Luong, A. (2018). Unified asymptotic results for maximum spacing and generalized spacing methods for continuous models. *Open Journal of Statistics*, *8*, 614–639.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, *1*, 295–313.

Ranneby, B. (1984). The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, *11*(2), 93–112.

Ranneby, B., & Ekström, M. (1997). *Maximum spacing estimates based on different metrics*. Research Report: Umeå University.

Ranneby, B., Jammalamadaka, S. R., & Teterukovskiy, A. (2005). The maximum spacing estimation for multivariate observations. *Journal of Statistical Planning and Inference*, *129*(1–2), 427–446.

Schilling, M. F. (1983). Goodness of fit testing in $\mathbb{R}^m$ based on the weighted empirical distribution of certain nearest neighbor statistics. *Annals of Statistics*, *11*(1), 1–12.

Schilling, M. F. (1986). Mutual and shared neighbor probabilities: Finite- and infinite-dimensional results. *Advances in Applied Probability*, *18*(2), 388–405.

van der Vaart, A. W. (2000). *Asymptotic statistics*. New York, NY: Cambridge University Press.

Zhou, S., & Jammalamadaka, S. R. (1993). Goodness of fit in multidimensions based on nearest neighbour distance. *Nonparametric Statistics*, *2*, 271–284.

How to cite this article: Kuljus K, Ranneby B. Asymptotic normality of generalized maximum spacing estimators for multivariate observations. *Scand J Statist*. 2020;47:968–989. https://doi.org/10.1111/sjos.12436

APPENDIX

Proof of Lemma 1

Proof. Recall that $\lambda_n(\theta) = E[\psi_n(\xi_i, \theta)]$ and $\lambda(\theta) = E[\psi(\xi_i, \theta)]$. Thus, we can suppress i in the notation and write $\psi_n(\xi, \theta)$, $\tilde{\psi}_n(\xi, \theta)$, $\psi(\xi, \theta)$, and $z_n(\theta_0)$ for the random quantities of interest. Since uniform convergence can be proved componentwise, we will write $\partial f_\theta(\xi)$ to emphasize that the same approach holds for any component $\partial/\partial\theta_j, j = 1, \ldots, q$. We suppress j also in vector notations $\psi_{n,j}$, ψ_j, etc. The uniform convergence of $\lambda_n(\theta)$ to $\lambda(\theta)$ in Θ_0 holds if

$$\sup_{\theta\in\Theta_0} |\lambda_n(\theta) - \lambda(\theta)| \to 0 \quad \text{as} \quad n \to \infty.$$

Thus, we will study

$$\sup_{\Theta_0} |E[\psi_n(\xi, \theta) - \psi(\xi, \theta)]| \leq \sup_{\Theta_0} |E[\psi_n(\xi, \theta) - \tilde{\psi}_n(\xi, \theta)]|$$
$$+ \sup_{\Theta_0} |E[\tilde{\psi}_n(\xi, \theta) - \psi(\xi, \theta)]| = \text{Term}_{\text{I}} + \text{Term}_{\text{II}}.$$

We will show that both terms converge to 0 under the assumptions of the lemma.

Term$_{\text{II}}$. Observe that

$$\tilde{\psi}_n(\xi,\theta)-\psi(\xi,\theta)=\left[\nu\left(z_n(\theta_0)\frac{f_\theta(\xi)}{g(\xi)}\right)-\nu\left(Z\frac{f_\theta(\xi)}{g(\xi)}\right)\right]\frac{\partial f_\theta(\xi)}{f_\theta(\xi)}.$$

We will exemplify the proof using h_3, the proof is similar for other h-functions. For h_3 we have $h_3'(z)z=-z+\sqrt{z}$, therefore

$$\tilde{\psi}_n(\xi,\theta)-\psi(\xi,\theta)=[Z-z_n(\theta_0)]\frac{f_\theta(\xi)}{g(\xi)}\frac{\partial f_\theta(\xi)}{f_\theta(\xi)}+\left[\sqrt{z_n(\theta_0)}-\sqrt{Z}\right]\sqrt{\frac{f_\theta(\xi)}{g(\xi)}}\frac{\partial f_\theta(\xi)}{f_\theta(\xi)}.$$

Because $\sqrt{f_\theta(\xi)}/\sqrt{g(\xi)}\le 1+f_\theta(\xi)/g(\xi)$, we obtain

$$\begin{aligned}\sup_{\Theta_0}|E[\tilde{\psi}_n(\xi,\theta)-\psi(\xi,\theta)]|&\le|E(z_n(\theta_0)-Z)|\sup_{\Theta_0}E\left[\frac{|\partial f_\theta(\xi)|}{g(\xi)}\right]\\&+\left|E\left(\sqrt{z_n(\theta_0)}-\sqrt{Z}\right)\right|\sup_{\Theta_0}E\left[\frac{|\partial f_\theta(\xi)|}{f_\theta(\xi)}+\frac{|\partial f_\theta(\xi)|}{g(\xi)}\right].\end{aligned}$$

Since $z_n(\theta_0)\xrightarrow{D}Z$ and $\sqrt{z_n(\theta_0)}\xrightarrow{D}\sqrt{Z}$, and since the corresponding moments of $z_n(\theta_0)$ and Z are finite, Term$_{\text{II}}\to 0$ follows because of convergence of the respective expected values and because of assumptions A3 and A4.

