The Annals of Statistics
1997, Vol. 25, No. 3, 1036–1049

ASYMPTOTIC PROPERTIES OF THE NPMLE OF A DISTRIBUTION FUNCTION BASED ON RANKED SET SAMPLES

By Jian Huang

University of Iowa

We show that the nonparametric maximum likelihood estimator (NPMLE) of a distribution function based on balanced ranked set samples is consistent, converges weakly to a Gaussian process and is asymptotically efficient. The covariance function of the limiting process is described in terms of the solution to a Fredholm integral equation of the second kind.

1. Introduction. A balanced ranked set sample (RSS) consists of independent observations $\{X_{rj}, r = 1, \dots, k; j = 1, \dots, m\}$, where, for every $1 \le r \le k$, $\{X_{rj}\colon j = 1, \dots, m\}$ is an i.i.d. subsample distributed as the rth order statistic of k independent random variables with a common distribution F_0. The total sample size is $n = mk$.

McIntyre (1952) first introduced a ranked set sampling procedure to estimate the mean of pasture yields. Measuring yields of pasture plots is a costly and time-consuming process, but ranking the yields visually can be done easily and accurately without measurement. In such applications, we can first select and rank k plots, measure the smallest yield, where k is usually a small number such as 2 or 3 so that ranking can be done visually without error. Next, select and rank another k plots, measure the second smallest yield and so on. Repeat this process m times; the resulting measurements constitute a balanced RSS. The advantage of such a sampling procedure is that its sample mean has smaller variance than the sample mean of a simple random sample of the same size in estimation of a population mean. This was proved by Takahasi and Wakimoto (1968). Further applications of RSS procedure in agriculture can be found in, for example, Cobby, Ridout, Bassett and Large (1985). Intuitively, a RSS consists of independent order satistics; it is more efficient than the simple random sampling procedure since the order statistics of a simple random sample are correlated. Kvam and Samaniego (1994) mentioned that RSS's also arise in reliability applications.

In the balanced RSS, the empirical distribution function

$$F_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \le x\}}$$

is a consistent estimator for F_0 and is asymptotically normal as m, and hence n, goes to infinity [see, e.g., Stokes and Sager (1988)]. However, as

Received January 1995; revised November 1995.

AMS 1991 *subject classifications*. 60G05, 62G20.

Key words and phrases. Asymptotic normality, consistency, efficiency, Fredholm integral equation, nonparametric maximum likelihood estimation, ranked set sample.

pointed out by Kvam and Samaniego (1994), F_n is not making full use of the information provided by the data and hence is not an efficient estimator of F_0. Based on these considerations, Kvam and Samaniego (1994) proposed the nonparametric maximum likelihood estimator (NPMLE) $\widehat{F}_n$ as an alternative estimator. They demonstrated via simulation that the NPMLE $\widehat{F}_n$ outperforms F_n in terms of mean squared error and Kolmogorov–Smirnov distance in a number of cases. Using the EM algorithm, they also derived self-consistency equations for $\widehat{F}_n$ and showed that, if starting from a consistent initial estimator, such as F_n in balanced RSS, each iteration in the EM algorithm yields a consistent estimator. They conjectured that the NPMLE itself is consistent.

In this paper, we confirm their conjecture for the balanced RSS. Furthermore, we prove that the NPMLE $\widehat{F}_n$ converges weakly to a Gaussian process and is asymptotically efficient with the assumption that F_0 is continuous. We also provide an expression for the covariance function of the limiting Gaussian process in terms of the solution to a Fredholm integral equation with a symmetric and positive definite kernel. With slightly stronger conditions that F_0 is continuous and strictly increasing, the integral equation can be simplified to an equation with a known symmetric and positive definite kernel. We point out that the asymptotic results are for $m \to \infty$ and fixed k. Since $k \geq 1$, we will simply write $n \to \infty$ without causing confusion.

2. Main results. Let F_{rk} be the distribution function of the rth order statistic from an i.i.d. sample of size k. It can be expressed in terms of the underlying distribution function F_0 as

$$dF_{rk}(x) = r\binom{k}{r} F_0^{r-1}(x)(1 - F_0(x))^{k-r}\, dF_0(x).$$

The likelihood function of a balanced ranked set sample $X_1, \dots, X_n$ is proportional to

$$\prod_{i=1}^{n} F^{r_i - 1}(X_i)(1 - F(X_i))^{k-r_i}\, dF(X_i).$$

So the log-likelihood function is, up to an additive constant,

$$l_n(F) = \sum_{i=1}^{n}\{(r_i - 1)\log F(X_i) + (k - r_i)\log(1 - F(X_i)) + \log dF(X_i)\}, \tag{1}$$

where $dF(x) = F(x) - F(x-)$ is the mass that F puts at x. The NPMLE is the $\widehat{F}_n$ that maximizes $l_n(F)$ in the class of distribution functions. Kvam and Samaniego (1994) showed that this optimization problem is well defined and has a unique solution for every fixed sample size.

