doi: 10.1111/j.1467-9469.2007.00574.x

 Vol 35: 186–192, 2008

A *Z*-theorem with Estimated Nuisance Parameters and Correction Note for 'Weighted Likelihood for Semiparametric Models and Two-phase Stratified Samples, with Application to Cox Regression'

NORMAN E. BRESLOW

Department of Biostatistics, University of Washington

JON A. WELLNER

Departments of Statistics and Biostatistics, University of Washington

ABSTRACT. We state and prove a limit theorem for estimators of a general, possibly infinite dimensional parameter based on unbiased estimating equations containing estimated nuisance parameters. The theorem corrects a gap in the proof of one of the assertions of our paper 'Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression' [*Scand. J. Statist*. 34 (2007) 86–102].

Key words: nuisance parameters, Z-theorem

1. Introduction

Breslow & Wellner (2007) cited a theorem of Pierce (1982) in deriving the asymptotic distribution of weighted likelihood estimators for parameters in semiparametric models fitted to two-phase stratified samples when sampling weights were estimated from the data. Li (2007, personal communication) noticed that one of Pierce's two key hypotheses had in fact not been established by us. In this note we develop a Z-theorem with estimated nuisance parameters that applies to infinite dimensional parameters and allows us to complete our earlier proof under a slight strengthening of our previous hypotheses. The derivations use empirical process techniques developed in van der Vaart & Wellner (1996) and related articles. In order to keep the exposition as short as possible, we assume familiarity with the notation and results in section 3.3 of van der Vaart & Wellner (1996) and section 6 of Breslow & Wellner (2007).

2. A *Z*-theorem with estimated nuisance parameters

Following van der Vaart & Wellner (1996, section 3.3), define random and fixed maps $\Psi_n(\theta,\alpha)$: $\mathcal{H}\to\mathbb{R}$, $\Psi(\theta,\alpha):\mathcal{H}\to\mathbb{R}$ for some index set $\mathcal{H}$ with $\Psi_n(\theta,\alpha),\Psi(\theta,\alpha)\in\ell^\infty(\mathcal{H})$. In most applications, including that in section 3, $\Psi_n(\theta,\alpha)h=\mathbb{P}_n\psi_{\theta,\alpha,h}$ and $\Psi(\theta,\alpha)h=P\psi_{\theta,\alpha,h}$ for given measurable functions $\psi_{\theta,\alpha,h}$ indexed by $\theta\in\Theta$, $\alpha\in\mathcal{A}$ and $h\in\mathcal{H}$. We do not insist on this in the general theorem, however.

Suppose that $\Psi(\theta_0,\alpha_0)=0$. Here α is to be regarded as a nuisance parameter; in the application, α is a finite-dimensional parameter while $\theta=(\nu,\eta)$ where ν is finite-dimensional and η is infinite-dimensional. Suppose we have available estimators $\hat\alpha=\hat\alpha_n$ of α, and then consider estimators $\hat\theta_n$ of θ satisfying

$$\sup_{h\in\mathcal{H}} |\Psi_n(\hat{\theta}_n, \hat{\alpha}_n)h| = o_p^*(n^{-1/2}). \tag{1}$$

We would like to establish limit theorems for $\sqrt{n}(\hat{\theta}_n - \theta_0)$ which are similar to those in the standard Z-theorem of van der Vaart (1995); see also theorem 3.3.1, p. 310, of van der Vaart & Wellner (1996). As argued in van der Vaart & Wellner (2007, p. 235), we can derive limit distributions of $\hat{\theta}_n$ based on $\{\Psi_n(\theta, \hat{\alpha}) : \theta \in \Theta\}$ from the corresponding theory for $\{\Psi_n(\theta, \alpha_0) : \theta \in \Theta\}$, if we know that $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$ and if we show that

$$\sup_{\theta\in\Theta} ||\sqrt{n}(\Psi_n - \Psi)(\theta, \hat{\alpha}_n) - \sqrt{n}(\Psi_n - \Psi)(\theta, \alpha_0)||_{\mathcal{H}} = o_p^*(1). \tag{2}$$

An alternative goal would be to relate the estimators $\hat{\theta}_n$ to estimators $\hat{\theta}_n^0$ that satisfy $\sup_{h\in\mathcal{H}} |\Psi_n(\hat{\theta}_n^0, \alpha_0)h| = o_p^*(n^{-1/2})$. This is accomplished in the third part of the following theorem which generalizes theorem 5.31, p. 60, of van der Vaart (1998); see also theorem 6.18, p. 407, of van der Vaart (2002).

