Statistica Sinica **12**(2002), 609-634

CENTRAL LIMIT THEOREMS FOR FUNCTIONAL Z-ESTIMATORS

Yihui Zhan

Insightful Corp.

Abstract: We establish central limit theorems for function-valued estimators defined as a zero point of a function-valued random criterion function. Our approach is based on a differential identity that applies when the random criterion function is linear in terms of the empirical measure. We do not require linearity of the statistical model in the unknown parameter, even though the result is most applicable for models with convex linearity that can be boundedly extended to the linear span of the parameter space. Three examples are given to illustrate the application of these theorems: a simplified frailty model which is nonlinear in the unknown parameter; the multiplicative censoring and double censoring models which are bounded convex linear in the unknown parameter.

Key words and phrases: Estimating equations, M-estimators, nonparametric maximum likelihood, stochastic equicontinuity, self-consistency, weak convergence.

1. Introduction

The methods for proving central limit theorems for maximum likelihood estimators defined as zero points of the likelihood equations (and thus Z-estimators) can be traced back to the classic theory due to Fisher (1922) and Cramér (1946). A recent development has introduced empirical process methodology and extended the classic results to general estimating equations that may not be likelihood equations, see Daniels (1961), Huber (1964, 1967), Pakes and Pollard (1989) and Pollard (1985, 1989). Another recent advance is the extension of the theory to models involving functional parameters, see Gill (1989) and Van der Vaart (1994, 1995). A delta method using compact differentiability is also given in Heesterman and Gill (1992). See also Appendix A.10 in Bickel, Klaassen, Ritov and Wellner (1993) and Section 3.3 in Van der Vaart and Wellner (1996).

The context for a central limit theorem for Z-estimators includes an empirical measure $\mathbb{P}_n$ for n i.i.d. observations and a score operator $B(\theta)$ depending on a parameter θ of interest. We are interested in proving a central limit theorem for Z-estimators $\{\hat{\theta}_n\}$ defined as the zero points of the estimating equations

$$\psi(\hat{\theta}_n, \mathbb{P}_n) \equiv \mathbb{P}_n B(\hat{\theta}_n) = o_{P^*}(n^{-1/2}). \tag{1.1}$$

Traditional Argument. Let P be the true probability. To prove a central limit theorem, the traditional argument assumes that the operator $\psi(\theta, P)$ is Fréchet differentiable in θ with respect to a norm $\|\cdot\|$. One expands $\psi(\theta, P)$ at the true θ_0 and evaluates the linear approximation at $\hat{\theta}_n$:

$$\psi(\hat{\theta}_n, P) - \psi(\theta_0, P) = \dot{\psi}(\theta_0)(\hat{\theta}_n - \theta_0) + o_{P^*}(\|\hat{\theta}_n - \theta_0\|). \tag{1.2}$$

Suppose $\psi(\theta_0, P) = PB(\theta_0) = 0$ and that the theory of empirical process can be used to show $\mathbb{G}_n B(\hat{\theta}_n) = \mathbb{G}_n B(\theta_0) + o_{P^*}(1)$, where $\mathbb{G}_n$ is the empirical process. By (1.1) and algebra, the difference $\sqrt{n}(\psi(\hat{\theta}_n, P) - \psi(\theta_0, P)) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(1)$ (see Lemma 2.3 for details), so that the linearization in (1.2) implies

$$\dot{\psi}(\theta_0)\Big(\sqrt{n}(\hat{\theta}_n - \theta_0)\Big) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(\sqrt{n}\|\hat{\theta}_n - \theta_0\|) + o_{P^*}(1). \tag{1.3}$$

This relates $\sqrt{n}(\hat{\theta}_n - \theta_0)$ directly to a weakly convergent process $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ with $\mathbb{G}_n B(\theta_0) \Rightarrow \mathbb{Z}_0$, where $\mathbb{Z}_0$ is a Brownian bridge process defined on an appropriate space.

By assuming the bounded invertibility of the operator $\dot{\psi}(\theta_0)$ with respect to the *same* norm $\|\cdot\|$, one can improve on the consistency of $\hat{\theta}_n$ and prove that $\hat{\theta}_n$ actually converges with a rate of $n^{-1/2}$, i.e., $\sqrt{n}\|\hat{\theta}_n - \theta_0\| = O_{P^*}(1)$ (Lemma 2.4). With this boundedness, the dominant error term $o_{P^*}(\sqrt{n}\|\hat{\theta}_n - \theta_0\|)$ in (1.3) vanishes as n goes to infinity. Hence, by the continuous mapping theorem, $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow -\dot{\psi}^{-1}(\theta_0)(\mathbb{Z}_0)$ is asymptotically normal.

Difficulty with the Traditional Argument. There are non-trivial examples involving both Euclidean parameters and functional parameters for which central limit theorems have been established using the traditional argument. However, there are other interesting examples for which only somewhat restrictive results can be obtained by a similar argument. One example is the double censoring model (Chang and Yang (1987), Chang (1990), and Gu and Zhang (1993)). Another example is the multiplicative censoring model (Vardi (1989), Vardi and Zhang (1992), and Van der Vaart (1994)).

The main difficulty with the traditional argument is that the derivative operator $\dot{\psi}(\theta_0)$ may not be boundedly invertible with respect to the norm $\|\cdot\|$ used in linearization (1.2). For example, the operator $\dot{\psi}(\theta_0)$ in the double censoring model is only invertible with respect to a weaker norm $\|\cdot\|_K$ rather than the stronger uniform norm $\|\cdot\|$. To improve the convergence rate from consistency to $n^{-1/2}$ and thereby validate the linearization and prove a central limit theorem by this argument, however, both the invertibility of $\dot{\psi}(\theta_0)$ and the differentiability of $\psi(\theta, P)$ have to be established with respect to *the same norm* $\|\cdot\|$. At this point one may wonder if the weaker norm $\|\cdot\|_K$ should be used in linearization given that the derivative operator is invertible with respect to it. The answer is no; in the double censoring model $\psi(\theta, P)$ is not differentiable with respect to the $\|\cdot\|_K$ norm.

Argument Through a Differential Identity. In an interesting class of models, there is an identity that connects $\sqrt{n}(\hat{\theta}_n - \theta_0)$ to the weakly convergent quantity $\mathbb{G}_n B(\theta_0)$: $\dot{\psi}(\hat{\theta}_n)(\sqrt{n}(\hat{\theta}_n - \theta_0)) = -\dot{P}_{\hat{\theta}_n}(\sqrt{n}(\hat{\theta}_n - \theta_0))B(\hat{\theta}_n)$ (With $\vartheta = \hat{\theta}_n$ and $a = \sqrt{n}(\hat{\theta}_n - \theta_0)$ in Lemma 2.1). This identity was derived, in an ad hoc manner, in the multiplicative censoring model (equation (2.5) in Vardi and Zhang (1992)), in the double censoring model (equation (2.11) in Gu and Zhang (1993)), and in the interval truncation model (equation (21) in Tsai and Zhang (1995)). Van der Laan (1992) also derived an identity to investigate the efficiency of the MLE. However, previous work does not provide general model conditions for the validity of the identity.

A common feature in these problems is that the probability measures P_θ are convex linearly indexed by θ (as in the double censoring model and the multiplicative censoring model), or nearly so up to a normalizing constant (as in the interval truncation model). For a general class of models in which convex linearity can be boundedly extended to the linear span of the parameter space, this linearity identity can be established via Fréchet differentiability of the likelihood equations $\psi(\theta, P_\theta) = 0$. See Section 2.1 for more details, also see Zhan (1996).

This identity allows a linearization applied to P_θ instead of $\psi(\theta, P)$ through its derivative operator $\dot{P}_\vartheta(\cdot)$. For models P_θ with bounded convex linearity, the differential $\dot{P}_{\hat{\theta}_n}(\sqrt{n}(\hat{\theta}_n - \theta_0))B(\hat{\theta}_n)$ *exactly* equals the difference $\sqrt{n}(\psi(\hat{\theta}_n, P_{\hat{\theta}_n}) - \psi(\hat{\theta}_n, P_{\theta_0}))$ (Lemma 2.5). Consequently, we have

$$\dot{\psi}(\hat{\theta}_n)\Big(\sqrt{n}(\hat{\theta}_n - \theta_0)\Big) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(1). \tag{1.4}$$

Unlike (1.3), where $\hat{\theta}_n$ must converge with an $n^{-1/2}$ rate with respect to $\|\cdot\|$ to validate the linearization, there is no need to require this condition in (1.4) because the linearization is perfect. The Z-estimators $\{\hat{\theta}_n\}$ still have to converge at the $n^{-1/2}$ rate, but they may converge in any norm as long as the derivative operator is invertible with respect to it. This circumvents the problem with the traditional argument by obtaining a weak convergence as well as the rate control in one step. Theorem 2.2 is a rigorous statement of this argument.

For a model P_θ that is not linearly parameterized, the linearity identity leads to

$$\dot{\psi}(\hat{\theta}_n)\Big(\sqrt{n}(\hat{\theta}_n - \theta_0)\Big) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(\sqrt{n}\|\hat{\theta}_n - \theta_0\|) + o_{P^*}(1).$$

The term $o_{P^*}(\sqrt{n}\|\hat{\theta}_n - \theta_0\|)$ comes from approximating $\dot{P}_{\hat{\theta}_n}(\sqrt{n}(\hat{\theta}_n - \theta_0))B(\hat{\theta}_n)$ by the difference $\sqrt{n}(\psi(\hat{\theta}_n, P_{\hat{\theta}_n}) - \psi(\hat{\theta}_n, P_{\theta_0}))$. In this case, the uniform boundedness of $\dot{\psi}(\theta)$ with respect to the norm $\|\cdot\|$ is required to improve the rate of convergence for $\hat{\theta}_n$ to make the linearization valid, and a central limit theorem follows. Theorem 2.1 formulates this argument precisely.

Other Motivations. One of the motivations for a functional central limit theorem lies in the fact that the NPMLE (nonparametric maximum likelihood estimator) is a Z-estimator in most cases. The NPMLE is a popular estimator for a functional parameter such as a distribution function or a hazard function. It is the estimate that is actually computed in most applications because it is a discrete function and is the solution of a well-defined optimization problem. For some models, such as the double censoring model, there are effective algorithms for computing the NPMLE as an alternative to solving the likelihood equations (Wellner and Zhan (1997)). More importantly, efficient algorithms also make it possible to estimate the covariance structure of the limiting processes by bootstrap means (Wellner and Zhan (1996)). Because of the implicit and complicated expression for the limiting process, the bootstrap is an important method to construct a confidence set in a functional central limit theorem.

Our paper is organized as follows. Section 2.1 presents a differential identity based on which the linearity identity is derived. The linearity identity is obtained by combining Lemma 2.1 in Section 2.1 and Lemma 2.5 in Section 2.4. Section 2.2 presents the uniform boundedness condition needed in the proof of the central limit theorem. The actual proofs of the central limit theorems are given in Section 2.3 and Section 2.4. Three examples are given in Section 3 to illustrate the applications of the theorems.

2. Central Limit Theorems

Let X_i, $i = 1, 2, \ldots$ be a sequence of independent observations from a distribution $P \in \mathcal{P}$ on a probability space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{P}$ denotes the set of all probability measures on $(\mathcal{X}, \mathcal{A})$. Suppose that the collection $\mathcal{P}$ is parametrized by $\theta \in \Theta$, where Θ is assumed to be a smooth surface in a Banach space $(\mathbf{B}, \|\cdot\|)$ with a norm $\|\cdot\|$. We are interested in estimating a functional parameter $\theta_0 \in \Theta$, the true parameter.

Let $l^\infty(\mathcal{H})$ denote the set of bounded functions from $\mathcal{H}$ to the real line R, for some set $\mathcal{H}$, and let $\|\cdot\|_{\mathcal{H}}$ denote the uniform norm on $l^\infty(\mathcal{H})$. Let $B(\theta)$ be a θ-indexed operator (the score operator for θ) from $\mathcal{H}$ to some subset $\mathcal{F}(\theta)$ of $L_2(P_\theta)$ for each $\theta \in \Theta$. Define the set $\mathcal{F}(\Theta) = \bigcup_{\theta \in \Theta} \mathcal{F}(\theta)$. For simplicity of notation, we omit Θ in $\mathcal{F}(\Theta)$ and simply write $\mathcal{F}$. (See Van der Vaart (1995) for a similar formulation).

The empirical measure for the first n observations is denoted by $\mathbb{P}_n = (1/n) \sum_{i=1}^n \delta_{X_i}$ and the empirical process by $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$. As usual we use linear functional notation, and write $Pf = \int f dP$ for $f \in \mathcal{F}$, and we consider $\mathbb{G}_n$ as indexed by the collection of functions $\mathcal{F} \subset L_2(P)$, $P \in \mathcal{P}$.