Let $\mathscr{H}$ be a class of uniformly bounded functions. For s close to zero, define a one-dimensional curve through $\widehat{F}_n$ by

$$d\widehat{F}_s(x) = \left(1 + s\left(h(x) - \int h\, d\widehat{F}_n\right)\right) d\widehat{F}_n(x), \qquad h \in \mathscr{H}.$$

It is clear that, for s close to zero, $\widehat{F}_s$ is a distribution function with $\widehat{F}_0 = \widehat{F}_n$, because its derivative with respect to $d\widehat{F}_n$ is nonnegative and integrates to 1. Since $\widehat{F}_n$ maximizes $l_n(F)$, it must satisfy

$$\frac{\partial}{\partial s} l_n(\widehat{F}_s)\bigg|_{s=0} = 0. \tag{2}$$

Let $h(x) = 1_{\{x \le t\}}$, the above equation can be written as

$$\begin{aligned} \widehat{F}_n(t) &= \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i \le t\}} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{r_i - 1}{\widehat{F}_n(X_i)} - \frac{k - r_i}{1 - \widehat{F}_n(X_i)}\right](\widehat{F}_n(X_i \wedge t) - \widehat{F}_n(X_i)\widehat{F}_n(t))\right\}. \end{aligned} \tag{3}$$

It can be verified that (3) agrees with the self-consistency equation (15) of Kvam and Samaniego (1994) derived via the EM algorithm, provided that both $F^{(1)}$ and $F^{(0)}$ are replaced by $\widehat{F}_n$. However, notice that the term in the first line of their equation (15) is misprinted as $Z_i + (k - Z_i)a(W_i, t; F^{(0)})I(W_i \le t)$; it should be $[Z_i + (k - Z_i)a(W_i, t; F^{(0)})]I(W_i \le t)$.

THEOREM 2.1 (Strong consistency).

$$\sup_{-\infty < t < \infty} |\widehat{F}_n(t) - F_0(t)| \to 0 \quad \textit{a.s. as } n \to \infty.$$

Assume our sample consists of nonnegative random variables, that is, $\sup\{t\colon F_0(t) = 0\} = 0$. Let $\tau = \inf\{t\colon F_0(t) = 1\}$; τ can be ∞. This restriction is for definiteness, the results below hold for other types of supports. Only obvious minor modifications are needed in the proofs.

Let $D[0, \tau]$ be the space of functions on $[0, \tau]$ that are right continuous and have left limits, endowed with the supremum norm $\|f\| = \sup_{0 \le x \le \tau} |f(x)|$ for any $f \in D[0, \tau]$. The convergence in distribution below is according to Hoffmann-Jørgensen (1984); see, for example, van der Vaart and Wellner (1996) for a description.

THEOREM 2.2 (Convergence in distribution and efficiency). *Suppose $F_0(x)$ is continuous. Then the following hold*:

(i)

$$\sqrt{n}(\widehat{F}_n - F_0) \Rightarrow_D Z \quad \textit{as } n \to \infty, \tag{4}$$

where Z is a Gaussian process in $D[0, \tau]$ with mean zero and covariance function

$$\begin{aligned} \mathrm{Cov}(Z(s), Z(t)) &= \int \nu(x, t)\nu(x, s)\, dF_0(x) \\ &\quad + (k-1)\int \frac{\int_0^x \nu(u, t)\, dF_0(u) \int_0^x \nu(u, s)\, dF_0(u)}{F_0(x)(1 - F_0(x))}\, dF_0(x), \end{aligned}$$

where, for any $0 \le x \le \tau$, $\nu(x, t)$ *as a function of* t *is the unique solution to the integral equation*

$$h(t) + (k-1) \int \frac{F_0(s \wedge t) - F_0(s)F_0(t)}{F_0(s)(1 - F_0(s))} h(s)\, dF_0(s) = 1_{\{x \le t\}} - F_0(t); \tag{5}$$

(ii) $\widehat{F}_n$ *is regular and asymptotically efficient.*

REMARK 2.1. When $k = 1$, $\nu(x, t) = 1_{\{x \le t\}} - F_0(t)$, the theorem reduces to the familiar results on the empirical distribution functions of simple random samples.

REMARK 2.2. Notice that the solution $h(t)$ of (5) satisfies the boundary conditions $h(0) = h(\tau) = 0$. Let $g(t) = h(t)/\sqrt{F_0(t)(1 - F_0(t))}$. Equation (5) can be rewritten as

$$g(t) + (k-1) \int k^*_{F_0}(s, t) g(s)\, dF_0(s) = \frac{1_{\{x \le t\}} - F_0(t)}{\sqrt{F_0(t)(1 - F_0(t))}},$$

where

$$k^*_{F_0}(s, t) = \frac{F_0(s \wedge t) - F_0(s)F_0(t)}{\sqrt{F_0(s)(1 - F_0(s))F_0(t)(1 - F_0(t))}}$$

is a symmetric and positive definite kernel. Suppose F_0 is continuous and strictly increasing. Denote its inverse by F_0^{-1}. Let $g_*(t) = g(F_0^{-1}(t))$. The above equation can be further reduced to

$$g_*(t) + (k-1) \int_0^1 \frac{s \wedge t - st}{\sqrt{s(1-s)t(1-t)}} g_*(s)\, ds = \frac{1_{\{F_0(x) \le t\}} - t}{\sqrt{t(1-t)}}.$$

This is an integral equation with a known symmetric and positive definite kernel

$$k_*(s, t) = \frac{s \wedge t - st}{\sqrt{s(1-s)t(1-t)}}.$$

REMARK 2.3. The definition of an efficient regular estimator sequence in a Banach space can be found in Bickel, Klaassen, Ritov and Wellner [(1993), pages 180–182].