Theorem 1

Suppose that $\hat{\theta}_n$ satisfies (1), that $\theta \mapsto \{\Psi(\theta, \alpha)h : h \in \mathcal{H}\}$ is uniformly Fréchet-differentiable in a neighbourhood of α_0 with derivative maps $\dot{\Psi}(\theta_0, \alpha)$ satisfying $\dot{\Psi}(\theta_0, \alpha) \to \dot{\Psi}(\theta_0, \alpha_0) \equiv \dot{\Psi}_0$ as $\alpha \to \alpha_0$ with $\dot{\Psi}_0$ continuously invertible. Suppose, moreover, that $\mathbb{Z}_n \equiv \sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0) \rightsquigarrow \mathbb{Z}_0$ in $\ell^\infty(\mathcal{H})$, that (2) holds and that

$$\left|\left|\sqrt{n}(\Psi_n - \Psi)(\hat{\theta}_n, \alpha_0) - \sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0)\right|\right|_{\mathcal{H}} = o_p^*(1 + \sqrt{n}||\hat{\theta}_n - \theta_0||). \tag{3}$$

(*i*) *If $||\sqrt{n}(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0))||_{\mathcal{H}} = O_p^*(1)$, then $\sqrt{n}||\hat{\theta}_n - \theta_0|| = O_p^*(1)$ and*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\dot{\Psi}_0^{-1}\sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0) - \dot{\Psi}_0^{-1}\left[\sqrt{n}(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0))\right] + o_p^*(1). \tag{4}$$

(*ii*) *If the map $\alpha \mapsto \{\Psi(\theta_0, \alpha)h : h \in \mathcal{H}\}$ is Fréchet-differentiable at α_0 with derivative map $\dot{\Psi}_\alpha$ and $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = \mathbb{G}_n\varphi + o_p^*(1)$ satisfies $(\mathbb{Z}_n, \sqrt{n}(\hat{\alpha}_n - \alpha_0)) \rightsquigarrow (\mathbb{Z}_0, \mathbb{G}\varphi)$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow -\dot{\Psi}_0^{-1}\left(\mathbb{Z}_0 + \dot{\Psi}_\alpha\mathbb{G}\varphi\right). \tag{5}$$

(*iii*) *Under the same hypotheses as in (i)*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\hat{\theta}_n^0 - \theta_0) - \dot{\Psi}_0^{-1}\sqrt{n}\left(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0)\right) + o_p^*(1). \tag{6}$$

(*iv*) *Under the same hypotheses as in (ii)*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\hat{\theta}_n^0 - \theta_0) - \dot{\Psi}_0^{-1}\dot{\Psi}_\alpha\sqrt{n}(\hat{\alpha}_n - \alpha_0) + o_p^*(1). \tag{7}$$

Proof. By the definition of $\hat{\theta}_n$,

$$\begin{aligned}
&\sqrt{n}\Big(\Psi(\hat{\theta}_n, \hat{\alpha}_n) - \Psi(\theta_0, \hat{\alpha}_n)\Big)\\
&= \sqrt{n}\left(\Psi(\hat{\theta}_n, \hat{\alpha}_n) - \Psi_n(\hat{\theta}_n, \hat{\alpha}_n)\right) - \sqrt{n}\left(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0)\right) + o_p^*(1)\\
&= -\sqrt{n}\Big(\Psi_n(\hat{\theta}_n, \hat{\alpha}_n) - \Psi(\hat{\theta}_n, \hat{\alpha}_n)\Big) + \sqrt{n}\big(\Psi_n(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \hat{\alpha}_n)\big)\\
&\quad - \sqrt{n}\big(\Psi_n(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \hat{\alpha}_n)\big) - \sqrt{n}\big(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0)\big) + o_p^*(1)\\
&= -\sqrt{n}(\Psi_n - \Psi)(\hat{\theta}_n, \alpha_0) + \sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0)\\
&\quad - \sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0) - \sqrt{n}(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0)) + o_p^*(1) \quad \text{by (2)}\\
&= o_p^*(1 + \sqrt{n}||\hat{\theta}_n - \theta_0||) - \sqrt{n}(\Psi_n - \Psi)(\theta_0, \alpha_0) - \sqrt{n}(\Psi(\theta_0, \hat{\alpha}_n) - \Psi(\theta_0, \alpha_0)) + o_p^*(1)
\end{aligned} \tag{8}$$