For each fixed $\theta \in \Theta$ and $P \in \mathcal{P}$, define the operator $\psi(\theta, P) = PB(\theta)$ from $\mathcal{H}$ to the real line R. Suppose that $B(\theta)$ is bounded in the sense that $\|\psi(\theta, P)\|_{\mathcal{H}} = \|PB(\theta)\|_{\mathcal{H}} < \infty$ for all $P \in \mathcal{P}$. Then $\psi(\theta, P) \in l^\infty(\mathcal{H})$ for each fixed $\theta \in \Theta$. The empirical process $\mathbb{G}_n B(\theta)$ acting on $B(\theta)$ is also a function in $l^\infty(\mathcal{H})$ for fixed $\theta \in \Theta$.

A *functional Z-estimator* for θ_0 is a sequence of estimates $\{\hat{\theta}_n\} \in \Theta$ which makes the "scores" $\mathbb{P}_n B(\theta)(h)$, $h \in \mathcal{H}$, approximately zero: $\|\psi(\hat{\theta}_n, \mathbb{P}_n)\|_{\mathcal{H}} = o_{P^*}(n^{-1/2})$, where P^* denotes the outer probability of P^∞ (See Van der Vaart and Wellner (1996) for more details on outer probability measures).

2.1. A differential identity

The function $\psi(\theta, P)$, as a map from Θ to $l^\infty(\mathcal{H})$, is Fréchet differentiable with respect to the norm $\|\cdot\|$ at a point $\vartheta \in \Theta$ if there is a bounded linear operator $\dot{\psi}(\vartheta, P_\vartheta)(\cdot)$ mapping from $(lin(\Theta), \|\cdot\|)$ to $(l^\infty(\mathcal{H}), \|\cdot\|_{\mathcal{H}})$ such that $\|\psi(\theta, P_\vartheta) - \psi(\vartheta, P_\vartheta) - \dot{\psi}(\vartheta, P_\vartheta)(\theta - \vartheta)\|_{\mathcal{H}} = o(\|\theta - \vartheta\|)$. Denote the operator $\dot{\psi}(\theta, P_\theta)$ by $\dot{\psi}(\theta)$: $\dot{\psi}(\theta) \equiv \dot{\psi}(\theta, P_\theta)$.

Recall that for a fixed $\vartheta \in \Theta$, the operator $B(\vartheta)$ is bounded in the sense that $\|PB(\vartheta)\|_{\mathcal{H}} < \infty$ for all $P \in \mathcal{P}$. Thus for a fixed $\vartheta \in \Theta$, the probability measure P_θ induces a mapping $\theta \mapsto P_\theta B(\vartheta)$ from Θ to $l^\infty(\mathcal{H})$. The map $P_\theta B(\vartheta)$, as a function of θ, is Fréchet differentiable with respect to the norm $\|\cdot\|$ at a point $\vartheta \in \Theta$ if there is a linear operator $\dot{P}_\vartheta(\cdot)$ such that $\dot{P}_\vartheta(\cdot)B(\vartheta)$ is bounded and $\|P_\theta B(\vartheta) - P_\vartheta B(\vartheta) - \dot{P}_\vartheta(\theta - \vartheta)B(\vartheta)\|_{\mathcal{H}} = o(\|\theta - \vartheta\|)$.

Lemma 2.1. *Assume that $\psi(\theta, P_\theta) \equiv 0$ for all $\theta \in \Theta$. For any $\vartheta \in \Theta$, suppose that $\psi(\theta, P)$ is Fréchet differentiable with respect to the norm $\|\cdot\|$ in a neighborhood of ϑ, and the operator $\dot{\psi}(\theta)$ is continuous as a function of θ at ϑ:*

$$\left\|\dot{\psi}(\theta) - \dot{\psi}(\vartheta)\right\| \equiv \sup_{\|a\| \le 1} \left\|\dot{\psi}(\theta)(a) - \dot{\psi}(\vartheta)(a)\right\|_{\mathcal{H}} \longrightarrow 0 \tag{2.1}$$

as $\|\theta - \vartheta\| \to 0$. If $P_\theta B(\vartheta)$ is Fréchet differentiable with respect to the norm $\|\cdot\|$ at $\vartheta \in \Theta$, then the operator $\psi(\theta, P_\theta)$ as a function of θ is Fréchet differentiable with respect to the norm $\|\cdot\|$ at $\vartheta \in \Theta$ and the following identity holds for all $a \in lin(\Theta)$:

$$\dot{\psi}(\vartheta)(a) + \dot{P}_\vartheta(a)B(\vartheta) = 0 \tag{2.2}$$

in $l^\infty(\mathcal{H})$.

Proof. For any $\vartheta \in \Theta$, write the difference

$$\begin{aligned} &\psi(\theta, P_\theta) - \psi(\vartheta, P_\vartheta) = P_\theta B(\theta) - P_\vartheta B(\vartheta) \\ &= P_\vartheta(B(\theta) - B(\vartheta)) + (P_\theta - P_\vartheta)B(\vartheta) + (P_\theta - P_\vartheta)(B(\theta) - B(\vartheta)). \end{aligned} \tag{2.3}$$

Since $\psi(\theta, P)$ is Fréchet differentiable at ϑ, the first term on the right-hand side of (2.3) can be written as $P_\vartheta(B(\theta) - B(\vartheta)) = \psi(\theta, P_\vartheta) - \psi(\vartheta, P_\vartheta) = \dot{\psi}(\vartheta, P_\vartheta)(\theta - \vartheta) + o(\|\theta - \vartheta\|)$. The map $P_\theta B(\vartheta)$ as a function of θ is Fréchet differentiable with respect to the norm $\|\cdot\|_D$ at ϑ, the second term in (2.3) is actually $(P_\theta - P_\vartheta)B(\vartheta) = \dot{P}_\vartheta(\theta - \vartheta)B(\vartheta) + o(\|\theta - \vartheta\|)$. Because P_θ acts on $B(\vartheta)$ linearly, the third term on the right-hand side of (2.3) can be written as

$$\begin{aligned}&(P_\theta - P_\vartheta)(B(\theta) - B(\vartheta)) = P_\theta(B(\theta) - B(\vartheta)) - P_\vartheta(B(\theta) - B(\vartheta))\\&= -\dot{\psi}(\theta)(\vartheta - \theta) - o(\|\vartheta - \theta\|) - \dot{\psi}(\vartheta)(\theta - \vartheta) - o(\|\theta - \vartheta\|).\end{aligned}$$

The first two terms in the last equality are obtained by applying the Fréchet differentiability of $\psi(\vartheta, P_\theta) = P_\theta B(\vartheta)$ at θ.

Thus, by the triangle inequality and the continuity of $\dot{\psi}(\theta)$, we have $\|(P_\theta - P_\vartheta)(B(\theta) - B(\vartheta))\|_{\mathcal{H}} = o(\|\theta - \vartheta\|)$. Therefore $\psi(\theta, P_\theta)$ as a function of θ is Fréchet differentiable with respect to $\|\cdot\|$ at ϑ and its Fréchet derivative is given by $\dot{\psi}(\vartheta, P_\vartheta)(a) + \dot{P}_\vartheta(a)B(\vartheta)$. But $\psi(\theta, P_\theta) \equiv 0$ and we have the identity in (2.2) by the uniqueness of the Fréchet derivative.

2.2. A condition of uniform boundedness

The uniform boundedness of the operators $\dot{\psi}(\theta)$ is needed to establish the rate of convergence for a sequence of *Z*-estimators $\{\hat{\theta}_n\}$. This property is also needed to asymptotically replace $\dot{\psi}^{-1}(\hat{\theta}_n)(\mathbb{G}_n B(\theta_0))$ by $\dot{\psi}^{-1}(\theta_0)(\mathbb{G}_n B(\theta_0))$ for a consistent $\hat{\theta}_n$, and thus allows us to apply the Continuous Mapping Theorem on $\dot{\psi}^{-1}(\theta_0)(\mathbb{G}_n B(\theta_0))$ to obtain a central limit theorem.

Given that Θ is a subset in a Banach space $(\mathbf{B}, \|\cdot\|)$, the closure $\overline{lin(\Theta)}$ is a Banach space with the same norm $\|\cdot\|$ (Lemma II.1.3 on page 50, Dunford and Schwartz (1988), Part I). Because $(l^\infty(\mathcal{H}), \|\cdot\|_{\mathcal{H}})$ is also a Banach space, the bounded operators $\dot{\psi}^{-1}(\theta)$ and $\dot{\psi}(\theta)$ can be uniquely extended to the closures of their domains by continuity (see, e.g., Lemma I.6.16 on page 23 of Dunford and Schwartz (1988), Part I).

The unique continuous extensions of $\dot{\psi}^{-1}(\theta)$ and $\dot{\psi}(\theta)$ on the closures of their domains are also denoted by $\dot{\psi}^{-1}(\theta)$ and $\dot{\psi}(\theta)$. The extension $\dot{\psi}^{-1}(\theta)$ on $\overline{\mathcal{R}(\dot{\psi})}$ is also the inverse of the extension $\dot{\psi}(\theta)$ on $\overline{lin(\Theta)}$. For the examples we deal with in Section 3, and for other examples, it is true that $\overline{\mathcal{R}(\dot{\psi})}$ does not depend on θ. We use $\overline{\mathcal{R}(\dot{\psi})}$ instead of $\overline{\mathcal{R}(\dot{\psi}(\theta))}$ to denote the common subspace on which every $\dot{\psi}^{-1}(\theta)$ resides.

Lemma 2.2. *Suppose that, for every fixed $\theta \in \Theta$, the operator $\dot{\psi}(\theta)$ mapping from $(\overline{lin(\Theta)}, \|\cdot\|_K)$ to $(l^\infty(\mathcal{H}), \|\cdot\|_{\mathcal{H}})$ has a bounded inverse $\dot{\psi}^{-1}(\theta)$ on a fixed* *subspace space* $\overline{\mathcal{R}(\dot\psi)} \subset l^\infty(\mathcal{H})$. *Further assume that* $\dot\psi^{-1}(\theta)$ *converges on* $\overline{\mathcal{R}(\dot\psi)}$ *to* $\dot\psi^{-1}(\theta_0)$ *with respect to a norm* $\|\cdot\|_K$*: for any* $f \in \overline{\mathcal{R}(\dot\psi)}$

$$\left\|\dot\psi^{-1}(\theta)(f) - \dot\psi^{-1}(\theta_0)(f)\right\|_K \longrightarrow 0 \tag{2.4}$$

as $\|\theta-\theta_0\| \to 0$. *Assume that* $\|\hat\theta_n - \theta_0\| \to_{P^*} 0$ *and that* $\mathbb{G}_n B(\theta_0) \Rightarrow \mathbb{Z}_0$ *in* $l^\infty(\mathcal{H})$ *as* $n \to \infty$. *Then* $\|(\dot\psi^{-1}(\hat\theta_n) - \dot\psi^{-1}(\theta_0))(\mathbb{G}_n B(\theta_0))\|_K = o_{P^*}(1)$.

Proof. For any compact set $C \subset \overline{\mathcal{R}(\dot\psi)} \subset l^\infty(\mathcal{H})$, let $C(\delta)$ be the δ-enlargement of C defined by $C(\delta) = \{f \in \overline{\mathcal{R}(\dot\psi)} : \ \|f - f'\|_{\mathcal{H}} \le \delta \text{ for some } f' \in C\}$. We show that

$$\sup\left\{ \left\|(\dot\psi^{-1}(\theta) - \dot\psi^{-1}(\theta_0))(f)\right\|_K : \ f \in C(\delta) \right\} \longrightarrow 0 \tag{2.5}$$

as $\|\theta - \theta_0\| \to 0$ and then $\delta \to 0+$.

In fact, by (2.4) and the Banach-Steinhaus Theorem, the operator norm of $\dot\psi^{-1}(\theta)$ is uniformly bounded: $\sup_{\|\theta-\theta_0\|\le\beta} \|\dot\psi^{-1}(\theta)\|_K \le M < \infty$ for some positive numbers $\beta > 0$ and $M > 0$. The uniform boundedness of the operators $\dot\psi^{-1}(\theta)$ is equivalent to their uniform continuity as mappings in Banach spaces, so that the pointwise convergence in (2.4) directly implies the uniform convergence in the norm $\|\cdot\|_K$.

Now since $\mathbb{G}_n B(\theta_0) \in \mathcal{R}(\dot\psi)$ converges weakly to $\mathbb{Z}_0$ in $(\overline{\mathcal{R}(\dot\psi)}, \|\cdot\|_{\mathcal{H}})$, it is asymptotically tight: for every $\epsilon > 0$ there exists a compact set $C \subset \overline{\mathcal{R}(\dot\psi)}$ such that $\liminf_{n\to\infty} P_* \{\mathbb{G}_n B(\theta_0) \in C(\delta)\} \ge 1 - \epsilon$ for every $\delta > 0$; see Van der Vaart and Wellner (1996), Section 1.3. Thus by (2.5), we have $\|(\dot\psi^{-1}(\hat\theta_n) - \dot\psi^{-1}(\theta_0))(\mathbb{G}_n B(\theta_0))\|_K = o_{P^*}(1)$ as $n \to \infty$ and then $\delta \to 0+$.

2.3. A central limit theorem

We need the following assumptions for a central limit theorem.

L.1 For all $\theta \in \Theta$, $\psi(\theta, P_\theta) = P_\theta B(\theta) \equiv 0$ in $l^\infty(\mathcal{H})$.