3. Proofs.

PROOF OF THEOREM 2.1. By Helly's selection theorem, there exists a non-decreasing function F_* such that any subsequence of $\widehat{F}_n(t)$ has a further subsequence converging to $F_*(t)$ for every t [Chung (1974), Excercise 5, page 86]. If we can show that $F_*(t) = F_0(t)$, then the whole sequence converges to F_0 pointwise. Without loss of generality, we assume that $\widehat{F}_n(t)$ converges to $F_*(t)$

for every t. Write the right-hand side of (3) as the sum of three terms $II_{1n}(t)$, $II_{2n}(t)$ and $-II_{3n}(t)$, where $II_{1n}(t) = (1/n)\sum_{i=1}^{n} 1_{\{X_i \le t\}}$,

$$II_{2n}(t) = k^{-1}\sum_{r=1}^{k} m^{-1}\sum_{j=1}^{m}\left\{(r-1)\frac{\widehat{F}_n(X_{rj}\wedge t) - \widehat{F}_n(X_{rj})\widehat{F}_n(t)}{\widehat{F}_n(X_{rj})}\right\}$$

and

$$II_{3n}(t) = k^{-1}\sum_{r=1}^{k} m^{-1}\sum_{j=1}^{m}\left\{(k-r)\frac{\widehat{F}_n(X_{rj}\wedge t) - \widehat{F}_n(X_{rj})\widehat{F}_n(t)}{1-\widehat{F}_n(X_{rj})}\right\}.$$

The first term II_{1n} is simply the empirical distribution function of a balanced ranked set sample; hence it converges to F_0. Since

$$\frac{\widehat{F}_n(x\wedge t) - \widehat{F}_n(x)\widehat{F}_n(t)}{\widehat{F}_n(x)} = 1_{[x\le t]} - \widehat{F}_n(t) + \frac{\widehat{F}_n(t)}{\widehat{F}_n(x)}1_{[x>t]},$$

and $(\widehat{F}_n(t)/\widehat{F}_n(x))1_{[x>t]}$ (as a function of x) has total variation bounded by 2, by uniform convergence of empirical distribution functions, and $\widehat{F}_n(t) \to F_*(t)$ for every t, it follows that $II_{2n}(t)$ converges to

$$\begin{aligned} &k^{-1}\sum_{r=1}^{k}\int \frac{F_*(x\wedge t) - F_*(x)F_*(t)}{F_*(x)}(r-1)\,dF_{r,k}(x)\\ &\quad = k^{-1}\int \frac{F_*(x\wedge t) - F_*(x)F_*(t)}{F_*(x)}\sum_{r=1}^{k}(r-1)\,dF_{r,k}(x) \qquad (6)\\ &\quad = (k-1)\int \frac{F_*(x\wedge t) - F_*(x)F_*(t)}{F_*(x)}F_0(x)\,dF_0(x), \end{aligned}$$

where the last equation follows from

$$\sum_{r=1}^{k} dF_{r,k}(x) = k\,dF_0(x) \quad\text{and}\quad \sum_{r=1}^{k} r\,dF_{r,k}(x) = k(1+(k-1)F_0(x))\,dF_0(x).$$

These two identities are equations (A.10) and (A.11) of Kvam and Samaniego (1994). They follow from the binomial expansion. Similarly, the third term $II_{3n}(t)$ converges to

$$(k-1)\int \frac{F_*(x\wedge t) - F_*(x)F_*(t)}{1-F_*(x)}(1-F_0(x))\,dF_0(x). \qquad (7)$$

Combine equations (6) and (7) and $II_{1n}(t)\to F_0(t)$ a.s., to get

$$F_*(t) = F_0(t) + (k-1)\int k_{F_*}(x,t)(F_0(x) - F_*(x))\,dF_0(x), \qquad (8)$$

where

$$k_F(x,t) = \frac{F(x\wedge t) - F(x)F(t)}{F(x)(1-F(x))}$$

for any distribution function F. It is proved in the Appendix (Lemma A.1) that F_* satisfies equation (8) if and only if $F_*(t) = F_0(t)$ for all t. So, for every t, $\widehat{F}_n(t) \to F_0(t)$ almost surely. By the lemma following Theorem 5.5 of Chung [(1974), page 133], to prove uniform convergence, it suffices to show that $\widehat{F}_n\{t\} \equiv \widehat{F}_n(t) - \widehat{F}_n(t-) \to F_0(t) - F_0(t-) \equiv F_0\{t\}$. Take $h(x) = 1_{\{x=t\}}$ in (2),

$$\widehat{F}_n\{t\} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i=t\}} + \widehat{F}_n\{t\}B_n(t),$$

where

$$B_n(t) = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\frac{r_i - 1}{\widehat{F}_n(X_i)} - \frac{k - r_i}{1 - \widehat{F}_n(X_i)}\right]\left[1_{\{X_i \le t\}} - \widehat{F}_n(X_i)\right]\right\}.$$

Since $\widehat{F}_n(t)$ converges to $F_0(t)$ almost surely for every t, it can be shown exactly the same way as the verification of (8) that $B_n(t) \to 0$ a.s., for every t. Finally, since the empirical distribution of a balanced ranked set sample converges to F_0 uniformly with probability 1, $(1/n)\sum_{i=1}^{n} 1_{\{X_i=t\}} \to F_0\{t\}$ a.s. It follows that $\widehat{F}_n\{t\} \to F_0\{t\}$ a.s. This completes the proof. □