by using (3) in the last line. By uniform differentiability of $\theta \mapsto \{\Psi(\theta,\alpha)h : h \in \mathcal{H}\}$ and uniform non-singularity of $\dot\Psi(\theta_0,\alpha)$, it follows that there is a constant $c>0$ such that, for all (θ,α) in a sufficiently small neighbourhood of (θ_0,α_0), $||\Psi(\theta,\alpha)-\Psi(\theta_0,\alpha)|| \ge c||\theta-\theta_0||$. Combining this with (8) yields

$$c\sqrt{n}||\hat\theta_n-\theta_0|| \le o_p^*(1+\sqrt{n}||\hat\theta_n-\theta_0||)+||\mathbb{Z}_n||+||\sqrt{n}(\Psi(\theta_0,\hat\alpha_n)-\Psi(\theta_0,\alpha_0))||$$

and hence that $\sqrt{n}||\hat\theta_n-\theta_0||=O_p^*(1)$ if the last term in the preceding display is $O_p(1)$. Now

$$\begin{aligned}\sqrt{n}(\hat\theta_n-\theta_0) &= -\dot\Psi_0^{-1}\left[\mathbb{Z}_n+\sqrt{n}\left(\Psi(\theta_0,\hat\alpha_n)-\Psi(\theta_0,\alpha_0)\right)\right]+o_p^*(1)\\ &= -\dot\Psi_0^{-1}\left(\mathbb{Z}_n+\dot\Psi_\alpha\mathbb{G}_n\varphi\right)+o_p^*(1)\end{aligned}$$

where in the first equation we have again used the uniform differentiability hypothesis and in the second the hypotheses of part (ii) of the theorem. As this converges to the claimed limit, this proves (i) and (ii). To prove (iii) and (iv), note that the standard Z-theorem yields $\sqrt{n}(\hat\theta_n^0-\theta_0)=-\dot\Psi_0^{-1}\mathbb{Z}_n+o_p^*(1)$. The claimed results follow by combining each line in the last display with (4).

Remark. Under the hypotheses of theorem 1, theorem 2.21 of Kato (1976, p. 205) implies that the derivative maps $\dot\Psi(\theta_0,\alpha)$ are continuously invertible for α in a neighbourhood of α_0.

3. Completion of the proof of Breslow & Wellner (2007)

In Breslow & Wellner (2007), $\mathcal{P}=\{P_{\theta,\eta}:\theta\in\Theta,\eta\in\Xi\}$ is a semiparametric model that satisfies five assumptions A1–A5. Here we slightly strengthen A1, which had already strengthened the hypotheses of van der Vaart (1998, theorem 25.90), to

A1* for (θ,η) in a δ-neighbourhood of (θ_0,η_0) the functions $\dot\ell_{\theta,\eta}$ and $\{B_{\theta,\eta}h-P_{\theta,\eta}B_{\theta,\eta}h, h\in\mathcal{H}\}$ are contained in a P_0-Donsker class $\mathcal{F}$ and have square-integrable envelope functions F_1 and F_2 respectively.

We also strengthen A3 to

A3* A3 holds and moreover the derivative maps $\dot\Psi_0=(\dot\Psi_{11},\dot\Psi_{12},\dot\Psi_{21},\dot\Psi_{22})$ have representations

$$\dot\Psi_{ij}(\theta_0,\eta_0)h=P_0(\dot\psi_{ij,\theta_0,\eta_0,h}),\quad i,j\in\{1,2\}$$

in terms of $L_2(P_0)$ derivatives of $\psi_{1,\theta,\eta,h}=\dot\ell_{\theta,\eta}$ and $\psi_{2,\theta,\eta,h}=B_{\theta,\eta}h-P_{\theta,\eta}B_{\theta,\eta}h$, i.e.

$$\sup_{h\in\mathcal{H}}\left\{P_0\left(\psi_{i,\theta,\eta_0,h}-\psi_{i,\theta_0,\eta_0,h}-\dot\psi_{i1,\theta_0,\eta_0,h}(\theta-\theta_0)\right)^2\right\}^{1/2}=o(||\theta-\theta_0||),$$

$$\sup_{h\in\mathcal{H}}\left\{P_0\left(\psi_{i,\theta_0,\eta,h}-\psi_{i,\theta_0,\eta_0,h}-\dot\psi_{i2,\theta_0,\eta_0,h}(\eta-\eta_0)\right)^2\right\}^{1/2}=o(||\eta-\eta_0||),$$

for $i=1,2$.