L.2 As $n \to \infty$, for any decreasing $\delta_n \downarrow 0$, the stochastic equicontinuity condition $\sup\{\|\mathbb{G}_n(B(\theta) - B(\theta_0))\|_{\mathcal{H}} : \ \|\theta - \theta_0\| \le \delta_n\} = o_{P^*}(1)$ holds at the point θ_0.

L.3 At the point θ_0, $\mathbb{G}_n B(\theta_0) \Rightarrow \mathbb{Z}_0$ in $l^\infty(\mathcal{H})$, where $\Rightarrow$ indicates weak convergence in $l^\infty(\mathcal{H})$ to a tight Borel measurable random element $\mathbb{Z}_0$.

L.4 For a fixed $\vartheta \in \Theta$, the operator $P_\theta B(\vartheta)$ as a function of θ is Fréchet differentiable with respect to the norm $\|\cdot\|$ at ϑ. Furthermore, the function $\theta \mapsto \psi(\theta, P)$ from Θ to $l^\infty(\mathcal{H})$ is Fréchet differentiable with respect to the norm $\|\cdot\|$. The operator $\dot\psi(\theta)$ is continuous as a function of θ in the sense of (2.1).

L.5 For every fixed $\theta \in \Theta$, the operator $\dot{\psi}(\theta)$ from $(\overline{lin(\Theta)}, \|\cdot\|)$ to $(l^\infty(\mathcal{H}), \|\cdot\|_{\mathcal{H}})$ has a bounded inverse $\dot{\psi}^{-1}(\theta)$ on a fixed subspace $\overline{\mathcal{R}(\dot{\psi})} \subset l^\infty(\mathcal{H})$. Furthermore $\dot{\psi}^{-1}(\theta)$ as an operator sequence converges to $\dot{\psi}^{-1}(\theta_0)$ as $\|\theta - \theta_0\| \to 0$:

$$\left\|\dot{\psi}^{-1}(\theta)(f) - \dot{\psi}^{-1}(\theta_0)(f)\right\| \longrightarrow 0 \tag{2.6}$$

for all $f \in \overline{\mathcal{R}(\dot{\psi})}$.

Theorem 2.1. *Let $\|\hat{\theta}_n - \theta_0\| \to_{P^*} 0$ be a sequence of consistent Z-estimators. Assume L.1 through L.5. Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow -\dot{\psi}^{-1}(\theta_0)(\mathbb{Z}_0)$ in $(\overline{lin(\Theta)}, \|\cdot\|)$.*

We begin to prove Theorem 2.1 with the following lemma. It asserts that the standardized estimating equations behave asymptotically as $\mathbb{G}_n B(\theta_0)$ under our assumptions.

Lemma 2.3. *Let L.1 and L.2 hold. Then $\sqrt{n}\psi(\hat{\theta}_n, P) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(1)$.*

Proof. Since $\psi(\theta, P_\theta) \equiv 0$ for all $\theta \in \Theta$, we have by the definitions of $\psi(\theta, P)$ and the Z-estimator $\|\sqrt{n}\mathbb{P}_n B(\hat{\theta}_n)\|_{\mathcal{H}} = o_{P^*}(1)$:

$$\mathbb{G}_n B(\theta_0) + \sqrt{n}(\psi(\hat{\theta}_n, P) - \psi(\theta_0, P)) = -\mathbb{G}_n(B(\hat{\theta}_n) - B(\theta_0)) + o_{P^*}(1).$$

By L.2, the consistency of $\hat{\theta}_n$, and

$$P^*\{\|\mathbb{G}_n(B(\theta) - B(\theta_0))\|_{\mathcal{H}} \geq \epsilon\} \leq P^*\{\sup_{\|\theta - \theta_0\| \leq \delta} \|\mathbb{G}_n(B(\theta) - B(\theta_0))\|_{\mathcal{H}} \geq \epsilon\} + P^*\{\|\hat{\theta}_n - \theta_0\| > \delta\},$$

it follows that $\|\mathbb{G}_n(B(\hat{\theta}_n) - B(\theta_0))\|_{\mathcal{H}} = o_{P^*}(1)$. Hence $\|\mathbb{G}_n B(\theta_0) + \sqrt{n}(\psi(\hat{\theta}_n, P) - \psi(\theta_0, P))\|_{\mathcal{H}} \leq \|-\mathbb{G}_n(B(\hat{\theta}_n) - B(\theta_0))\|_{\mathcal{H}} + o_{P^*}(1) = \Delta_n(\hat{\theta}_n) + o_{P^*}(1) = o_{P^*}(1)$. The conclusion of the lemma follows from this inequality and the fact that $\psi(\theta_0, P) = 0$.

Remark 2.1. In order to obtain the conclusion of Lemma 2.3, it suffices to verify $\Delta_n(\hat{\theta}_n) = o_{P^*}(1)$ for consistent $\hat{\theta}_n$, although this weaker condition is usually verified through the stronger L.2.

The next lemma shows that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is actually $O_{P^*}(1)$ under the assumptions.

Lemma 2.4. *Assume L.1 through L.5 and that $\hat{\theta}_n$ is consistent: $\|\hat{\theta}_n - \theta_0\| \to_{P^*} 0$. Then $\sqrt{n}\|\hat{\theta}_n - \theta_0\| = O_{P^*}(1)$.*

Proof. Mapping from the Banach space $\overline{\mathcal{R}(\dot{\psi})}$ to the Banach space $\overline{lin(\Theta)}$, the sequence of continuous linear operators $\dot{\psi}^{-1}(\theta)$ converges on $\overline{\mathcal{R}(\dot{\psi})}$ to $\dot{\psi}^{-1}(\theta_0)$ as

$\|\theta-\theta_0\| \to 0$ by L.5. Hence, by the Banach-Steinhaus Theorem (for example, Theorem II.3.6 on page 60 of Dunford and Schwartz (1988), or Theorem 2 on page 203 of Kantorovich and Akilov (1982)), the norm of the operators $\dot\psi^{-1}(\theta)$ is uniformly bounded: $\sup_{\|\theta-\theta_0\|\le\beta}\|\dot\psi^{-1}(\theta)\| \le 1/\alpha < \infty$ for some positive numbers $0<\alpha<\infty$ and $\beta>0$.

Thus for any $a \in \overline{lin(\Theta)}$, we have $\|a\| = \|\dot\psi^{-1}(\theta)(\dot\psi(\theta)(a))\| \le \|\dot\psi^{-1}(\theta)\| \times\|\dot\psi(\theta)(a)\|_{\mathcal{H}} \le (1/\alpha)\|\dot\psi(\theta)(a)\|_{\mathcal{H}}$. We have not made a distinction between the operator norm and the norm on $\mathbf{B}$. The meaning of $\|\cdot\|$ should be clear from the context. Hence

$$\alpha\|a\| \le \left\|\dot\psi(\theta)(a)\right\|_{\mathcal{H}} \tag{2.7}$$

for all θ such that $\|\theta-\theta_0\|\le\beta$.

Take $a=\sqrt{n}(\hat\theta_n-\theta_0)$ and $\vartheta=\hat\theta_n$ in identity (2.2). By the linearity of $\dot P_\vartheta(a)B(\vartheta)$ in a and the definition of Fréchet differentiability of $P_\theta B(\vartheta)$ as a function of θ,

$$\begin{aligned}\dot\psi(\hat\theta_n)(\sqrt{n}(\hat\theta_n-\theta_0)) &= -\dot P_{\hat\theta_n}(\sqrt{n}(\hat\theta_n-\theta_0))B(\hat\theta_n) = \sqrt{n}\dot P_{\hat\theta_n}(\theta_0-\hat\theta_n)B(\hat\theta_n)\\ &= \sqrt{n}(P_{\theta_0}B(\hat\theta_n) - P_{\hat\theta_n}B(\hat\theta_n)) + o_{P^*}(\sqrt{n}\|\hat\theta_n-\theta_0\|)\\ &= \sqrt{n}\psi(\hat\theta_n,P) + o_{P^*}(\sqrt{n}\|\hat\theta_n-\theta_0\|).\end{aligned}$$

We have used P to denote P_{θ_0} in the last equality. Therefore, by the boundedness in (2.7) we obtain: $\alpha\sqrt{n}\|\hat\theta_n-\theta_0\| \le \|\dot\psi(\hat\theta_n)(\sqrt{n}(\hat\theta_n-\theta_0))\| \le \sqrt{n}\|\psi(\hat\theta_n,P)\|_{\mathcal{H}} + o_{P^*}(1)\cdot\sqrt{n}\|\hat\theta_n-\theta_0\|$ in P^*-probability when n is sufficiently large. The conclusion of the lemma follows from Lemma 2.3 and L.3 which assert that the term $\sqrt{n}\|\psi(\hat\theta_n,P)\|_{\mathcal{H}}$ is of an order of $O_{P^*}(1)+o_{P^*}(1)$.

Proof of Theorem 2.1. By the Fréchet differentiability of $P_\theta B(\vartheta)$ at ϑ we have $P_\theta B(\vartheta) - P_\vartheta B(\vartheta) - \dot P_\vartheta(\theta-\vartheta)B(\vartheta) = o(\|\theta-\vartheta\|)$. Substituting $\hat\theta_n$ for ϑ and θ_0 for θ, and using P to denote P_{θ_0}, we obtain $\dot P_{\hat\theta_n}(\hat\theta_n-\theta_0)B(\hat\theta_n) = P_{\hat\theta_n}B(\hat\theta_n) - P_{\theta_0}B(\hat\theta_n) + o_{P^*}(\|\hat\theta_n-\theta_0\|) = -\psi(\hat\theta_n,P) + o_{P^*}(\|\hat\theta_n-\theta_0\|)$. Note that $\dot\psi(\hat\theta_n) \equiv \dot\psi(\hat\theta_n, P_{\hat\theta_n})$, and by the identity (2.2) we have

$$\begin{aligned}\dot\psi(\hat\theta_n)(\sqrt{n}(\hat\theta_n-\theta_0)) &= -\sqrt{n}\dot P_{\hat\theta_n}(\hat\theta_n-\theta_0)B(\hat\theta_n) = \sqrt{n}\psi(\hat\theta_n,P) + o_{P^*}(\sqrt{n}\|\hat\theta_n-\theta_0\|)\\ &= \sqrt{n}\psi(\hat\theta_n,P) + o_{P^*}(1).\end{aligned}$$

The last equality follows from the consistency of $\hat\theta_n$, L.1 through L.5 and Lemma 2.4.

Note that by L.5 the operator sequence $\dot\psi^{-1}(\theta)$ converges to $\dot\psi^{-1}(\theta_0)$ on $\overline{\mathcal{R}(\dot\psi)}$ as $\|\theta-\theta_0\|\to 0$. Hence the Banach-Steinhaus Theorem and the consistency of $\hat\theta_n$ imply that the operator norm of $\dot\psi^{-1}(\hat\theta_n)$ is uniformly bounded in P^*-probability when n is sufficiently large. It maps a term of $o_{P^*}(1)$ in the $\|\cdot\|_{\mathcal{H}}$-norm into a term of $o_{P^*}(1)$ in $\|\cdot\|$-norm: $\dot{\psi}^{-1}(\hat{\theta}_n)(o_{P^*}(1)) = o_{P^*}(1)$. This means that

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n - \theta_0) &= \dot{\psi}^{-1}(\hat{\theta}_n)(\sqrt{n}\psi(\hat{\theta}_n, P) + o_{P^*}(1)) = \dot{\psi}^{-1}(\hat{\theta}_n)(-\mathbb{G}_n B(\theta_0) + o_{P^*}(1)) \\ &= -\dot{\psi}^{-1}(\hat{\theta}_n)(\mathbb{G}_n B(\theta_0)) + o_{P^*}(1).\end{aligned}$$

The second equality follows from Lemma 2.3.

By the triangle inequality and Lemma 2.2 (applied with the K-norm replaced by $\|\cdot\|$) we obtain $\dot{\psi}^{-1}(\hat{\theta}_n)(\mathbb{G}_n B(\theta_0)) = \dot{\psi}^{-1}(\theta_0)(\mathbb{G}_n B(\theta_0)) + o_{P^*}(1)$. Hence we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow -\dot{\psi}^{-1}(\theta_0)(\mathbb{Z}_0)$ in $(\overline{lin(\Theta)}, \|\cdot\|)$ as $n \to \infty$ by the Continuous Mapping Theorem.

2.4. Bounded convex linearity

The parametrization $\theta \mapsto P_\theta$ is said to be *convex linear* if $\theta = \lambda_1\theta_1 + \lambda_2\theta_2 \in lin(\Theta)$ implies $P_\theta = \lambda_1 P_{\theta_1} + \lambda_2 P_{\theta_2} \in \mathcal{P}$ for any $\theta_1, \theta_2 \in \Theta$ and any real numbers λ_1 and λ_2 such that $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$. Convex linearity is referred to as bounded with respect to a norm $\|\cdot\|$ on $lin(\Theta)$ if

L.6 For any $\theta_1, \ldots, \theta_k$ in Θ, and any real numbers $\lambda_1, \ldots, \lambda_k$, $k \geq 1$, there is a constant $C < \infty$ such that

$$\Big\| \sum_{i=1}^{k} \lambda_i P_{\theta_i} B(\vartheta) \Big\|_{\mathcal{H}} \leq C \Big\| \sum_{i=1}^{k} \lambda_i \theta_i \Big\| \tag{2.8}$$

holds for every fixed $\vartheta \in \Theta$, where $B(\vartheta)$ is the score operator mapping from $\mathcal{H}$ to $\mathcal{F}$.