We now prepare for the proof of Theorem 2.2. Let

$$\psi_F(r, x; t) = 1_{\{x \le t\}} - F_0(t) + \left[\frac{r-1}{F(x)} - \frac{k-r}{1-F(x)}\right](F(x \wedge t) - F(x)F(t)).$$

By (3), $\widehat{F}_n$ satisfies the score equation $S_n(\widehat{F}_n) = 0$, where for any distribution function F we define

$$S_n(F) \equiv F(t) - F_0(t) - \frac{1}{n}\sum_{i=1}^{n} \psi_F(r_i, X_i; t). \tag{9}$$

As in the verification of (8), it can be shown that the limiting version of $S_n(F)$,

$$\begin{aligned} S(F) &\equiv E[S_n(F)] \\ &= F(t) - F_0(t) - (k-1)\int k_F(x, t)(F_0(x) - F(x))\, dF_0(x). \end{aligned} \tag{10}$$

Following the general theorem of van der Vaart (1995) on asymptotics of infinite-dimensional M-estimators, suppose we can prove that the following hold:

(a) $\sqrt{n}(S_n - S)(F_0) \Rightarrow_D Z_0$, where Z_0 is a tight random map in $D[0, \tau]$;
(b) $\|\sqrt{n}(S_n - S)(\widehat{F}_n) - \sqrt{n}(S_n - S)(F_0)\| = o_p(1)(1 + \|\widehat{F}_n - F_0\|)$;
(c) there exists a continuously invertible linear map $\dot{S}_0$ such that

$$\|S(F) - S(F_0) - \dot{S}_0(F - F_0)\| = o(1)(\|F - F_0\|) \quad \text{as } \|F - F_0\| \to 0.$$

Then

$$\sqrt{n}(\widehat{F}_n - F_0) = \dot{S}_0^{-1}\sqrt{n}(S_n - S)(F_0) + o_p(1) \Rightarrow_D -\dot{S}_0^{-1}Z_0.$$

This general approach has been used by several authors to establish convergence in distribution in several difficult estimation problems. See, for example, Murphy (1995) and van der Vaart (1994) for asymptotics of the maximum likelihood estimators of the frailty model and partially censored data model, respectively.

To prove part (i) of Theorem 2.2, it suffices to prove (a)–(c) and then identify the covariance function of the limiting process. We now prove (b) and (c) in the following two lemmas. In the remainder of this section, it is assumed that F_0 is continuous.

LEMMA 3.1. *For S_n and S defined above,*

$$\|\sqrt{n}(S_n - S)(\widehat{F}_n) - \sqrt{n}(S_n - S)(F_0)\| = o_p(1).$$

Notice that this implies (b).

PROOF. Let P_{rm} be the empirical measure of random variables $X_{r1}, \dots, X_{rm}$ that are i.i.d. F_{rk}, and let P_r be the probability measure induced by F_{rk}, $r = 1, \dots, k$. Then we can write

$$(S_n - S)(\widehat{F}_n) - (S_n - S)(F_0) = \frac{1}{k}\sum_{r=1}^{k}(P_{rm} - P_r)(\psi_{\widehat{F}_n}(r, x; t) - \psi_{F_0}(r, x; t)).$$

It suffices to show that

$$\sup_{0\le t\le\tau} |(P_{rm} - P_r)(\psi_{\widehat{F}_n}(r, x; t) - \psi_{F_0}(r, x; t))| = o_p(n^{-1/2})$$

for $r = 1, \dots, k$. After some algebraic manipulations, we can write

$$\begin{aligned}
&(P_{rm} - P_r)(\psi_{\widehat{F}_n}(r, x; t) - \psi_{F_0}(r, x; t))\\
&\quad = (r-1)(P_{rm} - P_r)\left[\frac{\widehat{F}_n(t)}{\widehat{F}_n(x)} - \frac{F_0(t)}{F_0(x)}1_{\{x>t\}}\right]\\
&\qquad + (k-r)(P_{rm} - P_r)\left[\left(\frac{1-\widehat{F}_n(t))}{1-\widehat{F}_n(x)} - \frac{1-F_0(t)}{1-F_0(x)}\right)1_{\{x\le t\}}\right]\\
&\quad \equiv (r-1)A_n(t) + (k-r)B_n(t).
\end{aligned}$$

Since $A_n(t)$ and $B_n(t)$ can be dealt with similarly, we will only prove that $\sup_{0\le t\le\tau}|A_n(t)| = o_p(n^{-1/2})$. For a sequence of numbers $\varepsilon_n \searrow 0$ sufficiently slowly, define the class of functions

$$\mathscr{H}_n = \left\{h(x)\colon h(x) = \left[\frac{F(t)}{F(x)} - \frac{F_0(t)}{F_0(x)}\right]1_{\{x>t\}},\ \|F - F_0\| \le \varepsilon_n,\ t\in[0,\tau]\right\}.$$