Term$_{\text{I}}$. According to Proposition 5, $\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)\xrightarrow{D}0$ for each θ. This implies that for any $\theta_1,\dots,\theta_m\in\Theta_0$, where m is any finite number, the respective finite-dimensional distribution converges to a zero-vector of length m in distribution. Observe that $\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)$ are continuous functions of θ in a neighborhood of θ_0. We will prove tightness of $\{\tilde{\psi}_n(\xi,\theta)\}$ and $\{\psi_n(\xi,\theta)\}$ and use that this together with the convergence of finite-dimensional distributions implies

$$\sup_{\Theta_0}|\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)|\xrightarrow{D}0.$$

Since $\sup_{\Theta_0}|\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)|$ is uniformly integrable due to our assumptions, we then obtain

$$\sup_{\Theta_0}|E[\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)]|\le E\left[\sup_{\Theta_0}|\psi_n(\xi,\theta)-\tilde{\psi}_n(\xi,\theta)|\right]\to 0.$$

Since both $\psi_n(\xi,\theta_0)$ and $\tilde{\psi}_n(\xi,\theta_0)$ converge to $\psi(\xi,\theta_0)$ in distribution, $\psi_n(\xi,\theta_0)$ and $\tilde{\psi}_n(\xi,\theta_0)$ are tight according to Prohorov's theorem. To show tightness of $\tilde{\psi}_n(\xi,\theta)$, take now arbitrary $\varepsilon>0$ and $\eta>0$. Choose a compact set $K\subset\mathbb{R}^d$ and a constant $M>0$ such that $P(\xi\in K)>1-\eta/4$, $\sup_n P(z_n(\theta_0)>M)<\eta/4$. Consider arbitrary θ_1,θ_2 in Θ_0. Applying the equality

$$a(\theta_1)b(\theta_1)-a(\theta_2)b(\theta_2)=(a(\theta_1)-a(\theta_2))\,b(\theta_1)+(b(\theta_1)-b(\theta_2))\,a(\theta_2),$$

we obtain

$$\begin{aligned}|\tilde{\psi}_n(\xi,\theta_1)-\tilde{\psi}_n(\xi,\theta_2)|\,I_K(\xi)I(z_n(\theta_0)\le M)&\le\left\{\left|\nu\left(z_n(\theta_0)\frac{f_{\theta_1}(\xi)}{g(\xi)}\right)-\nu\left(z_n(\theta_0)\frac{f_{\theta_2}(\xi)}{g(\xi)}\right)\right|\sup_{\Theta_0}\frac{|\partial f_\theta(\xi)|}{f_\theta(\xi)}\right.\\&\left.+\left|\frac{\partial f_{\theta_1}(\xi)}{f_{\theta_1}(\xi)}-\frac{\partial f_{\theta_2}(\xi)}{f_{\theta_2}(\xi)}\right|\sup_{\Theta_0}\left|\nu\left(z_n(\theta_0)\frac{f_\theta(\xi)}{g(\xi)}\right)\right|\right\}I_K(\xi)I(z_n(\theta_0)\le M).\end{aligned}$$

Since our functions of interest are uniformly continuous on $K \times \Theta_0$, we obtain that there exists $\delta_1 > 0$ such that

$$|\tilde{\psi}_n(\xi, \theta_1) - \tilde{\psi}_n(\xi, \theta_2)|\, I_K(\xi) I(z_n(\theta_0) \leq M) < \varepsilon,$$

whenever $|\theta_1 - \theta_2| < \delta_1$.

Tightness of $\{\psi_n(\xi, \theta)\}$ follows analogously, but now we also need to bring in the set A, where $R_n(\omega) \to 0$ uniformly. Therefore, if n is large enough and $\omega \in A \cap \{\xi \in K\} \cap \{z_n(\theta_0) \leq M\}$, there exists δ_2 such that whenever $|\theta_1 - \theta_2| < \delta_2$,

$$\left| \frac{1}{P_{\theta_0}(B_n)} \int_{B_n} \frac{f_{\theta_1}(y)}{g(y)} dP_{\theta_0}(y) - \frac{1}{P_{\theta_0}(B_n)} \int_{B_n} \frac{f_{\theta_2}(y)}{g(y)} dP_{\theta_0}(y) \right|,$$
$$\left| \frac{1}{P_{\theta_1}(B_n)} \int_{B_n} \partial f_{\theta_1}(y) dy - \frac{1}{P_{\theta_2}(B_n)} \int_{B_n} \partial f_{\theta_2}(y) dy \right|$$

become sufficiently small. ∎

Proof of Lemma 2

Proof. The assumptions of the lemma ensure uniform integrability of the random variables $\sup_{\Theta_0} \left| \frac{\partial}{\partial \theta_l} \psi_{n,j}(\xi, \theta) \right|$. Thus,

$$E \left[\sup_{\Theta_0} \left| \frac{\partial}{\partial \theta_l} \psi_{n,j}(\xi, \theta) \right| \right] < \infty. \tag{A1}$$

Therefore, we can differentiate under the integral sign and

$$\frac{\partial}{\partial \theta_l} \lambda_{n,j}(\theta) = E \left[\frac{\partial}{\partial \theta_l} \psi_{n,j}(\xi, \theta) \right].$$

Since $\frac{\partial}{\partial \theta_l} \psi_{n,j}(\xi, \theta)$ are continuous functions of θ, it follows from Equation (A1) and the Lebesgue dominated convergence theorem that $V_n(\theta)$ is continuous in θ. Proposition 5 implies $\frac{\partial}{\partial \theta_l} \psi_{n,j}(\xi, \theta) \xrightarrow{D} \frac{\partial}{\partial \theta_l} \psi_j(\xi, \theta)$. The uniform integrability gives $V_n(\theta) \to V(\theta)$ for every $\theta \in \Theta_0$. The uniform convergence of $V_n(\theta)$ can be proved in the same way as the uniform convergence of $\lambda_n(\theta)$ in Lemma 1. ∎