Lemma 2.5. *If the parametrization $\theta \mapsto P_\theta$ is boundedly convex linear, then the mapping $P_\theta B(\vartheta)$ is Fréchet differentiable with respect to the norm $\|\cdot\|$ at all $\vartheta \in \Theta$ and the derivative operator $\dot{P}_\vartheta(\cdot)B(\vartheta)$ is given by $\dot{P}_\vartheta(\theta_1 - \theta_2)B(\vartheta) = P_{\theta_1}B(\vartheta) - P_{\theta_2}B(\vartheta)$ for any θ_1, θ_2 and ϑ in Θ.*

Proof. Let $\theta = \sum_{i=1}^{k} \lambda_i\theta_i \in lin(\Theta)$ be a linear combination of the θ_i's. We want to prove that

$$L_\theta B(\vartheta) = \sum_{i=1}^{k} \lambda_i P_{\theta_i} B(\vartheta) \tag{2.9}$$

is a bounded linear extension of $P_\theta B(\vartheta)$ to $lin(\Theta)$.

First by (2.8), if a linear combination of the elements $\theta_1, \ldots, \theta_k$ is equal to the zero element $\sum_{i=1}^{k} \lambda_i\theta_i = 0$, then $\sum_{i=1}^{k} \lambda_i P_{\theta_i} B(\vartheta) = 0$ as well. From this observation, the value of the mapping $L_\theta B(\vartheta)$ is uniquely determined by

$\theta \in lin(\Theta)$. It is not hard to verify that $L_\theta B(\vartheta)$ is a linear mapping from $lin(\Theta)$ to $l^\infty(\mathcal{H})$. The boundedness $\|L_\theta B(\vartheta)\|_\mathcal{H} \le C\|\theta\|$ of $L_\theta B(\vartheta)$ follows from (2.8). And it is easy to verify that $L_\theta B(\vartheta)$ is an extension of $P_\theta B(\vartheta)$ to $lin(\Theta)$ with $L_\theta B(\vartheta) \equiv P_\theta B(\vartheta)$ for all $\theta \in \Theta$ by (2.9).

For any bounded linear mapping $A : lin(\Theta) \mapsto l^\infty(\mathcal{H})$, the Fréchet derivative of A is simply A itself because $A(\theta') - A(\theta) = A(\theta' - \theta)$ is bounded and linear. Now the mapping $L_\theta B(\vartheta) : lin(\Theta) \mapsto l^\infty(\mathcal{H})$ is bounded and linear, hence it is Fréchet differentiable at $\theta \in lin(\Theta)$, and its derivative operator is given by $L_a B(\vartheta)$ for $a \in lin(\Theta)$.

Since $L_\theta B(\vartheta) = P_\theta B(\vartheta)$ for any $\theta \in \Theta$, we have $\dot{P}_\theta(a)B(\vartheta) = L_a B(\vartheta)$ by the uniqueness of the Fréchet derivative. Therefore, for $a = \theta_1 - \theta_2$ with θ_1 and θ_2 belonging to Θ, we have $\dot{P}_\theta(\theta_1 - \theta_2)B(\vartheta) = L_{(\theta_1-\theta_2)}B(\vartheta) = L_{\theta_1}B(\vartheta) - L_{\theta_2}B(\vartheta) = P_{\theta_1}B(\vartheta) - P_{\theta_2}B(\vartheta)$, which completes the proof of the lemma.

In view of Lemma 2.5, the differential identity (2.2) for models with bounded convex linearity can be improved to

$$\dot{\psi}(\vartheta)(\theta_1 - \theta_2) = -(P_{\theta_1} - P_{\theta_2})B(\vartheta). \tag{2.10}$$

for any θ_1, θ_2, and $\vartheta \in \Theta$.

For these models, a strong enough norm $\|\cdot\|$ may be used to obtain the differentiability of $\psi(\theta, P_\theta)$ and condition L.6, and therefore the identity (2.10). Then a weaker norm $\|\cdot\|_K$ can be used to establish the invertibility of $\dot{\psi}^{-1}(\theta)$ and the pointwise convergence in (2.4). The difference on the right of (2.10) also implies that no rate control, such as that in Lemma 2.4, is needed. Because of this, we can actually obtain asymptotic normality with the weaker norm. This usually improves the applicability of the Central Limit Theorem. See Section 3.2 and 3.3 for two examples. To be more specific, the assumptions replacing L.4 and L.5 are the followings.

L.4' The function $\psi(\theta, P)$ as a map from Θ to $l^\infty(\mathcal{H})$ is Fréchet differentiable with respect to the norm $\|\cdot\|$. The operator $\dot{\psi}(\theta)$ is continuous as a function of θ in the sense of (2.1).

L.5' For every fixed $\theta \in \Theta$, the operator $\dot{\psi}(\theta)$ from $(\overline{lin(\Theta)}, \|\cdot\|_K)$ to $(l^\infty(\mathcal{H}), \|\cdot\|_\mathcal{H})$ has a bounded inverse $\dot{\psi}^{-1}(\theta)$ on a fixed subspace $\overline{\mathcal{R}(\dot{\psi})} \subset l^\infty(\mathcal{H})$. Furthermore $\dot{\psi}^{-1}(\theta)$ as operator sequence converges to $\dot{\psi}^{-1}(\theta_0)$ as $\|\theta - \theta_0\| \to 0$:

$$\left\|\dot{\psi}^{-1}(\theta)(f) - \dot{\psi}^{-1}(\theta_0)(f)\right\|_K \longrightarrow 0 \tag{2.11}$$

for all $f \in \overline{\mathcal{R}(\dot{\psi})}$.

Theorem 2.2. *For a model with bounded convex linearity specified in L.6, assume L.1 through L.3, L.4' and L.5'. Then a sequence of consistent Z-estimators $\hat{\theta}_n$ such that $\|\hat{\theta}_n - \theta_0\| \to_{P^*} 0$ is actually asymptotically normal: $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow -\dot{\psi}^{-1}(\theta_0)(\mathbb{Z}_0)$ in $(\overline{lin(\Theta)}, \|\cdot\|_K)$.*

Proof. By Lemma 2.5, take $\theta_1 = \hat{\theta}_n$, $\theta_2 = \theta_0$ and $\vartheta = \hat{\theta}_n$ in (2.10), and use P to denote P_{θ_0}, we obtain $\dot{\psi}(\hat{\theta}_n)(\sqrt{n}(\hat{\theta}_n - \theta_0)) = \sqrt{n}\psi(\hat{\theta}_n, P) = -\mathbb{G}_n B(\theta_0) + o_{P^*}(1)$. The last equality follows from Lemma 2.3 and the term $o_{P^*}(1)$ in the above denotes a term whose $\|\cdot\|_{\mathcal{H}}$-norm is of order $o_{P^*}(1)$.

Since $\dot{\psi}^{-1}(\theta)$ converges to $\dot{\psi}^{-1}(\theta_0)$ on $\overline{\mathcal{R}(\dot{\psi})}$, the Banach-Steinhaus Theorem implies that the operator norm of $\dot{\psi}^{-1}(\hat{\theta}_n)$ is uniformly bounded in P^*-probability when n is sufficiently large. It then maps a term of $o_{P^*}(1)$ in the $\|\cdot\|_{\mathcal{H}}$-norm into a term of $o_{P^*}(1)$ in K-norm: $\dot{\psi}^{-1}(\hat{\theta}_n)(o_{P^*}(1)) = o_{P^*}(1)$. This means that $\sqrt{n}(\hat{\theta}_n - \theta_0) = \dot{\psi}^{-1}(\hat{\theta}_n)(-\mathbb{G}_n B(\theta_0) + o_{P^*}(1)) = -\dot{\psi}^{-1}(\hat{\theta}_n)(\mathbb{G}_n B(\theta_0)) + o_{P^*}(1)$. By Lemma 2.2, $\dot{\psi}^{-1}(\hat{\theta}_n)(\mathbb{G}_n B(\theta_0)) = \dot{\psi}^{-1}(\theta_0)(\mathbb{G}_n B(\theta_0)) + o_{P^*}(1)$. Hence $\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow -\dot{\psi}^{-1}(\theta_0)(\mathbb{Z}_0)$ in $(\overline{lin(\Theta)}, \|\cdot\|_K)$ as $n \to \infty$ by the Continuous Mapping Theorem.

Theorem 2.2 is mainly motivated by Vardi and Zhang (1992) on the multiplicative censoring model, and Gu and Zhang (1993) on the double censoring model. The proof of Theorem 2.2 is an extension of their arguments of asymptotic normality from these two important examples to a general model. We give a differential identity for models with convex linearity, and use that as the basis of our argument (Lemma 2.1). The key assumptions are formulated in L.4' and L.5' which are not explicit in these two papers. Other assumptions such as L.1, L.2 and L.3 are mainly from the traditional arguments, see Huber (1964, 1967), Pakes and Polard (1989), Polard (1985, 1989), and Van der Vaart (1994, 1995). For models not necessarily convex linear, Theorem 2.1 generalizes the argument. A non-trivial application of Theorem 2.1 is given in the next section.

3. Applications

In this section we give three examples to illustrate the application of Theorem 2.1 and Theorem 2.2. The first example concerns a simplified frailty model without convex linearly parameterization. Because of its nice analytical properties, the asymptotic normality of the MLE can be established by traditional arguments. The same conclusion can also be obtained by Theorem 2.1. The second example deals with the multiplicative censoring model which has a convex linear parameterization. In this model the likelihood equations are only differentiable with respect to a $\|\cdot\|_D$ norm that is stronger than the uniform norm. The derivative operator, however, is only invertible with respect to the uniform norm.

The third example concerns the double censoring model which has a convex linear parameterization. Here, the likelihood equations are only differentiable with respect to the uniform norm, and the derivative operator is only invertible with respect to a weaker $\|\cdot\|_K$ norm. Because of these difficulties, the traditional argument does not apply in these examples. Central limit theorems are instead established by invoking Theorem 2.2.

3.1. A simplified frailty model

Let $Z \sim \text{Gamma}(\nu_0, 1)$ be a known gamma frailty. Conditional on $Z = z$, we observe independent random variables (X, Y) with a common, absolutely continuous hazard function $z\Lambda_0$. Based on n i.i.d. observations (X_i, Y_i) with distribution $P\{X > x, Y > y\} = 1/[1 + \Lambda_0(x) + \Lambda_0(y)]^{\nu_0}$, we are interested in estimating Λ_0 on $[0, \tau]$, where $\tau < \infty$ is a real number such that $\Lambda_0(\tau) < \infty$.

Let $\Theta \subset l^\infty(\mathcal{H}_p)$ be the parameter space, where $\mathcal{H}_p$ is a set of real functions h defined on $[0, \infty)$ with bounded variation $\|h\|_v < p$ on $[0, \tau]$ and identical to zero on (τ, ∞). The set $\mathcal{H}_p$ is considered as a space equipped with the variation norm $\|\cdot\|_v$ defined by $\|h\|_v \equiv |h(0)| + \vee_0^\tau(h)$. A bounded linear functional $\Lambda(h) \in l^\infty(\mathcal{H}_p)$ is given by $\Lambda(h) = \int_{[0,\infty)} h(x)d\Lambda(x)$ with $\|\Lambda\|_{\mathcal{H}_p} = \sup_{h\in\mathcal{H}_p} |\int_{[0,\infty)} h(x)d\Lambda(x)| < \infty$. The parameter space Θ can thus be identified with all absolutely continuous integrated hazard functions Λ restricted to the interval $[0, \tau]$, such that $\Lambda(u) \equiv \Lambda_0(u)$ for $u > \tau$. We will not distinguish between a functional $\Lambda \in \Theta$ and a hazard function $\Lambda(u)$.

The score operator $B(\Lambda)$ is obtained by differentiating the log-likelihood along a curve passing through $\Lambda \in \Theta$. It is a function of Λ mapping from $\mathcal{H}_p$ to a set $\mathcal{F}$ of $L_2(P)$ functions defined on the sample space:

$$B(\Lambda)(h)(x,y) = h(x) + h(y) - (\nu_0 + 2)\frac{\int_{[0,x]} h(u)d\Lambda(u) + \int_{[0,y]} h(u)d\Lambda(u)}{1 + \Lambda(x) + \Lambda(y)}. \quad (3.1)$$

Murphy (1995) considers this model in the context of counting processes and proves the asymptotic normality of the MLE following the traditional argument. Van der Vaart (1995) uses this model as an example to motivate the Central Limit Theorem for functional parameters, also following the traditional argument. In this section, we show that the same asymptotic results are obtainable from Theorem 2.1.

Verification of L.1, L.2 *and* L.3. It is straightforward to verify that $\psi(\Lambda, P_\Lambda)(h) = \int_{R^+} B(\Lambda)(h)(x,y)dP_\Lambda(x,y) = 0$ for all $\Lambda \in \Theta$ and $h \in \mathcal{H}_p$ by interchanging the order of the integrations involved.