Breslow & Wellner (2007) showed in (25) on p. 92 that for $\pi=\pi_0$

$$\sqrt{N}\begin{pmatrix}\hat\theta_N(\alpha_0)-\theta_0\\ \hat\alpha_N-\alpha_0\end{pmatrix}=\sqrt{N}\begin{pmatrix}\mathbb{P}_N^\pi\tilde\ell_0\\ \mathbb{Q}_N\tilde\ell_0^\alpha\end{pmatrix}+o_p(1) \tag{9}$$

and thus that the quantity on the LHS has an asymptotic $N(0,V)$ distribution where

$$V=\begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix},\qquad V_{11}=\tilde I_0^{-1}+\tilde P_0\left(\tfrac{1-\pi_0}{\pi_0}\tilde\ell_0^{\otimes 2}\right),$$

$$V_{22}=\left(\tilde P_0 1_{\mathcal{V}_0^c}\frac{\dot\pi_0^{\otimes 2}}{\pi_0(1-\pi_0)}\right)^{-1},\qquad V_{12}=V_{21}^T=\tilde P_0\left(\tfrac{\xi}{\pi_0}\tilde\ell_0\tilde\ell_0^\alpha\right)=\tilde P_0\left(\tilde\ell_0\tilde\ell_0^\alpha\right).$$

Thus Pierce's (1982) first hypothesis (1.1) is satisfied with $\hat{\theta}(\alpha_0) - \theta_0$ the statistic of interest, T_n in his notation, and α the estimated nuisance parameter, Pierce's (1982) λ.

Pierce's (1982) second hypothesis (1.2) is that with $\hat{T}_n = T_n(\hat{\lambda}_n)$

$$\sqrt{n}\hat{T}_n = \sqrt{n}T_n + B\sqrt{n}(\hat{\lambda}_n - \lambda) + o_p(1) \tag{10}$$

for some matrix B. A further hypothesis is that $\hat{\lambda}_n$ is efficient; i.e. $V_{22} = I_{\lambda}^{-1}$. Then Pierce shows that $\sqrt{n}\hat{T}_n \rightsquigarrow N(0, V_{11} - BV_{22}B^T)$. Breslow & Wellner (2007) also showed in (26) on p. 92 that

$$\sqrt{N}(\mathbb{P}_N^{\hat{\pi}} - \mathbb{P}_N^{\pi_0})\tilde{\ell}_0 = -\tilde{P}_0\left(\mathbf{1}_{\mathcal{V}_0^c}\frac{\tilde{\ell}_0\dot{\pi}_0^T}{\pi_0}\right)\sqrt{N}(\hat{\alpha}_N - \alpha_0) + o_p(1). \tag{11}$$

However, as pointed out by Li, this does not yet prove that

$$\sqrt{N}(\hat{\theta}_N(\hat{\alpha}) - \theta_0) = \sqrt{N}(\hat{\theta}_N(\alpha_0) - \theta_0) + B\sqrt{N}(\hat{\alpha}_N - \alpha_0) + o_p(1) \tag{12}$$

for some matrix B as is needed to verify (10). This result does, however, follow from part (iv) of theorem 1.

As in section 6 of Breslow & Wellner (2007), suppose that $\hat{\alpha} = \hat{\alpha}_N$ denotes the maximum likelihood estimator of parameters in the model $\pi_\alpha(v)$ for the sampling probabilities and that $\hat{\theta}_N \equiv \hat{\theta}_N(\hat{\alpha})$, $\hat{\eta}_N \equiv \hat{\eta}_N(\hat{\alpha})$ solve $\mathbb{P}_N^{\hat{\pi}}\dot{\ell}_{\theta,\eta} = 0$ and $\mathbb{P}_N^{\hat{\pi}}B_{\theta,\eta}h - P_{\theta,\eta}B_{\theta,\eta}h = 0$ for all $h \in \mathcal{H}$ where

$$\mathbb{P}_N^{\hat{\pi}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\hat{\pi}_i}\delta_{X_i},$$

and where $\hat{\pi}_i \equiv \pi_{\hat{\alpha}}(V_i)$, $i = 1, \ldots, N$.