Since every function of $\mathscr{H}_n$ has total variation bounded above by 4, $\mathscr{H}_n$ is a subclass of the class of uniformly bounded variation functions. It follows that for any probability measure Q, the $L_2(Q)$ δ-covering entropy for $\mathscr{H}_n$ is of order $1/\delta$ [see, e.g., van de Geer (1993)]. This implies $\mathscr{H}_n$ is manageable in the sense of Pollard (1989). Since $\|\widehat{F}_n - F_0\| \to 0$, a.s., $\widehat{F}_n \in \mathscr{H}_n$ for n sufficiently large. Furthermore, it can be verified that

$$\sup_{\mathscr{H}_n} P_r|h| \to 0 \quad \text{as } n \to \infty.$$

This can be done by a truncation argument: consider the integral separately on the intervals $[t, t+\eta]$ and $[t+\eta, \tau]$ for a small $\eta > 0$. It follows from Theorem 4.4 of Pollard (1989) that

$$\sup_{\mathscr{H}_n} |(P_{rm} - P_r)h| = o_p(n^{-1/2}).$$

This implies the assertion of the lemma. □

Recall $k_{F_0}(x, t)$ is defined in (8). Define integral operator K on $D[0, \tau]$ by $Kf = \int k_{F_0}(x, t) f(x)\, dF_0(x)$. Let $\dot{S}_0 = I + (k-1)K$, where I is the identity operator. Notice that $\dot{S}_0$: $D[0, \tau] \to D[0, \tau]$ is a Fredholm integral operator of the second kind. The Fredholm integral equation has also been used by Chang (1990) in showing weak convergence of a self-consistent estimator of a survival function with doubly censored data.

LEMMA 3.2.

$$\|S(F) - S(F_0) - \dot{S}_0(F - F_0)\| = o(\|F - F_0\|) \quad \textit{as } \|F - F_0\| \to 0.$$

Furthermore, $\dot{S}_0$ *is continuously invertible.*

PROOF. The proof of the differentiability of $S(F)$ can be based on the fact that the kernel k_F is bounded and the assumption that F_0 is continuous. It is omitted. We prove that $\dot{S}_0$ is continuously invertible. First, similarly to the proof of Lemma A.1, it can be shown that

$$\dot{S}_0 h(t) \equiv h(t) + (k-1)Kh(t) = 0$$

if and only if $h(t) = 0$ for every t. So by the Fredholm theory [see, e.g., Kress (1989), Theorem 3.4], it suffices to show that K is compact. Write

$$\begin{aligned} Kh(t) &= \int_{x>t} \frac{F_0(t)}{F_0(x)} h(x)\, dF_0(x) + \int_{x \le t} \frac{1 - F_0(t)}{1 - F_0(x)} h(x)\, dF_0(x) \\ &\equiv K_1 h(t) + K_2 h(t). \end{aligned}$$

We first show that K_1 is compact. For any bounded $h \in D[0, \tau]$: $\|h\| \le C$ and s, t,

$$\begin{aligned} |K_1 h(t) - K_1 h(s)| &\le \int \frac{|F_0(t) 1_{[x>t]} - F_0(s) 1_{[x>s]}|}{F_0(x)} |h(x)|\, dF_0(x) \\ &\le 2C \int \frac{|\sqrt{F_0(t)} 1_{[x>t]} - \sqrt{F_0(s)} 1_{[x>s]}|}{\sqrt{F_0(x)}}\, dF_0(x) \\ &\le 4C\left(\left| \sqrt{F_0(t)} - \sqrt{F_0(s)} \right| + |F_0(t) - F_0(s)| \right). \end{aligned}$$

Hence $K_1 h(t)$ is equicontinuous by (uniform) continuity of F_0. So K_1 is compact by the Arzelà-Ascoli theorem. Similarly, K_2 is compact. It follows that K is compact. □

For the proof of regularity and efficiency of $\widehat{F}_n$, we will apply the general theorems of Bickel, Klaassen, Ritov and Wellner [(1993), Chapter 3] or van der Vaart (1995). We need to define the tangent space. This in turn requires the definition of the score operator of the model considered. To identify the covariance function of the limiting Gaussian process of $\widehat{F}_n$, we also need the second derivatives of the likelihood function.

Denote $\mathbf{X} = (X_{1k}, \dots, X_{kk})$, where $X_{1k}, \dots, X_{kk}$ are independent and X_{rk} is distributed as the rth order statistic from k independent random variables with common distribution F_0. A ranked set sample of size $n = mk$ is the same as the collection of m random vectors $\mathbf{X}_1, \dots, \mathbf{X}_m$ that are independent and identically distributed as $\mathbf{X}$. The joint density of $\mathbf{X}$ is

$$dP_{F_0}(\mathbf{x}) = C \prod_{r=1}^{k} F_0(x_{rk})^{r-1}(1 - F_0(x_{rk}))^{k-r}\, dF_0(x_{rk}),$$

where C is the normalizing constant. For real numbers s and t close to zero and $a, b \in BL_2^0(F_0) \equiv \{h\colon \int h\, dF_0 = 0 \text{ and } h \in L_2(F_0), \text{ and } |h| \le M\}$, where $M < \infty$, define a two-dimensional curve $\{F_{(s,t)}\}$ through F_0 by $dF_{s,t}(x) = (1 + sa(x) + tb(x))\, dF_0(x)$. Denote the score operator by

$$\begin{aligned}(\dot{l}a)(\mathbf{x}) &= \frac{\partial}{\partial s} \log dP_{F_{(s,t)}}(\mathbf{x})\Big|_{s=0,\, t=0} \\ &= \sum_{r=1}^{k} \left\{ a(x_{rk}) + \left[\frac{r-1}{F_0(x_{rk})} - \frac{k-r}{1 - F_0(x_{rk})} \right] \int_0^{x_{rk}} a(u)\, dF_0(u) \right\}.\end{aligned}$$