To verify L.2, let $\mathcal{C}_M$ be the function class defined on the sample space R^{2+} given by

$$\mathcal{C}_M = \{B(\Lambda)(h)(x,y) : \ h \in \mathcal{H}_p, \ \|\Lambda\|_{\mathcal{H}_p} \le M\}, \quad (3.2)$$

with $0 < M < \infty$ a fixed number. Now consider a small difference class $\mathcal{F}_{\delta_n} = \{B(\Lambda)(h) - B(\Lambda_0)(h) : \; h \in \mathcal{H}_p, \|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \le \delta_n\}$. Set $a = \Lambda - \Lambda_0$. As $n \to \infty$, bounding the difference function in $\mathcal{F}_{\delta_n}$ by $\|a\|_{\mathcal{H}_P}$ leads to $|(B(\Lambda)(h) - B(\Lambda_0)(h))(x,y)| \le K\delta_n$ for $n \ge N_0$, where $K = [2(\nu_0+2)(p+1)+C]$ with a constant C such that $o(\|a\|_{\mathcal{H}_p}) \le C\|a\|_{\mathcal{H}_p}$ for $n \ge N_0$.

Since $B(\Lambda)(h)(x,y)$ in (3.1) has a constant envelope $2p + (\nu_0+2)p$, the set $\mathcal{F}_{\delta_n}$ itself has a constant envelope $2p(\nu_0+4)$. Its $L_2(P)$-norm boils down to $\sqrt{2p(\nu_0+4)}$ times its $L_1(P)$-norm: $\|B(\Lambda)(h) - B(\Lambda_0)(h)\|_{L_2(P)} \le \sqrt{2p(\nu_0+4)K\delta_n} \equiv \eta_n$. Thus if $\mathcal{F}'_{\eta_n} = \{B(\Lambda)(h) - B(\Lambda_0)(h) : \; \|B(\Lambda)(h) - B(\Lambda_0)(h)\|_{L_2(P)} \le \eta_n, h \in \mathcal{H}_p\}$, then $\mathcal{F}_{\delta_n} \subset \mathcal{F}'_{\eta_n}$ for all $n \ge N_0$.

For a given sequence $\delta_n \downarrow 0$, take $M = \|\Lambda_0\|_{\mathcal{H}_p} + \delta_1 < \infty$. Then $\mathcal{C}_M$ as defined in (3.2) is a universal Donsker class of functions with a constant envelope $p(\nu_0+4)$. Thus $\mathbb{G}_n$ must be asymptotically uniformly equicontinuous in probability with respect to ρ_P (the centered $L_2(P)$-norm) over $\mathcal{C}_M$. Since $\|P\|_{\mathcal{C}_M} = \sup_{h \in \mathcal{C}_M} |\int h dP| \le p(\nu_0+4)$ is finite, this is equivalent to $\mathbb{G}_n$ being asymptotically uniformly equicontinuous in probability with respect to the $L_2(P)$-norm. Hence $\|\mathbb{G}_n\|_{\mathcal{F}'_{\eta_n}} \to_{P^*} 0$. But $\mathcal{F}_{\delta_n} \subset \mathcal{F}'_{\eta_n}$ for all $n \ge N_0$ and it follows that $\sup_{\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \le \delta_n} \|\mathbb{G}_n(B(\Lambda) - B(\Lambda_0))\|_{\mathcal{H}_p} = \|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \le \|\mathbb{G}_n\|_{\mathcal{F}'_{\eta_n}} \to_{P^*} 0$ as $n \to \infty$, which verifies L.2.

To verify L.3, note that $\{B(\Lambda_0)(h) : \; h \in \mathcal{H}_p\}$ is a subset of $\mathcal{C}_M$ for a sufficiently large M, it is a universal Donsker class of functions. Thus we also have $\mathbb{G}_n B(\Lambda_0) \Longrightarrow \mathbb{Z}_0$ in $l^\infty(\mathcal{H}_p)$.

Verification of L.4 and L.5. Since $B(\Lambda)$ for any fixed $\Lambda \in \Theta$ is a bounded operator, we only need to verify that $P_\Lambda B$ is Fréchet differentiable for a bounded $B : \mathcal{H}_p \mapsto \mathcal{F}$. Let $a = \Lambda_1 - \Lambda$ for two points $\Lambda_1, \Lambda \in \Theta$. Straightforward calculation shows that the difference $P_{\Lambda_1}B(h) - P_\Lambda B(h)$ is a sum of a bounded linear operator of a and a higher order term $o(\|a\|_{\mathcal{H}_p})$. Hence $P_\Lambda B$ is Fréchet differentiable with respect to Λ.

Since we can write the difference $\psi(\Lambda_1, P_\Lambda)(h) - \psi(\Lambda, P_\Lambda)(h)$ as an operator $\dot\psi(\Lambda)(a)(h)$ linear in a and a remainder term $R = R(\Lambda_1, \Lambda, h)$ that is of order $o(\|a\|_{\mathcal{H}_p})$, $\psi(\Lambda_1, P_\Lambda)(h) - \psi(\Lambda, P_\Lambda)(h) = \dot\psi(\Lambda)(a)(h) + R(\Lambda_1, \Lambda, h)$, we can verify that $\psi(\Lambda, P)$ is Fréchet differentiable at Λ. The derivative operator $\dot\psi(\Lambda)(a)(h)$ is then obtained as

$$\dot\psi(\Lambda)(a)(h) = -2\nu_0 \int_{[0,\tau]} \sigma(\Lambda, h)(u) da(u), \tag{3.3}$$

and $\sigma(\Lambda, h)$ is given by

$$\sigma(\Lambda,h)(u) = \Big(\frac{1}{1+\Lambda(u)}\Big)^{\nu_0+1} \cdot h(u) - \frac{\nu_0+1}{\nu_0+3}\Big[\frac{1}{[1+\Lambda(u)]^{\nu_0+2}} \int_{[0,u]} h(v) d\Lambda(v) + \int_{(u,\tau]} \frac{h(v)d\Lambda(v)}{[1+\Lambda(v)]^{\nu_0+2}} + \int_{[0,\tau]} \frac{h(v)d\Lambda(v)}{[1+\Lambda(v)+\Lambda(u)]^{\nu_0+2}}\Big] \tag{3.4}$$

which maps from $(\mathcal{H}_p, \|\cdot\|_v)$ into itself. These conditions verify L.4.

To verify L.5, note that the operator σ is actually a continuously invertible operator J minus a compact operator Q: $\sigma = J - Q$, with $J(h)(u)$ given by the first term in (3.4) and $Q(h)(u)$ given by the second term in (3.4). This structure of σ implies that it is a continuously invertible operator with range $\mathcal{R}(\sigma) = \mathcal{H}_\infty$ for every $\Lambda \in \Theta$, as shown below.

Lemma 3.1. *For every fixed $\Lambda \in \Theta$ the operator $\sigma(\Lambda)$ is one-to-one and maps onto $(\mathcal{H}_\infty, \|\cdot\|_v)$, hence is continuously invertible with the range space $\mathcal{H}_\infty$.*

Proof. First we show that σ is one-to-one. Fix $\Lambda \in \Theta$. Suppose that $\sigma(\Lambda, h)(u) \equiv 0$ for all $u \in [0, \tau]$. Since $1/[1 + \Lambda(u)]^{\nu_0+1} > 0$ for all $u \in [0, \tau]$, it follows that

$$h(u) - \int_{[0,\tau]} K(u, v)h(v)d\Lambda(v) \equiv 0 \tag{3.5}$$

for all $u \in [0, \tau]$, where the continuous kernel function $K(u, v)$ is given by

$$K(u, v) = \frac{\nu_0 + 1}{\nu_0 + 3}\Big[\frac{1_{[0,u]}(v)}{1 + \Lambda(u)} + \frac{[1 + \Lambda(u)]^{\nu_0+1}}{[1 + \Lambda(v)]^{\nu_0+2}}1_{(u,\tau]}(v) + \frac{[1 + \Lambda(u)]^{\nu_0+1}}{[1 + \Lambda(v) + \Lambda(u)]^{\nu_0+2}}\Big].$$

The operator K defined by the the integral $K(h)(u) = \int_{[0,\tau]} K(u, v)h(v)d\Lambda(v)$ maps from $C[0, \tau]$ (equipped with the supremum norm) to itself. Its operator norm can be bounded by a number strictly less than 1: $\|K\| = \sup_{u\in[0,\tau]} \int_{[0,\tau]} |K(u, v)|d\Lambda(v) \le [(\nu_0 + 1)/(\nu_0 + 3)][\Lambda(\tau)/(1 + \Lambda(\tau))] + 2/(\nu_0 + 3) < 1$. Thus $(I - K)$ is actually an invertible operator in $C[0, \tau]$ and hence $(I - K)(h) = 0$ implies $h = 0$.

Next we show that $Q(h)$ is a compact operator in $(\mathcal{H}_\infty, \|\cdot\|_v)$ for every $\Lambda \in \Theta$. Let $\{h_n\}$ be a sequence in $(\mathcal{H}_1, \|\cdot\|_v)$ which is a typical bounded set in $(\mathcal{H}_\infty, \|\cdot\|_v)$. By Helly's Selection Theorem and the Dominated Convergence Theorem, we can show that $\{h_n\}$ actually contains a subsequence $\{h_{n_k}\}$ such that $\{Q(h_{n_k})\}$ is convergent. Hence, the operator Q is a compact operator defined on $\mathcal{H}_\infty$.

Now σ is one-to-one and is a sum of a continuously invertible operator J and a compact operator Q. To show $J - Q$ maps onto $\mathcal{H}_\infty$, we only have to show that $I - J^{-1}Q = I - T$ maps onto $\mathcal{H}_\infty$ where $T = J^{-1}Q$ is again a compact operator.

Suppose that $I - T$ does not map onto $\mathcal{H}_\infty$. Let M_n denote the range space of $(I - T)^n$ for $n \ge 1$, then M_n is a closed subspace of $\mathcal{H}_\infty$ (because T is compact). Furthermore, $M_{n+1} \subset M_n$ and M_{n+1} is a proper subspace of M_n. To see this, suppose $M_{n+1} \equiv M_n$. Then since M_n is defined by $M_n = \{(I - T)^n h : h \in \mathcal{H}_\infty\}$, for every $h \in \mathcal{H}_\infty$ there is a $g \in \mathcal{H}_\infty$ such that $(I - T)^{n+1}g = (I - T)^n h$. Thus $(I - T)^n(h - (I - T)g) = 0$. But since $(I - T)$ is one-to-one, using this fact repeatedly leads to $h - (I - T)g = 0$. This means that $(I - T)$ maps onto $\mathcal{H}_\infty$, because $h \in \mathcal{H}_\infty$ is arbitrary.

Once we have a sequence of closed subspaces M_{n+1} that are proper subspaces of M_n, there is a $g_n \in M_n$ such that $\|g_n\|_v \le 2$, $\|g_n - h\|_v \ge 1$ for all $h \in M_{n+1}$ (see Lemma 4.22 on page 106 of Rudin (1991)). Now define $z = Tg_m + (I-T)g_n$ for $m > n > 2$. Then, since $(I-T)g_n \in M_{n+1}$ and $Tg_m \in M_{n+1}$ (because $T(I-T) = (I-T)T$), we have $z \in M_{n+1}$. Thus $\|Tg_m - Tg_n\|_v = \|z - g_n\|_v \ge 1$. The sequence $\{Tg_n\}$ has therefore no convergent subsequences, although $\{g_n\}$ is bounded. This contradicts the compactness of T. The contradiction shows that $(I-T)$ is onto and so is $(J-Q)$.

Theorem 3.1. *For every $\Lambda \in \Theta$ and every fixed $0 < p < \infty$, the Fréchet derivative $\dot\psi(\Lambda)(a)$ maps $\overline{lin(\Theta)}$ onto $\overline{\mathcal{R}(\dot\psi)} \subset l^\infty(\mathcal{H}_p)$ and is continuously invertible.*

Proof. By Lemma 3.1, the operator $\sigma(\Lambda, h)$ is one-to-one and maps $(\mathcal{H}_\infty, \|\cdot\|_v)$ onto itself. Hence the operator norm of σ^{-1} is bounded: $\|\sigma^{-1}\| < \infty$. Therefore, for any fixed $p > 0$, $\sigma^{-1}(\mathcal{H}_q) \subset \mathcal{H}_p$ with $q = p/\|\sigma^{-1}\| > 0$. This leads to

$$\begin{aligned}\left\|\dot\psi(\Lambda)(a)\right\|_{\mathcal{H}_p} &\ge \sup_{h\in\sigma^{-1}(\mathcal{H}_q)}\left|2\nu_0\int_{[0,\tau]}\sigma(\Lambda,h)(u)da(u)\right| \\ &= 2\nu_0 \sup_{g\in\mathcal{H}_q}\left|\int_{[0,\tau]} gda\right| = 2\nu_0\|a\|_{\mathcal{H}_q}.\end{aligned}$$

Notice that $g = (q/p)h \in \mathcal{H}_q$ for an $h \in \mathcal{H}_p$,

$$\|a\|_{\mathcal{H}_p} = \sup_{h\in\mathcal{H}_p}\left|\int_{[0,\tau]} hda\right| = \frac{p}{q}\sup_{h\in\mathcal{H}_p}\left|\int_{[0,\tau]}\frac{q}{p}hda\right| \le \frac{p}{q}\sup_{g\in\mathcal{H}_q}\left|\int_{[0,\tau]} gda\right| = \frac{p}{q}\,\|a\|_{\mathcal{H}_q}.$$

This implies that $\|\dot\psi(\Lambda)(a)\|_{\mathcal{H}_p} \ge 2\,\nu_0\|a\|_{\mathcal{H}_q} \ge (2\nu_0 q)/p\,\|a\|_{\mathcal{H}_p}$, i.e., $\dot\psi(\Lambda)(a)$ is continuously invertible on its range $\mathcal{R}(\dot\psi)$. The above inequality also implies that $\dot\psi(\Lambda)(a)$ is a continuously invertible operator from $\overline{lin(\Theta)}$ to $\overline{\mathcal{R}(\dot\psi)}$.