Theorem 2

Suppose the semiparametric model $\mathcal{P}$ satisfies A1 and A3* above and A2, A4 and A5 of Breslow & Wellner (2007) and that the model $\pi_\alpha(V)$ for the conditional distribution of ξ given X, V satisfies the hypotheses of theorem 5.39 of van der Vaart (1998). Suppose moreover that π_α satisfies (42) of Breslow & Wellner (2007):*

$$\left|\frac{1}{\pi_\alpha(v)} - \frac{1}{\pi_{\alpha_0}(v)} - \frac{-\dot{\pi}_0^T(v)}{\pi_0^2(v)}(\alpha - \alpha_0)\right| \le \psi(v)|\alpha - \alpha_0|^{1+\zeta} \tag{13}$$

for α in a neighbourhood of α_0 where $\zeta > 0$ and ψ satisfies $E\psi^2(V) < \infty$. Then

$$\sqrt{N}(\hat{\theta}_N(\hat{\alpha}) - \theta_0) \rightsquigarrow Z \sim N_p(0, \Sigma)$$

where, as in (27) of Breslow & Wellner (2007),

$$\Sigma = \mathrm{Var}\left(\frac{\xi}{\pi_0}\tilde{\ell}_0\right) - \tilde{P}_0 1_{\mathcal{V}_0^c}\frac{\tilde{\ell}_0\dot{\pi}_0^T}{\pi_0}\left(\tilde{P}_0 1_{\mathcal{V}_0^c}\frac{\dot{\pi}_0^{\otimes 2}}{\pi_0(1-\pi_0)}\right)^{-1}\tilde{P}_0 1_{\mathcal{V}_0^c}\frac{\dot{\pi}_0\tilde{\ell}_0^T}{\pi_0}.$$

Proof. We use theorem 1 with θ there replaced by (θ, η). We proceed by verifying the conditions of the theorem, beginning with (2). Recall that $W = (X, U)$ and $V = (\tilde{X}, U)$ where $\tilde{X} = \tilde{X}(X)$. Consider the classes of functions

$$\psi_{1;\theta,\eta,\alpha}(w, \xi) = \frac{\xi}{\pi_\alpha(v)}\dot{\ell}_{\theta,\eta}(x), \tag{14}$$

$$\psi_{2;\theta,\eta,\alpha,h}(w, \xi) = \frac{\xi}{\pi_\alpha(v)}B_{\theta,\eta}h(x), \tag{15}$$

for $\theta \in \Theta$, $\eta \in \Xi$, and $h \in \mathcal{H}$. Then showing (2) is equivalent to showing that

$$\sup_{\theta\in\Theta,\ \eta\in\Xi} |\mathbb{G}_N(\psi_{1;\theta,\eta,\hat\alpha} - \psi_{1;\theta,\eta,\alpha_0})| \to_{p^*} 0 \quad \text{and} \tag{16}$$

$$\sup_{\theta\in\Theta,\ \eta\in\Xi,\ h\in\mathcal{H}} |\mathbb{G}_N(\psi_{2;\theta,\eta,\hat\alpha,h} - \psi_{2;\theta,\eta,\alpha_0,h})| \to_{p^*} 0. \tag{17}$$

Under the condition (13) imposed by Breslow & Wellner (2007), (16) holds by virtue of

$$\begin{aligned}
&\psi_{1;\theta,\eta,\hat\alpha}(w,\xi) - \psi_{1;\theta,\eta,\alpha_0}(w,\xi)\\
&\quad = \left(\frac{\xi}{\pi_{\hat\alpha}(v)} - \frac{\xi}{\pi_{\alpha_0}(v)}\right)\dot\ell_{\theta,\eta}(x)\\
&\quad = \xi\left(\frac{1}{\pi_{\hat\alpha}(v)} - \frac{1}{\pi_{\alpha_0}(v)} - \frac{-\dot\pi_0(v)^T}{\pi_0^2(v)}(\hat\alpha-\alpha_0)\right)\dot\ell_{\theta,\eta} - \ \xi\frac{\dot\pi_0(v)^T}{\pi_0^2(v)}(\hat\alpha-\alpha_0)\dot\ell_{\theta,\eta}.
\end{aligned}$$