Denote the second derivative of the log-likelihood function by

$$\begin{aligned}\ddot{l}[a, b](\mathbf{x}) &= \frac{\partial^2}{\partial s \partial t} \log dP_{F_{(s,t)}}(\mathbf{x})\Big|_{(s=0,\, t=0)} \\ &= -\sum_{r=1}^{k} \left\{ a(x_{rk})b(x_{rk}) + \left[\frac{r-1}{F_0^2(x_{rk})} - \frac{k-r}{(1 - F_0(x_{rk}))^2} \right] \right. \\ &\qquad \left. \times \left[\int_0^{x_{rk}} a(u)\, dF_0(u) \right] \left[\int_0^{x_{rk}} b(u)\, dF_0(u) \right] \right\}.\end{aligned}$$

Let $L_2^0(F_0) = \{h\colon \int h dF_0 = 0, h \in L_2(F_0)\}$. Since $BL_2^0(F_0)$ is a dense subset of $L_2^0(F_0)$, we can extend the domain of $\dot{l}$ and $\ddot{l}$ to any $a, b \in L_2^0(F_0)$. So for any $a, b \in L_2^0(F_0)$, $\dot{l}a$ and $\ddot{l}[a, b]$ are well defined. The closure of the linear span of $\{\dot{l}a\colon a \in L_2^0(F_0)\}$ is called the tangent space.

It can be verified that

$$\begin{aligned}E[\ddot{l}[a, b](\mathbf{X})] &= -k \int a(x)b(x)\, dF_0(x) \\ &\quad - k(k-1) \int \frac{\int_0^x a(u)\, dF_0(u) \int_0^x b(u)\, dF_0(u)}{F_0(x)(1 - F_0(x))}\, dF_0(x).\end{aligned} \tag{11}$$

Since $a,\ b \in L_2^0(F_0)$, by the Cauchy–Schwarz inequality, it can be shown that the second integral on the right-hand side of (11) is bounded. This also implies that $E[(\dot{l}a)(\mathbf{X})]^2 < \infty$. Actually, since a ranked set sample can be considered as incomplete observation of all the underlying random variables [see Kvam and Samaniego (1994)] for a detailed description], $\dot{l}a$ can be expressed as a conditional expectation of a given the observed data [see, e.g., Bickel, Klassen, Ritov and Wellner (1993), pages 271–272]. This implies that $E[(\dot{l}a)(\mathbf{X})]^2 \le \int a^2\,dF_0 < \infty$, since $a \in L_2^0(F_0)$. Furthermore, as in the finite dimensional parametric model case, we have

$$E[(\dot{l}a)(\mathbf{X})(\dot{l}b)(\mathbf{X})] = -E[\ddot{l}[a,b](\mathbf{X})]. \tag{12}$$

PROOF OF THEOREM 2.2(i). Let $a_t^*(x) = 1_{\{x\le t\}} - F_0(t)$. Since $(\dot{l}a_t^*)(\mathbf{x}) = \sum_{r=1}^k \psi_{F_0}(r, x_{rk}; t)$ and $S(F_0) \equiv 0$, we have

$$\sqrt{n}[S_n(F_0) - S(F_0)](t) = \sqrt{n}S_n(F_0)(t) = \sqrt{\frac{m}{k}}\mathbf{P}_m \dot{l}a_t^*,$$

where $\mathbf{P}_m$ is the empirical measure of $\mathbf{X}_1, \dots, \mathbf{X}_m$. It can be verified that functions $\dot{l}a_t^*$, $t \in [0,\tau]$ are in the uniformly bounded variation class, so $\sqrt{n}S_n(F_0)(t)$, which are empirical processes indexed by $\dot{l}a_t^*$, $t \in [0,\tau]$, converge in distribution to a Gaussian process in $D[0,\tau]$. Since S_0^{-1} is a continuous linear operator, by the continuous mapping theorem [see, e.g., Pollard (1984), Theorem 12, page 70, or Wellner (1989)], $\sqrt{n}S_0^{-1}S_n(F_0)$ converges weakly to a stochastic process Z in $D[0,\tau]$. That Z is also a Gaussian process follows from the fact that $\dot{S}_0^{-1}$ is a linear operator. In view of Lemmas 3.1 and 3.2 and the general theorem of van der Vaart (1995) described earlier, convergence in distribution of $\sqrt{n}(\widehat{F}_n - F_0)$ in $D[0,\tau]$ is proved.