The last stage of verifying L.5 requires showing

$$\left\|\dot\psi^{-1}(\Lambda)(f) - \dot\psi^{-1}(\Lambda_0)(f)\right\|_{\mathcal{H}_p} \to 0 \tag{3.6}$$

for any $f \in \overline{\mathcal{R}(\dot\psi)}$ as $\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \to 0$.

Let $a_0 = \dot\psi^{-1}(\Lambda_0)(f)$. Because $\|\dot\psi(\Lambda)\| \le \epsilon$, we can rewrite $\|\dot\psi^{-1}(\Lambda)(f) - \dot\psi^{-1}(\Lambda_0)(f)\|_{\mathcal{H}_p} = \|\dot\psi^{-1}(\Lambda)(\dot\psi(\Lambda_0)(a_0) - \dot\psi(\Lambda)(a_0))\|_{\mathcal{H}_p} \le \epsilon\|\dot\psi(\Lambda_0)(a_0) - \dot\psi(\Lambda)(a_0)\|_{\mathcal{H}_p}$. Thus we only need to show that $\|\dot\psi(\Lambda_0)(a_0) - \dot\psi(\Lambda)(a_0)\|_{\mathcal{H}_p} \to 0$ for any a_0 as $\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \to 0$. Because $|\int f(u)da(u)| \le \sup_u |f(u)| \cdot \|a\|_v$, this amounts to show that $\sup_{h\in\mathcal{H}_p}\sup_{u\in[0,\tau]}|\sigma(\Lambda,h)(u) - \sigma(\Lambda_0,h)(u)| \to 0$ as $\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \to 0$.

In fact, $\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \to 0$ implies $\sup_{u\in[0,\tau]}|\Lambda(u) - \Lambda_0(u)| \to 0$. Hence $\sup_{u\in[0,\tau]}|J(\Lambda)h - J(\Lambda_0)h| \to 0$ because $\|h\|_v < p < \infty$. In a similar way, all three terms in $(Q(\Lambda) - Q(\Lambda_0))(h)$ can be shown to approach zero as $\|\Lambda - \Lambda_0\|_{\mathcal{H}_p} \to 0$.

The overall conclusion thus follows from Theorem 2.1: $\sqrt{n}(\hat{\Lambda}_n - \Lambda_0) \Rightarrow \mathbb{Z}_0$ in $l^\infty(\mathcal{H}_p)$ as $n \to \infty$.

3.2. The multiplicative censoring model

Let $0 < p \leq 1$ be a fixed number. The censoring distribution G is a mixture of a unit mass δ_1 at 1 and a uniform distribution $U(0,1)$: $G = p\delta_1 + (1-p)U(0,1)$. Corresponding to a non-negative random variable $Z \sim F_0$ and a censoring variable $W \sim G$, the observed random variables (X, Δ) are given by a "censored" observation $X = ZW$ and an indicator of whether Z has been censored or not: $\Delta = 1_{[W=1]}$. We are interested in estimating the unknown distribution function F_0 on $[0,\infty)$ on n of i.i.d. observations (X_i, Δ_i) from

$$P_F^{(0)}(x) = P_F\{X \leq x, \Delta = 0\} = q\Big(\int_{[0,x]} dF(u) + \int_{(x,\infty)} xz^{-1} dF(z)\Big), \quad (3.7)$$

$$P_F^{(1)}(x) = P_F\{X \leq x, \Delta = 1\} = pF(x). \quad (3.8)$$

It is worth noting that $dP_F^{(0)}(x) = qf_F(x)dx$ where $f_F(x)$ is a density with respect to Lebesgue measure defined for $x \geq 0$, and given by $f_F(x) = \int_{(x,\infty)} z^{-1} dF(z)$. The value $f_F(0)$ is defined as the limit of $f_F(x)$ as x approaches to 0 from the right.

Let Θ be the set of all distribution functions on the positive real line $[0,\infty)$, equipped with the uniform norm $\|\cdot\|$. For a fixed $F \in \Theta$, the score operator $B(F)$ is obtained by differentiating the log-likelihood along a curve passing through F:

$$B(F)(h_t)(x,\delta) = \delta \cdot 1_{[0,t]}(x) - F(t) + (1-\delta) \cdot \Big[1 - \frac{f_F(t)}{f_F(x)}\Big] \cdot 1_{(0,t]}(x), \quad (3.9)$$

where $h_t = 1_{[0,t]} \in \mathcal{H}$, $\mathcal{H}$ being the set of all indicator functions h_t for $t \geq 0$.

Integrating with respect to P_{F_0} leads to the ψ operator:

$$\psi(F, P_{F_0})(h_t) = P_{F_0}^{(1)}(t) - F(t) + \int_{(0,t]} \Big[1 - \frac{f_F(t)}{f_F(x)}\Big] dP_{F_0}^{(0)}(x). \quad (3.10)$$

The set of all solutions $\hat{F}_n$ to $\psi(\hat{F}_n, \mathbb{P}_n)(h_t) = 0$ for all t are the MLE defined in Vardi and Zhang (1992), page 1025. The set of all Z-estimators $\hat{F}_n$ defined by $\|\psi(\hat{F}_n, \mathbb{P}_n)\|_{\mathcal{H}} = o_{P^*}(n^{-1/2})$ certainly contains this set. The consistency of $\hat{F}_n$ is assumed to be available, see Vardi and Zhang (1992), or Zhan (1996) for a proof.

Verification of L.1, L.2 *and* L.3. Verifying L.1 is trivial, simply substitute P_{F_0} with P_F for any $F \in \Theta$.

To verify L.2 note that, for any t and F, the class of functions defined on the sample space $\{f_{F,t}(x,\delta) = (1-\delta) \cdot [1 - f_F(t)/f_F(x)] \cdot 1_{(0,t]}(x) : F \in \Theta, t \in [0,\infty)\}$

only contains monotone functions in x for $\delta = 0, 1$. They are uniformly bounded by a constant 1. It is a universal Donsker class of functions. Hence the function class

$$\mathcal{F} = \Big\{ B(F)(h_t)(x, \delta) : \; F \in \Theta \text{ and } t \in [0, \infty) \Big\} \tag{3.11}$$

is a universal Donsker class of functions with a constant envelope 1 (see Van der Vaart and Wellner (1996), Section 2.10.2 for details).

Since the difference $|(B(F) - B(F_0))(h_t)(x, \delta)|$ is uniformly bounded by 2, the $L_2(P)$-norm of it boils down to 2 times the square root of its $L_1(P)$-norm. Because $f_F(t)/f_F(x) \le 1$ for $x \le t$,

$$\int_{(0,t]} \Big| \frac{f_F(t)}{f_F(x)} - \frac{f_{F_0}(t)}{f_{F_0}(x)} \Big| dP^{(0)}(x) \le q \int_{(0,\infty)} |f_F(x) - f_{F_0}(x)| dx + 2q \|F_0 - F\| \longrightarrow 0.$$

The fact that the first term above approaches zero as $\|F - F_0\| \to 0$ is verified by integration by parts (Shorack and Wellner (1986), page 868) and Scheffé's Theorem.

Thus $\sup_{t \ge 0} \|(B(F) - B(F_0))(h_t)\|_{L_2(P)} \to 0$ as $\|F - F_0\| \to 0$. Hence $\|\mathbb{G}_n(B(\hat{F}_n) - B(F_0))\|_{\mathcal{H}} = o_{P^*}(1)$ for any consistent $\hat{F}_n \to_{P^*} F_0$ in the uniform norm (see Van der Vaart (1994), Corollary 2.2). By Remark 2.1, we have verified L.2.

Since the function class $\{B(F_0)(h_t) : \; t \in [0, \infty)\}$ is a subset of $\mathcal{F}$, it is universal Donsker. Hence $\mathbb{G}_n B(F_0) \Rightarrow \mathbb{Z}_0$ in $l^\infty(\mathcal{H})$, where $\mathbb{Z}_0$ is a tight Gaussian random element in $l^\infty(\mathcal{H})$. This verifies L.3.

Verification of L.4' and L.5'. The operator $\psi(F, P)$ as a function of F is not Fréchet differentiable with respect to the uniform norm $\|\cdot\|$ on Θ. However, it is Fréchet differentiable with respect to a slightly stronger norm $\|\cdot\|_D$ defined in the following.

Let $D(x)$ be a positive-valued function defined on $[0, \delta)$ for an arbitrary but fixed $\delta > 0$. Assume that $D(x)$ is right-continuous at the origin: $D(0+) = D(0) = 0$. Define a set $\Theta_\alpha \subset \Theta$ for a positive constant $\alpha > 0$:

$$\Theta_{\alpha,\delta} = \Big\{ F \in \Theta : \; \frac{|F(x) - F_0(x)|}{\|F - F_0\|} \le \alpha D(x), \; x \in [0, \delta) \Big\}.$$

The set $\Theta_{\alpha,\delta}$ is mapped into a subset $A_{\alpha,\delta} \subset lin(\Theta)$ by $a(x) = (F(x) - F_0(x))/\|F - F_0\|$ with $A_{\alpha,\delta}$ defined by $A_{\alpha,\delta} = \{a : \; |a(x)| \le \alpha D(x) \text{ for } x \in [0, \delta) \text{ and } \|a\| \le 1\}$.

The norm $\|\cdot\|_D$ is defined by $\|a\|_D = \sup_{x \in [0,\delta), \, D(x) \neq 0} (|a(x)|/D(x) \, \|a\|) + \|a\|$ which is stronger than $\|\cdot\|$ and suffices to guarantee the Fréchet differentiability of the function $\psi(F, P)$ at any $F \in \Theta$.

The Fréchet differentiability of $\psi(F,P)$ boils down to the continuity of the operator $S_F(\cdot)$ indexed by $F \in \Theta$:

$$S_F(a)(h_t) = \int_{(0,t]} \Big(1 - \frac{f_F(t)}{f_F(x)}\Big)\Big(\int_{(x,\infty)} z^{-1} da(z)\Big) dx, \tag{3.12}$$

which maps $lin(\Theta)$ to $l^\infty(\mathcal{H})$. A transformation defined by

$$u(a)(x) = x \int_{(x,\infty)} z^{-2} a(z)\, dz \tag{3.13}$$

helps with the proof of a number of properties of the operator $S_F(\cdot)$. The function u maps an $a \in lin(\Theta)$ to a $u \in C_0[0,\infty)$, where $C_0[0,\infty)$ denotes the set of all continuous functions defined on $[0,\infty)$ that vanish at 0 and ∞: $u(a)(0+) = a(0+) = 0$ and $u(a)(\infty) = \lim_{x\to\infty} a(x) = 0$.

Lemma 3.2. *For any distribution function $F \in \Theta$, let $S_F(\cdot)$ be the operator from $lin(\Theta)$ to $l^\infty(\mathcal{H})$ defined in* (3.12).

(i) *S_F can be written as a linear operator from $(C_0[0,\infty), \|\cdot\|)$ to $l^\infty(\mathcal{H})$:*

$$S_F(a)(h_t) = \int_{(0,t]} u(a)(x) d\Big(\frac{f_F(t)}{f_F(x)}\Big) \equiv \bar{S}_F(u(a)),$$

with $u(a)(x)$ given by (3.13) *and $\|u(a)\| \le \|a\|$. For any fixed $u \in C_0[0,\infty)$, or equivalently for corresponding $a \in lin(\Theta)$, we have $\|\bar{S}_F(u(a)) - \bar{S}_{F_0}(u(a))\|_{\mathcal{H}} = \|S_F(a) - S_{F_0}(a)\|_{\mathcal{H}} \to 0$ as $\|F - F_0\| \to 0$.*

(ii) *For any $\alpha > 0$, the operator S_F as a function of F is also continuous in the sense that $\sup_{a\in A_\alpha} \|S_F(a) - S_{F_0}(a)\|_{\mathcal{H}} \to 0$ as $\|F - F_0\| \to 0$.*

Proof. The proof of (i) follows from Vardi and Zhang (1992) and is omitted here.