Then

$$\begin{aligned}
&\sqrt{N}(\mathbb{P}_N - P)(\psi_{1;\theta,\eta,\hat\alpha} - \psi_{1;\theta,\eta,\alpha_0})\\
&\quad = \sqrt{N}(\mathbb{P}_N - P)\left(\xi\left(\frac{1}{\pi_{\hat\alpha}(v)} - \frac{1}{\pi_{\alpha_0}(v)} - \frac{-\dot\pi_0(v)^T}{\pi_0^2(v)}(\hat\alpha-\alpha_0)\right)\dot\ell_{\theta,\eta}\right)\\
&\qquad + \sqrt{N}(\mathbb{P}_N - P)\left(-\xi\frac{\dot\pi_0(v)^T}{\pi_0^2(v)}(\hat\alpha-\alpha_0)\dot\ell_{\theta,\eta}\right)\\
&\quad \equiv R_N + S_N
\end{aligned}$$

where, using (13),

$$\begin{aligned}
|R_N| &\le \sqrt{N}(\mathbb{P}_N + P)\left\{\left|\left(\frac{1}{\pi_{\widehat\alpha}(v)} - \frac{1}{\pi_{\alpha_0}(v)} - \frac{-\dot\pi_0(v)^T}{\pi_0^2(v)}(\widehat\alpha-\alpha_0)\right)\dot\ell_{\theta,\eta}\right|\right\}\\
&\le \mathbb{P}_N\left(\psi(V)|\dot\ell_{\theta,\eta}|\right)\sqrt{N}|\hat\alpha-\alpha_0|^{1+\zeta}\\
&\quad + P\left(\psi(V)|\dot\ell_{\theta,\eta}|\right)\sqrt{N}|\widehat\alpha-\alpha_0|^{1+\zeta}\\
&\le \left\{\sqrt{\mathbb{P}_N\psi^2(V)\cdot\mathbb{P}_N F_1^2} + \sqrt{P\psi^2(V)\cdot PF_1^2}\right\}\sqrt{N}|\widehat\alpha-\alpha_0|^{1+\zeta}\\
&= O_p(1)o_p(1)
\end{aligned}$$

uniformly in $\theta \in \Theta$, $\eta \in \Xi$. Here F_1 is a square integrable envelope function for the class of functions $\{\dot\ell_{\theta,\eta} : \theta \in \Theta, \eta \in \Xi\}$, which exists by A1*.

To handle S_N, note that

$$\begin{aligned}
|S_N| &= \left|\sqrt{N}(\mathbb{P}_N - P)\left(-\xi\frac{\dot\pi_0^T}{\pi_0^2}(\hat\alpha-\alpha_0)\dot\ell_{\theta,\eta}\right)\right|\\
&= \left|(\mathbb{P}_N - P)\left(-\xi\dot\ell_{\theta,\eta}\frac{\dot\pi_0^T}{\pi_0^2}\right)\sqrt{N}(\hat\alpha-\alpha_0)\right|\\
&\le \sup_{\theta\in\Theta,\eta\in\Xi}\left|(\mathbb{P}_N - P)\left(-\xi\dot\ell_{\theta,\eta}\frac{\dot\pi_0^T}{\pi_0^2}\right)\right|\sqrt{N}|\hat\alpha-\alpha_0|\\
&= o_p^*(1)O_p(1) = o_p^*(1)
\end{aligned}$$

uniformly in $\theta \in \Theta$, $\eta \in \Xi$ as the class of functions $\{\xi\dot\ell_{\theta,\eta}\dot\pi_0^T/\pi_0^2 : \theta \in \Theta,\ \eta \in \Xi\}$ is a Glivenko–Cantelli class of functions. Here is the argument: as the class $\{\dot\ell_{\theta,\eta} : \theta \in \Theta, \eta \in \Xi\}$ is P-Donsker, it is P-Glivenko–Cantelli. Furthermore the (vector of) function(s) $\{-\xi\dot\pi_0^T/\pi_0^2\}$ is square-integrable: for $1 \le j \le q = \dim(\alpha)$

$$E\left(\xi\frac{\dot\pi_{0j}}{\pi_0^2}\right)^2 = E\frac{\dot\pi_{0j}^2}{\pi_0^3} \le \frac{1}{\sigma^2}E\frac{\dot\pi_{0j}^2}{\pi_0(1-\pi_0)} = \frac{1}{\sigma^2}E\dot\ell_{\alpha,j}^2 < \infty$$

by our assumptions on the model $\pi_\alpha(v)$. Thus the Glivenko–Cantelli preservation theorem of van der Vaart & Wellner (2000) applies by taking $\varphi(u,v)=uv$, $\mathcal{F}_1=\{\xi\dot\pi_{0j}/\pi_0^2\}$, $\mathcal{F}_2=\dot\ell_{\theta,\eta}$, and noting that $\mathcal{F}_1\cdot\mathcal{F}_2$ has integrable envelope function $\left(\dot\pi_{0j}/\pi_0^2\right)F_1$. A similar argument works for (17) using the square integrable envelope F_2 for $\{B_{\theta,\eta}h:\theta\in\Theta, h\in\mathcal{H}, \eta\in\Xi\}$.