We now identify the covariance function of Z. By Fubini's theorem,

$$S_0^{-1}\int_0^{\cdot} a_t^*(u)\,dF_0(u) = \int_0^{\cdot} S_0^{-1}a_t^*(u)\,dF_0(u),$$

we have

$$\sqrt{n}S_0^{-1}S_n(F_0)(t) = \sqrt{\frac{m}{k}}\mathbf{P}_m \dot{l}\nu(\cdot, t),$$

where $\nu(\cdot,t) = S_0^{-1}a_t^*$, that is, $\nu(x,t)$ satisfies integral equation (5). Since $|a_t^*(x)| \le 2$ and $\int a_t^*(x)dF_0(x) = 0$, by (5) and continuous invertibility of S_0,

$$\int \nu^2(x,t)\,dF_0(x) < \infty \quad \text{and} \quad \int \nu(x,t)\,dF_0(x) = 0. \tag{13}$$

It remains to prove that for $\mathbf{X}$ with density $dP_{F_0}(\mathbf{x})$ and for any $a(x), b(x) \in L_2(F_0)$,

$$\begin{aligned} E[(\dot{l}a)(\mathbf{X})(\dot{l}b)(\mathbf{X})] &= k\int a(x)b(x)\,dF_0(x) \\ &\quad + k(k-1)\int \frac{\int_0^x a(u)\,dF_0(u)\int_0^x b(u)\,dF_0(u)}{F_0(x)(1-F_0(x))}\,dF_0(x). \end{aligned} \tag{14}$$

However, this follows from (12) and (11). □

PROOF OF THEOREM 2.2(ii). First, by considering the present problem as a missing data problem, it follows from Proposition A.12 of van der Vaart (1988) that

$$\int\left[\frac{dP^{1/2}_{F_{(s,0)}}(\mathbf{x}) - dP^{1/2}_{F_0}(\mathbf{x})}{s} - \frac{1}{2}(\dot{l}a)(\mathbf{x})\,dP^{1/2}_{F_0}(\mathbf{x})\right]^2 \to 0. \tag{15}$$

Furthermore, since $\widehat{F}_n$ is asymptotically linear and (12) and (15) hold, the proof that $\widehat{F}_n$ is regular follows almost line by line from the proof of Theorem 3.1 of van der Vaart (1995).

Let $\dot{\mathbf{P}}$ denote the closed linear span of $\dot{\mathbf{P}}^0 \equiv \{(\dot{l}a)\colon a \in L^0_2(F_0)\}$ in $L_2(F_0)$; $\dot{\mathbf{P}}$ is the tangent space. Equation (13) implies that for every t, $\nu(\cdot, t) \in L^0_2(F_0)$, so $\dot{l}(\cdot, t) \in \dot{\mathbf{P}}^0$ and hence $\dot{l}(\cdot, t) \in \dot{\mathbf{P}}$. By Corollary 1(A) of Bickel, Klaassen, Ritov and Wellner [(1993), page 183] or Proposition 3.3. of van der Vaart (1995), $\widehat{F}_n$ is efficient. □

APPENDIX

LEMMA A.1. *F_* satisfies (8) if and only if $F_*(t) = F_0(t)$ for every t.*

PROOF. Let $h(t) = F_*(t) - F_0(t)$. Then if F_* satisfies (8),

$$h(t) = -(k-1)\int k_{F_*}(x, t)h(x)\,dF_0(x).$$

If $k = 1$, then $h(t) \equiv 0$. So it suffices to consider the cases when $k \geq 2$. The above equation can be rewritten as

$$\begin{aligned} h(t) = &-(k-1)\int_{x\leq t}\frac{1-F_*(t)}{1-F_*(x)}h(x)\,dF_0(x) \\ &-(k-1)\int_{x>t}\frac{F_*(t)}{F_*(x)}h(x)\,dF_0(x). \end{aligned} \tag{16}$$

First, for any t, it is true that

$$\text{if } F_*(t) = 1 \text{ or } 0, \quad \text{then } h(t) = 0. \tag{17}$$

Since if $F_*(t) = 1$, (16) implies

$$1 - F_0(t) = -(k-1)\int_{x>t}(1-F_0(x))dF_0(x).$$

This equation forces $F_0(t) = 1$. So $h(t) = 0$. Similarly, it holds that if $F_*(t) = 0$, then $h(t) = 0$.

We now show that, for $t \in \{t\colon 0 < F_*(t) < 1\}$, $h(t) = 0$. By (16),

$$dh(t) = (k-1)g(t-)\,dF_*(t), \tag{18}$$

where $dF_*(t) = F_*(t) - F_*(t-)$ and $dh(t) = h(t) - h(t-)$, and

$$g(t) = \int_{x \le t} \frac{h(x)}{1 - F_*(x)}\, dF_0(x) - \int_{x>t} \frac{h(x)}{F_*(x)}\, dF_0(x). \tag{19}$$

Notice that $F_*(t)$ may not be right continuous. Hence $h(t)$ may also not be right continuous. However, we can still define $dF_*(t) = F_*(t) - F_*(t-)$ and $dh(t) = h(t) - h(t-)$. We show that if $h(t_0) > 0$ and $0 < F_*(t_0) < 1$, it will lead to a contradiction. The proof is similar to the proof of Lemma 1 of Gu and Zhang (1993). Define

$$t_1 = \sup\{t \le t_0\colon h(t) \le 0\}, \qquad t_2 = \inf\{t \ge t_0\colon h(t) \le 0\}.$$
$$J = \{t\colon h(t) > 0, t_1 \le t \le t_2\}.$$

Then $t_0 \in J$ and $(t_1, t_2) \subset J \subset [t_1, t_2]$.

We first show that $g(t) = g(t-) = 0$ on J. By (19), g is a right continuous function with

$$dg(t) = h(t)\left\{\frac{dF_0(t)}{1 - F_*(t)} + \frac{dF_0(t)}{F_*(t)}\right\} \ge 0 \quad \text{on } J. \tag{20}$$

We show that $g(t) \ge 0$ and $g(t-) \ge 0$. The proof of the cases that $g(t) \le 0$ and $g(t-) \le 0$ is similar and is omitted.