For the proof of (ii), let $u(A_{\alpha,\delta}) \subset C_0[0,\infty)$ denote the image of $A_{\alpha,\delta}$ under the mapping u for some fixed $\alpha > 0$. First we show that the set $u(A_{\alpha,\delta})$ is relatively compact in $C_0[0,\infty)$. Since $\|u\| \le \|a\| \le 1$ for $a \in A_{\alpha,\delta}$, the set $u(A_{\alpha,\delta})$ is uniformly bounded. By the Arzelá-Ascoli Theorem (see Kirillov and Gvishiani (1982) on page 180 for a generalization that does not require the domain of the functions to be compact), we only need to show that elements in $u(A_{\alpha,\delta})$ are equicontinuous on $[0,\infty)$.

Notice that $\lim_{x\downarrow 0}\ x\int_{(x,\delta)} z^{-2}D(z)dz = \lim_{x\downarrow 0} D(x) = D(0+) = 0$. For any given $\epsilon > 0$ we can find an $\eta > 0$, $\eta \le (\epsilon\delta)/2$, such that $\alpha x\int_{(x,\delta)} z^{-2}D(z)dz \le \epsilon/4$ for any $0 \le x \le \eta$. Hence we have $|u(a)(x) - u(a)(y)| \le \epsilon/2 + |x-y|/\delta \le \epsilon/2 + \eta/\delta \le \epsilon$ for $x, y \in (0,\eta]$. For $x, y \in (\eta,\infty)$ and $x < y$, we have $|u(a)(x) - u(a)(y)| \le (2|x-y|)/\eta$ by the fact that $\|a\| \le 1$. Thus when $|x-y| \le (\eta\epsilon)/2$, $|u(a)(x) - u(a)(y)| \le \epsilon$. Therefore, $u(A_{\alpha,\delta})$ is relatively compact.

The operators $\{\bar{S}_F : F \in \Theta\}$ are uniformly bounded and defined on a relatively compact set $u(A_{\alpha,\delta})$. They are pointwise continuous by (i), and hence they are uniformly continuous, i.e., we have $\sup_{u\in A_{\alpha,\delta}} \|(\bar{S}_F - \bar{S}_{F_0})(u)\|_{\mathcal{H}} \to 0$ as $\|F - F_0\| \to 0$.

Now we are ready to verify the Fréchet differentiability of $\psi(F,P)$ in F with respect to $\|\cdot\|_D$. In fact, the difference $(\psi(F_1,P) - \psi(F,P))(h_t)$ can be written as the sum of a linear operator $-\dot{\psi}(F)(a)(h_t)$ given by

$$pa(t) + q\int_{(0,t]} \Big(1 - \frac{f_F(t)}{f_F(x)}\Big)\Big(\int_{(x,\infty)} z^{-1}da(z)\Big)dx = (pI + qS_F)(a)(h_t) \quad (3.14)$$

and a remainder term $R(t)=q\int_{(0,t]}[f_{F_1}(t)/f_{F_1}(x)-f_F(t)/f_F(x)]\int_{(x,\infty)} z^{-1}d(F_1(z)-F(z))dx$.

Let $a(x) = (F_1(x) - F(x))/\|F_1 - F\|$, then $R(t) = q\|F - F_0\| \cdot (S_F(a) - S_{F_1}(a))(h_t)$. For an $\alpha > 0$, when $\|F - F_0\|_D \le \alpha$, we certainly have $a \in A_{\alpha,\delta}$. Hence when $\|F - F_0\|_D$ is sufficiently small $\sup_{t\in[0,\infty)} |R(t)| \le q\|F_1 - F\| \cdot \sup_{a\in A_{\alpha,\delta}} \|(S_{F_1}(a) - S_F(a))\|_{\mathcal{H}}$. Since $\|F_1 - F\| \le \|F_1 - F\|_D$, we have $\sup_{t\in[0,\infty)} |R(t)| = o(\|F_1 - F\|_D)$ by Lemma 3.2 (ii).

To verify the invertibility of $\dot{\psi}(F)$ and the pointwise continuity of $\dot{\psi}^{-1}(F)$, let $D_0[0,\infty)$ be the Banach space of all real functions $f(\cdot)$ on $[0,\infty)$ that are right-continuous with left limits, satisfying $f(0) = 0$ and $f(\infty) = 0$, equipped with the uniform norm $\|\cdot\|$. Since the operator $\dot{\psi}(F)$ in (3.14) is the operator R in equation (3.4) in Vardi and Zhang (1992), Lemma 3 in their paper implies that the linear operator $\dot{\psi}(F)$ is a one-to-one mapping from $D_0[0,\infty)$ *onto* $D_0[0,\infty)$ for any $F \in \Theta$. Furthermore, the operator norm $\|\dot{\psi}^{-1}(F)\| \le 2/p^2$. Let $a_0 = \dot{\psi}^{-1}(F_0)(f)$, it is now straightforward to verify $\|\dot{\psi}^{-1}(F)(f) - \dot{\psi}^{-1}(F_0)(f)\| = \|\dot{\psi}^{-1}(F)(\dot{\psi}(F)(a_0) - \dot{\psi}(F_0)(a_0))\| \le \|\dot{\psi}^{-1}(F)\| \cdot \|\dot{\psi}(F)(a_0) - \dot{\psi}(F_0)(a_0)\|_{\mathcal{H}} \le qc_p\|S_F(a_0) - S_{F_0}(a_0)\|_{\mathcal{H}} \to 0$. Thus the assumption L.5' has been verified (the $\|\cdot\|_K$-norm in L.5' is the weaker uniform norm $\|\cdot\|$ here).

Verification of L.6. Since P_F can be identified as $(P_F^{(0)}(x), P_F^{(1)}(x))$ in (3.7) and (3.8), P_F is obviously convex linear in F for any $F \in \Theta$. For a fixed $\bar{F}$, let $f_t(x,\delta) = B(\bar{F})(h_t)(x,\delta)$ denote the image of $h_t = 1_{[0,t]}$ under the score operator $B(\bar{F})$. Then both $f_t(x,0)$ and $f_t(x,1)$ are functions of bounded variation. In addition, they are uniformly bounded by 1, see (3.9).

Now let λ_i be real numbers for $i = 1,\ldots,k$. We can write $dP_F^{(0)}(x) = qf_F(x)dx = q\,du(F)(x)$ with $u(a)(x)$ defined in (3.13). Since $u(F)$ is linear in F and $f_t(x,0)$ is left-continuous, integration by parts shows that $\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(x,0)\, dP_{F_i}^{(0)}(x) = -q\int_{[0,\infty)} u\Big(\sum_{i=1}^k \lambda_i F_i\Big)(x+)df_t(x,0)$, because $f_t(\infty,0) = 0$ and $u(\sum_{i=1}^k \lambda_i F_i)(0+) = 0$.

Noting that $\|u(F)\| \le \|F\|$ and $f_t(x,0)$ is uniformly bounded by 1, we have $\sup_{t\ge 0}|\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(x,0) dP_{F_i}^{(0)}(x)| \le q\|\sum_{i=1}^k \lambda_i F_i\|$, and $\sup_{t\ge 0}|\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(x,1) dP_{F_i}^{(1)}(x)| \le p\|\sum_{i=1}^k \lambda_i F_i\|$. Therefore, by the triangle inequality, we have $\sup_{t\ge 0}|\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(x,\delta)\, dP_{F_i}(x,\delta)| \le 2\|\sum_{i=1}^k \lambda_i F_i\| \le 2\|\sum_{i=1}^k \lambda_i F_i\|_D$.

After verifying all the conditions, it follows from Theorem 2.2 that $\sqrt{n}(\hat{F}_n - F_0) \Rightarrow -\dot{\psi}^{-1}(F_0)(\mathbb{Z}_0)$ for any Z-estimator $\{\hat{F}_n\}$. It is worthwhile to note that the linearity identity in this example reduces to $\dot{\psi}(F)(F - F_0) = -\psi(F,P)$. Now let $\hat{F}_n$ be a Z-estimator to see that $\dot{\psi}(\hat{F}_n)(\sqrt{n}(\hat{F}_n - F_0)) = -\sqrt{n}\psi(\hat{F}_n, P)$. This is equation (2.5) in Vardi and Zhang (1992) (with $m/(n+m)$ in their equation the same as p in our equation).

3.3. The double censoring model

Let $X \sim F_0$ be a non-negative random variable. Let (Y,Z) be a pair of non-negative random censoring times independent of the random variable X that satisfy $P\{Y \le Z\} = 1$. We observe a pair of random variables (W,Δ), where $(W,\Delta) \sim P$ defined by

$$(W,\Delta) = \begin{cases} (X,1) \text{ if } Y < X \le Z, \\ (Z,2) \text{ if } X > Z, \\ (Y,3) \text{ if } X \le Y. \end{cases}$$

We are interested in estimating the distribution function F_0 from i.i.d. pairs $(W_i,\Delta_i) \sim P$, $i = 1,\ldots,n$.

Likelihood equations. Let $G_Y(t) = P\{Y \le t\}$ and $G_Z(t) = P\{Z \le t\}$ be the marginal distribution functions of Y and Z, respectively. Let $K(t) = G_Y(t) - G_Z(t)$. The distribution P_{F_0} is now equivalent to the following three marginals for $\Delta = 1,2,3$,

$$P_F^{(1)}(t) \equiv P_F\{W \le t, \Delta = 1\} = \int_{[0,t]} K(u-)\, dF(u), \tag{3.15}$$

$$P_F^{(2)}(t) \equiv P_F\{W \le t, \Delta = 2\} = \int_{[0,t]} (1 - F(u)) dG_Z(u), \tag{3.16}$$

$$P_F^{(3)}(t) \equiv P_F\{W \le t, \Delta = 3\} = \int_{[0,t]} F(u) dG_Y(u). \tag{3.17}$$

Let $H_P(t) = \sum_{j=1}^3 P_{F_0}^{(j)}(t)$ denote the marginal distribution of W under the true F_0.

Let Θ be the set of all distribution functions on $[0,\infty)$. The score operator $B(F)$ can be obtained by the differentiating the log-likelihood function along a curve indexed by bounded measurable functions $h_t \in \mathcal{H} = \{h_t = 1_{[0,t]}(\cdot) : t \in [0,\infty)\}$:

$$B(F)(h_t)(w,\delta)=\Big(1_{[0,t]}(w)-F(t)\Big)-1_{[\delta=2,w\le t]}\,\frac{1-F(t)}{1-F(w)}+1_{[\delta=3,w>t]}\,\frac{F(t)}{F(w)}. \tag{3.18}$$

The likelihood equations are given by the ψ operator

$$\psi(F,P)(h_t)=H_P(t)-F(t)-\int_{[0,t]}\frac{1-F(t)}{1-F(u)}dP^{(2)}_{F_0}(u)+\int_{(t,\infty)}\frac{F(t)}{F(u)}dP^{(3)}_{F_0}(u) \tag{3.19}$$

for $t \in [0,\infty)$. The set of all Z-estimators $\hat{F}_n$ in this model contains the set of all self-consistent estimators defined by $\psi(\hat{F}_n,\mathbb{P}_n)(h_t) \equiv 0$ for all $t \ge 0$. It is well known that $\hat{F}_n$ is consistent in the uniform norm, see Gu and Zhang (1993), Chang and Yang (1987), and Zhan (1996).

Verification of L.1*,* L.2 *and* L.3*.* The first assumption $\psi(F,P_F)(h_t) \equiv 0$ follows from (3.15), (3.16) (3.17) and integration by parts.

To verify L.2, let $\mathcal{F}$ denote the function class $\mathcal{F} = \{B(F)(h_t)(w,\delta) : F \in \Theta,\, t \in [0,\infty)\}$. The function $B(F)(h_t)(w,\delta)$ is a sum of three functions given in (3.18). The first function is VC-class of functions uniformly bounded by 1, and the second and the third functions are uniformly bounded monotone functions. They are VC-hull classes of functions with a uniform bound 1 and hence they are universal Donsker functions. Since the pointwise sum of a finite number of Donsker classes of functions is Donsker by the permanence of the Donsker property (see Van der Vaart and Wellner (1996), Section 2.10.2 and Example 2.10.7), $\mathcal{F}$ is a universal Donsker class of functions.

Now consider the difference class $\mathcal{F}_{\delta_n} = \{(B(F)-B(F_0))(h_t)(w,\delta) :\ \|F - F_0\| \le \delta_n$ and $t \in [0,\infty)\}$ for a sequence of positive numbers δ_n approaching zero, where $\|\cdot\|$ is the uniform norm defined by $\|a\| = \sup_{t\ge 0}|a(t)|$. To complete the verification of L.2, it suffices to verify

$$\sup_{t\ge 0} P(B(F)(h_t)-B(F_0)(h_t))^2 \to 0 \quad \text{as} \quad \|F-F_0\| \le \delta_n \to 0. \tag{3.20}$$

Since $B(F)(h_t)(w,\delta)$ is bounded by 1 for any F and t, the absolute difference $|(B(F)-B(F_0))(h_t)(w,\delta)|$ is bounded by 2. The $L_2(P)$-norm of the difference function

$$\begin{aligned}
&\|(B(F)-B(F_0))(h_t)\|_{L_2(P)} \le \Big[2\int |(B(F)-B(F_0))(h_t)(w,\delta)|dP(w,\delta)\Big]^{1/2}\\
&\le \Big[2\delta_n\Big(\int dH_P(u)+\int\frac{2}{1-F_0(u)}dP^{(2)}_{F_0}(u)+\int\frac{2}{F_0(u)}dP^{(3)}_{F_0}(u)\Big)\Big]^{1/2}\\
&\le \sqrt{10\delta_n} \equiv \eta_n \to 0.
\end{aligned}$$

This shows that (3.20) holds, and verifies L.2 by an application of Lemma 3.3.5, Van der Vaart and Wellner (1996), page 311.