Now note that A1*, A2, A3* and A4 imply that $\mathbb{G}_n\psi_{\theta_0,\eta_0,\alpha_0}\rightsquigarrow\mathbb{G}\psi_{\theta_0,\eta_0,\alpha_0}$ in $\ell^\infty(\mathcal{H})$ and that (3) holds as α_0 is fixed in both cases. The hypothesized uniform Fréchet differentiability holds under (13) and A3*: writing θ for (θ,η) in the spirit of theorem 1,

$$
\begin{aligned}
&||\Psi(\theta,\alpha)h-\Psi(\theta_0,\alpha)h-\dot\Psi(\theta_0,\alpha)(\theta-\theta_0)h||_{\mathcal{H}}\\
&=\sup_{h\in\mathcal{H}}\left|P_0\left\{\frac{\pi_{\alpha_0}}{\pi_\alpha}\left(\psi_{\theta,h}-\psi_{\theta_0,h}-\dot\psi_{\theta_0,h}(\theta-\theta_0)\right)\right\}\right|\\
&\le\left\{P_0\left(\frac{\pi_{\alpha_0}}{\pi_\alpha}\right)^2\right\}^{1/2}\sup_{h\in\mathcal{H}}\left\{P_0\left(\psi_{\theta,h}-\psi_{\theta_0,h}-\dot\psi_{\theta_0,h}(\theta-\theta_0)\right)^2\right\}^{1/2}\\
&\le Ko(||\theta-\theta_0||)
\end{aligned}
$$

by using the assumed regularity of π_α, (13) and (3) of Breslow & Wellner (2007) to bound $P_0(\pi_{\alpha_0}/\pi_\alpha)^2$ uniformly in a neighbourhood of α_0 and using A3* to bound the second term. The additional hypotheses in (ii) and (iv) also follow from the regularity of π_α and (13).

To complete the proof, write $\psi_{\theta,\eta,\alpha,h}=(\psi_{1;\theta,\eta,\alpha},\psi_{2;\theta,\eta,\alpha,h})$ as defined in (14) and (15). Then the corresponding components of

$$
\dot\Psi_\alpha h=\frac{\partial}{\partial\alpha^T}P_0\psi_{\theta,\eta,\alpha,h}\Big|_{\alpha=\alpha_0}
$$

are

$$
\begin{aligned}
&\dot\Psi_{1,\alpha}=-P_0\left(\dot\ell_0\mathbf{1}_{\mathcal{V}_0^c}\frac{\dot\pi_0^T}{\pi_0}\right)\quad\text{and}\\
&\dot\Psi_{2;\alpha}h=-P_0\left(B_0h\mathbf{1}_{\mathcal{V}_0^c}\frac{\dot\pi_0^T}{\pi_0}\right),\quad h\in\mathcal{H}.
\end{aligned}
$$

Consequently, operating with the partitioned (assumption A5, η a measure) version of $\dot\Psi_0$ on both left- and right-hand sides of (7) we find

$$
\begin{aligned}
&-I_0\sqrt{N}\left(\hat\theta_N-\theta_0\right)-\sqrt{N}\left(\hat\eta_N-\eta_0\right)B_0^*\dot\ell_0\\
&\quad=-I_0\sqrt{N}\left(\hat\theta_N^0-\theta_0\right)-\sqrt{N}\left(\hat\eta_N^0-\eta_0\right)B_0^*\dot\ell_0+P_0\left(\dot\ell_0\mathbf{1}_{\mathcal{V}_0^c}\frac{\dot\pi_0^T}{\pi_0}\right)\sqrt{N}\left(\hat\alpha_N-\alpha_0\right)+o_p^*(1)
\end{aligned}
$$

and

$$
\begin{aligned}
&-P_0(B_0h)\dot\ell_0^T\sqrt{N}\left(\hat\theta_N-\theta_0\right)-\sqrt{N}\left(\hat\eta_N-\eta_0\right)B_0^*B_0h\\
&\quad=-P_0(B_0h)\dot\ell_0^T\sqrt{N}\left(\hat\theta_N^0-\theta_0\right)-\sqrt{N}\left(\hat\eta_N^0-\eta_0\right)B_0^*B_0h+P_0\left(B_0h\mathbf{1}_{\mathcal{V}_0^c}\frac{\dot\pi_0^T}{\pi_0}\right)\sqrt{N}\left(\hat\alpha_N-\alpha_0\right)\\
&\qquad+o_p^*(1).
\end{aligned}
$$