Case 1. $t_1 = -\infty$ or $F_*(t_1) = 0$. By (17), $h(t_1) = 0$. This implies

$$\int_{x \le t} (1 - F_*(x))^{-1} h(x)\, dF_0(x) \ge 0$$

in J. So $\int_{x>t} (h(x)/F_*(x))\, dF_0(x) \le 0$ by (16). Since $t_1 \notin J$, $g(t) \ge 0$ and $g(t-) \ge 0$ by (19).

For Cases 2 and 3, we show that $g(t_1-) \ge 0$ for $t_1 \in J$ and that $g(t_1) \ge 0$ for $t_1 \notin J$, which will imply $g(t) \ge 0$ and $g(t-) \ge 0$ on J by (20).

Case 2. $h(t_1) > 0, t_1 \in J$. In this case, $dh(t_1) > 0$ by the definition of t_1. Thus $g(t_1-) > 0$ by (18).

Case 3. $h(t_1) \le 0, t_1 \notin J$. Since $t_1 < t_0$, there exists $\{t_n\} \subset J$ with $t_n \downarrow t_1$ and $dh(t_n) > 0$. Therefore, $g(t_n-) \ge 0$ by (18), and hence $g(t_n) \ge 0$ by (20). This implies $g(t_1) \ge 0$ by the right continuity of g.

So we have proved that $g(t) = 0$ in J. By (18), this implies

$$h(t) = h(t_0) > 0 \quad \text{in } J \text{ and } t_0 \in J = (t_1, t_2). \tag{21}$$

Case 1. $F_*(t_2) < 1$. by (21), $h(t_2) \le 0$ and $h(t_2-) = h(t_0) > 0$, so by (18), $g(t_2-) < 0$, which is a contradiction of Step 1.

Case 2. $t_1 = -\infty$ and $F_*(t_2) = 1$. Then also $F_0(t_2) = 1$ by (17). In this case, the right-hand side of (16) is nonpositive for all t. However, the left-hand side of (16), $h(t)$, is positive in J, which is again a contradiction.

Case 3. $-\infty < t_1 < \infty$ and $F_*(t_2) = 1$. (21) implies that $h(t_1) \le 0$ and $h(t_1+) = h(t_0) > 0$. So $h(t)$ is not right continuous at t_1. Evaluate (16) at t_1+ and take the difference between $h(t_1+)$ and $h(t_1)$ to get

$$h(t_1+) - h(t_1) = (k-1)\left[\int_{x \le t_1} \frac{h(x)}{1 - F_*(x)}\, dF_0(x) - \int_{x > t} \frac{h(x)}{F_*(x)}\, dF_0(x)\right] \times (F_*(t+) - F_*(t)).$$

Since $F_*(t_1+) - F_*(t_1) = h(t_1+) - h(t_1) > 0$, this and (21) imply

$$1 = (k-1)\int_{x \le t_1} \frac{h(x)}{1 - F_*(x)}\, dF_0(x).$$

However, by the definition of t_1, the right-hand side of the this equation is less than or equal to 0, which is a contradiction. □

Acknowledgments. I thank two referees and an Associate Editor for their very constructive suggestions and comments. I especially thank one referee for pointing out an error in the proof of Lemma A.1 of an earlier version of this paper and for suggesting the current much better and shorter proof of Lemma 3.1. I also thank Jon A. Wellner for his advice.

REFERENCES

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press.

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

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic Press, New York.

COBBY, J. M., RIDOUT, M. S., BASSETT, P. J. and LARGE, R. V. (1985). An investigation into the use of ranked set sampling on grass and grass–clover swards. *Grass and Forage Science* **40** 257–263.

VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44.

GU, M. G. and ZHANG, C.-H. (1993). Asymptotic properties of self-consistent estimation based on doubly censored data. *Ann. Statist.* **21** 611–624.

HOFFMANN-JØRGENSEN, J. (1984). Stochastic process on Polish spaces. Unpublished manuscript.

KRESS, R. (1989). *Linear Integral Equations.* Springer, New York.

KVAM, P. H. and SAMANIEGO, F. J. (1994). Nonparametric maximum likelihood estimation based on ranked set samples. *J. Amer. Statist. Assoc.* **89** 526–537.

MCINTYRE, G. A. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agriculture Research* **3** 385–390.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

POLLARD, D. (1989). Asymptotics via empirical processes (with discussion). *Statist. Sci.* **4** 341–366.

STOKES, S. L. and SAGER, T. W. (1988). Characterization of a ranked-set sample with application to estimating distribution functions. *J. Amer. Statist. Assoc.* **83** 374–381.

TAKAHASI, K. and WAKIMOTO, K. (1968). On unbiased estimates of the population mean based on samples stratified by means of ordering. *Ann. Inst. Statist. Math.* **20** 1–31.

VAN DER VAART, A. W. (1988). *Statistical Estimation in Large Sample Parameter Spaces.* CWI Tract, Centrum voor Wiskunde en Informatice, Amsterdam.

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.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

WELLNER, J. A. (1989). Discussion of "Non- and semiparametric maximum likelihood estimators and the von Mises method (part I)" by Richard Gill. *Scand. J. Statist.* **16** 124–127.

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
E-MAIL: jian@stat.uiowa.edu