Now assumption L.3 holds easily from the preceding arguments, since the class of functions $\{B(F_0)(h_t)(w,\delta) : t \in [0,\infty)\}$ is also universal Donsker. Thus $\mathbb{G}_n B(F_0) \Rightarrow \mathbb{Z}_0$ in $l^\infty(\mathcal{H})$, where $\mathbb{Z}_0$ is a tight Gaussian random element in $l^\infty(\mathcal{H})$.

Verification of L.4' and L.5'. In this model, the ψ operator is indeed Fréchet differentiable with respect to the uniform norm $\|\cdot\|$ and its Fréchet derivative operator is given by $-\dot\psi(F)(a)(h_t) = (Ka)(h_t) + A(F,G_Y,G_Z)(a)(h_t)$, where $(Ka)(h_t) = K(t)a(t)$ and

$$A \equiv A(F,G_Y,G_Z)(a)(h_t) = \int_{[0,t]} \frac{1-F(t)}{1-F(u)} a(u)dG_Z(u) + \int_{(t,\infty)} \frac{F(t)}{F(u)} a(u)dG_Y(u).$$

The formal proof proceeds by bounding the remainder term R in the equation $(\psi(F_1,P_F) - \psi(F,P_F))(h_t) = \dot\psi(F)(F_1 - F)(h_t) + R$, and is omitted here.

The operator $\dot\psi(F) = K + A$ is in general not invertible with respect to the uniform norm $\|\cdot\|$ without assuming $\inf_{\tau_0 \le t \le \tau_1} K(t) > 0$. To see this, note that A is a compact operator (see the proof of Lemma 2 in Gu and Zhang (1993)). Without the above condition, the range space of K is not closed, and therefore not invertible with respect to the uniform norm $\|\cdot\|$. Thus $\dot\psi(F) = K + A$ is not invertible with respect to the uniform norm in general.

However, under certain conditions as shown in Gu and Zhang (1993), the operator $\dot\psi(F) = K + A$ is indeed invertible with respect to a weaker $\|\cdot\|_K$-norm defined by $\|a\|_K = \sup_{t\ge 0} |K(t)a(t)|$. The classic argument would still apply if the likelihood equation $\psi(F,P)$ were differentiable with respect to the weaker norm $\|\cdot\|_K$. However, the following example shows that non-differentiability with respect to the K-norm $\|\cdot\|_K$ can occur.

Example 3.1. (Non-Fréchet differentiability of ψ with respect to $\|\cdot\|_K$). Suppose that F is continuous. Let $\tau_1 = \inf\{t : F(t) = 1\} = 1$ be the upper endpoint of the support set of F. Then there exist choices of G_Y, G_Z and F_n in a neighborhood of F for which $\sup_t |R_1(t)|/\|F_n - F\|_K \to C$ as $\|F_n - F\|_K \to 0$, where $C > 0$ is a constant depending on how F_n approaches F.

In fact, let F_n be a sequence of distribution functions defined by

$$F_n(t) = \begin{cases} F(t) & \text{if } t \le \tau_1 - \frac{1}{n}, \\ F(\tau_1 - \frac{1}{n}) & \text{if } t \in (\tau_1 - 1/n, \tau_1), \\ 1 & \text{if } t = \tau_1. \end{cases}$$

Let $G_Y(t) \equiv 1$ be the degenerate distribution at $Y \equiv 0$ and $G_Z(u) = u$ be the uniform $(0,1)$-distribution. Take F to be the uniform $(0,1)$-distribution as well and note that $\tau_1 = 1$, $\sup_t |R_1^n(t)| = \int_{(1-\frac{1}{n},1)}(u-1+1/n)du = 1/2n^2$. On the other hand, the K-norm of $a_n = F_n - F$ is given by $\|a_n\|_K = 1/4n^2$. Hence $\sup_t |R_1^n(t)|/\|a_n\|_K \equiv 2$. The ψ operator is not Fréchet differentiable with respect to the K-norm.

We are now back to verifying L.5'. Let $\tau_0 = \sup\{t : F_0(t) = 0\}$ and $\tau_1 = \inf\{t : F_0(t) = 1\}$. Let $D_0[\tau_0, \tau_1]$ be the Banach space of all real-valued functions defined on $[\tau_0, \tau_1]$ which are right-continuous and have left-limits:

$$D_0[\tau_0,\tau_1] = \{a : F_0(t)=0 \Rightarrow a(t)=0, F_0(t-)=1 \Rightarrow a(t-)=0, F_0(t)=1 \Rightarrow a(t)=0\}.$$

Let $(D_K[\tau_0,\tau_1], \|\cdot\|_K)$ denote the completion of $D_0[\tau_0,\tau_1]$ under the K-norm $\|a\|_K = \|Ka\|$. Further restrict Θ to be all distribution functions on $[0,\infty)$ such that $F \in \Theta$ implies $F - F_0 \in D_0[\tau_0,\tau_1]$.

The operator $\dot\psi(F)(\cdot)$ can be regarded as a mapping from $D_K[\tau_0,\tau_1]$ into $D_0[\tau_0,\tau_1]$. To verify L.5', we need to verify that $\dot\psi(F)$ is invertible and $\dot\psi^{-1}(F)(f)$ converges to $\dot\psi^{-1}(F_0)(f)$ as $\|F - F_0\| \to 0$. But this follows from Lemma 2 in Gu and Zhang's (1993). Briefly, if G_Y, G_Z and F_0 satisfy the following conditions:

$$K(t-) > 0 \text{ on } \{t : \; F_0(t) > 0 \; or \; F_0(t-) < 1\} \tag{3.21}$$

and, for any $0 < \eta < 1$,

$$\int_{0<F_0(u)<1-\eta} \frac{dG_Z(u)}{G_Y(u) - G_Z(u)} + \int_{\eta<F_0(u)<1} \frac{dG_Y(u)}{G_Y(u) - G_Z(u)} < \infty, \tag{3.22}$$

then for any F such that $F - F_0 \in D_0[\tau_0,\tau_1]$, $\dot\psi(F)$ has a bounded inverse on $D_0[\tau_0,\tau_1]$: $\dot\psi^{-1}(F) : D_0[\tau_0,\tau_1] \mapsto D_K[\tau_0,\tau_1]$. Furthermore $\dot\psi^{-1}(F)$ is continuous in F: $\|\dot\psi^{-1}(F)(f) - \dot\psi^{-1}(F_0)(f)\|_K \to 0$. for any $f \in \mathcal{R}(\dot\psi) = D_0[\tau_0,\tau_1]$ and F such that $\|F - F_0\| \to 0$ and $F - F_0 \in D_0[\tau_0,\tau_1]$.

Verification of L.6. Let $\|\cdot\|$ denote the uniform norm over the positive real line $[0,\infty)$. Obviously P_F is convex linear in F by (3.15), (3.16) and (3.17).

Let $f_t(w,\delta) = B(\bar F)(h_t)(w,\delta)$ be the image of h_t under the score operator $B(\bar F)$ for an $\bar F \in \Theta$. Let λ_i be real numbers for $i = 1,\ldots,k$. To verify L.6, we need to establish that

$$\sup_{t\geq 0}\Big|\sum_{i=1}^k \lambda_i \sum_{l=1}^3 \int_{[0,\infty)} f_t(w,l)dP_{F_i}^{(l)}(w)\Big| \leq C \sup_{t\geq 0}\Big|\sum_{i=1}^k \lambda_i F_i(t)\Big|. \tag{3.23}$$

We only have to show by the triangle inequality for some positive numbers C_l, $l = 1,2,3$, that $\sup_{t\geq 0}\Big|\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(w,l)dP_{F_i}^{(l)}(w)\Big| \leq C_l \sup_{t\geq 0}\Big|\sum_{i=1}^k \lambda_i F_i(t)\Big| = C_l\Big\|\sum_{i=1}^k \lambda_i F_i\Big\|$.

For $l = 1$, let $r(w-) = f_t(w,1)K(w-)$. Then $r(w)$ is a function with bounded variation because it is a product of the function $K(w-)$ and the function $f_t(w,1)$, both of which are of bounded variation by (3.18). Integration by parts gives $\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(w,1)\, dP_{F_i}^{(1)}(w) = -\int_{[0,\infty)} \sum_{i=1}^k \lambda_i F_i(w)\, dr(w)$, because $\sum_{i=1}^k \lambda_i F_i(0) = 0$ and $r(\infty) = 0$. Now since $|\int_{[0,\infty)} dr(w)| \leq 1$, we have $\sup_{t\geq 0} |\sum_{i=1}^k \lambda_i \int_{[0,\infty)} f_t(w,1)\, dP_{F_i}^{(l)}(w)| \leq \| \sum_{i=1}^k \lambda_i F_i \|$. The cases $l = 2, 3$ can be verified in a similar way.

Hence the asymptotic normality of $\hat{F}_n$ follows from Theorem 2.2 under conditions in (3.21) and (3.22). It is worthwhile to note that, by Lemma 2.5, we have $\dot{\psi}(\hat{F}_n)(\sqrt{n}(\hat{F}_n - F_0)) = \sqrt{n}\psi(\hat{F}_n, P)$. This is the equation (2.11) in Gu and Zhang (1993) with $\dot{\psi}(F) = K - A_S$ being R_S in (2.9) of their paper with $S = 1 - F$.

Acknowledgements

I am grateful to the two referees and the associate editor for valuable suggestions that significantly improved the quality of the paper.

References

Bickel, P., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for semiparametric Models.* John Hopkins Univ. Press, Baltimore.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18**, 391-404.

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15**, 1536-1547.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press, Princeton.

Daniels, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 151-163. University of California Press, Berkeley.

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators, Part I: General Theory.* Wiley Classic Library Edition, New York.

Fisher, R. A. (1922). The mathematical foundations of theoretical statistics. *Phil. Trans. Ser. A*, **222**, 309-368.

Gill, R. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. of Statist.* **16**, 97-128.

Gu, M. G. and Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611-624.

Heesterman, C. C. and Gill, R. D. (1992). A central limit theorem for M-estimators by the von Mises Method. *Statistica Neerlandica* **46**, 165-177.

Huang, J. (1993). Central limit theorems for *M*-estimates. Tech. Rept., Department of Statistics, University of Washington.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings Fifth Berkeley Symposium Mathematical Statistics and Probability* **1** (Edited by L. M. LeCam and J. Neyman), 221-233. University of California Press, Berkeley.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Statist.* **35**, 73-101.

Kantorovich, L. V. and Akilov, G. P. (1982). *Functional Analysis.* Pergamon Press Ltd, Oxford.

Kirillov, A. A. and Gvishiani, A. D. (1982). *Theorems and Problems in Functional Analysis.* Springer-Verlag, New York.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182-198.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712-731.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Parner, E. (1996). Asymptotic normality in the correlated gamma-frailty model. *Tech. Rept.* **346**, Department of Theoretical Statistics, University of Arhrus. Submitted to *Ann. Statist.*.

Pollard, D. (1989). Asymptotics via empirical processes. *Statist. Sci.* **4**, 341-366.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1**, 295-314.

Rudin, W. (1991). *Functional Analysis.* McGraw-Hill, Inc., New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Application to Statistics.* John Wiley, New York.

Tsai, W. Y. and Zhang, C.-H. (1995). Asymptotic properties of nonparametric maximum likelihood estimator for interval-truncated data. *Scand. J. Statist.* **22**, 361-370.

Van der Laan, M. (1992). *Efficient and Inefficient Estimation in Semiparametric Models.* Ph.D. Dissertation, University of Utrecht.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Van der Vaart, A. W. (1995). Efficiency of infinite dimensional *M*-estimators. *Statistica Neerlandica* **49**, 9-30.

Van der Vaart, A. W. (1994). Maximum likelihood estimation with partially censored data. *Ann. Statist.* **22**, 1896-1916.

Vardi, Y. and Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022-1039.

Vardi, Y. (1989). Multiplicative censoring, renewal process, deconvolution, and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751-761.

Wellner, J. A. and Zhan, Y. (1996). Bootstrapping *Z*-Estimators. Accepted by *Bernoulli*.

Wellner, J. A. and Zhan, Y. (1997). A Hybrid Algorithm for NPMLE from Incomplete Data. *J. Amer. Statist. Assoc.* **92**, 945-959.

Zhan, Y. (1996). *Bootstrapping M-Estimators.* Ph.D. Dissertation. Department of Statistics, University of Washington.

Insightful Corp., Suite 500, 1700 Westlake Ave. N., Seattle, WA 98109.

E-mail: yzhan@insightful.com

(Received March 1999; accepted October 2001)