Following closely again section 25.12 of van der Vaart (1998) we choose $h=\left(B_0^*B_0\right)^{-1}B_0^*\dot\ell_0$ and subtract the first equation from the second to find

$$
\begin{aligned}
&P_0\left[\left(I-B_0(B_0^*B_0)^{-1}B_0^*\right)\dot\ell_0\dot\ell_0^T\right]\sqrt{N}\left(\hat\theta_N-\theta_0\right)\\
&=P_0\left[\left(I-B_0(B_0^*B_0)^{-1}B_0^*\right)\dot\ell_0\dot\ell_0^T\right]\sqrt{N}\left(\hat\theta_N^0-\theta_0\right)\\
&\quad-P_0\left[\left(I-B_0(B_0^*B_0)^{-1}B_0^*\right)\dot\ell_0\mathbf{1}_{\mathcal{V}_0^c}\frac{\dot\pi_0^T}{\pi_0}\right]\sqrt{N}\left(\hat\alpha_N-\alpha_0\right)+o_p^*(1).
\end{aligned}
$$

Recognizing that $I_0 = P_0 \dot{\ell}_0 \dot{\ell}_0^T$ is the ordinary information for θ, $(I - B_0(B_0^* B_0)^{-1} B_0^*)\dot{\ell}_0$ is the efficient score and $\tilde{I}_0 = P_0[(I - B_0(B_0^* B_0)^{-1} B_0^*)\dot{\ell}_0 \dot{\ell}_0^T]$ is the efficient information, we multiply both sides of the preceding equation by $\tilde{I}_0^{-1}$ to find

$$\sqrt{N}\left(\hat{\theta}_N(\hat{\alpha}) - \theta_0\right) = \sqrt{N}\left(\hat{\theta}_N(\alpha_0) - \theta_0\right) - P_0\left(\tilde{\ell}_0 \mathbf{1}_{\mathcal{V}_0^c} \frac{\dot{\pi}_0^T}{\pi_0}\right)\sqrt{N}\left(\hat{\alpha}_N - \alpha_0\right) + o_p(1) \tag{18}$$

which is the second hypothesis (1.2) of Pierce (1982), equivalent to our (12) above, with $B = -P_0\left(\tilde{\ell}_0 \mathbf{1}_{\mathcal{V}_0^c} \dot{\pi}_0^T / \pi_0\right)$. Theorem 2 now follows from theorem 1 via (9) and (18). The resolution of the gap in Breslow & Wellner's (2007) argument, namely the demonstration that

$$\sqrt{N}(\hat{\theta}_N(\hat{\alpha}) - \theta_0) = \sqrt{N}\mathbb{P}_N^{\hat{\pi}} \tilde{\ell}_0 + o_p(1),$$

is obtained from (9), (11) and (18) as a corollary to theorem 2.

Acknowledgement

We thank Zhiguo Li for pointing out the gap in the arguments in Breslow & Wellner (2007).

References

Breslow, N. E. & Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Statist.* **34**, 86–102.
Kato, T. (1976). *Perturbation theory for linear operators*, 2nd edn. Springer-Verlag, Berlin.
Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**, 475–478.
van der Vaart, A. W. (1995). Efficiency of infinite-dimensional M-estimators. *Statist. Neerlandica* **49**, 9–30.
van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.
van der Vaart, A. W. (2002). Semiparametric statistics. In *Lectures on probability theory and statistics. Ecole d'Eté de Probabilités de Saint-Flour XXIX – 1999* (eds E. Bolthausen, E. Perkins & A. van der Vaart), 331–457. Springer-Verlag, Berlin.
van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.
van der Vaart, A. W. & Wellner, J. A. (2000). Preservation theorems for Glivenko–Cantelli and uniform Glivenko–Cantelli classes. In *High dimensional probability II* (eds E. Giné, D. M. Mason & J. A. Wellner), 115–133. Birkhäuser, Boston.
van der Vaart, A. W. & Wellner, J. A. (2007). Empirical processes indexed by estimated functions. *IMS Lecture Notes Monogr. Ser.* **55**, 234–252.

Received July 2007, in final form August 2007

Norman E. Breslow, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA.
E-mail: norm@u.washington.edu