HAL
open science

Martingale representations of the Lynden-Bell estimator with applications

E. Strzalkowska-Kominiak, W. Stute

To cite this version:

E. Strzalkowska-Kominiak, W. Stute. Martingale representations of the Lynden-Bell estimator with applications. Statistics and Probability Letters, 2009, 79 (6), pp.814. 10.1016/j.spl.2008.10.038 . hal-00511176

HAL Id: hal-00511176

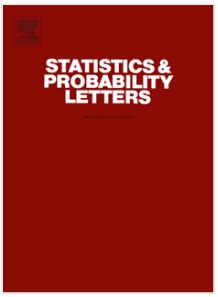
https://hal.science/hal-00511176v1

Submitted on 24 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript



Martingale representations of the Lynden-Bell estimator with applications

E. Strzalkowska-Kominiak, W. Stute

PII: S0167-7152(08)00520-8
DOI: 10.1016/j.spl.2008.10.038
Reference: STAPRO 5272

To appear in: *Statistics and Probability Letters*

Received date: 14 July 2008
Accepted date: 31 October 2008

Please cite this article as: Strzalkowska-Kominiak, E., Stute, W., Martingale representations of the Lynden-Bell estimator with applications. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.038

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Martingale Representations of the Lynden-Bell Estimator with Applications

E. Strzalkowska-Kominiak

W. Stute

Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany

Abstract

We derive a martingale representation for the Lynden-Bell estimator F_n and show that F_n fulfills linear upper and lower bounds.

Key words: Truncated data, Lynden-Bell, martingales, linear bounds

1. Introduction And Main Results

Let X and Z be two independent random variables with unknown distribution functions (d.f.'s) F and G, respectively. Under truncation from the right we observe (X, Z) only if $X \leq Z$. Truncation typically creates some dependence between the observed X and Z. Also the distribution function of X becomes

$$F^*(x) = \mathbb{P}(X \leq x | X \leq Z) = \alpha^{-1} \int\limits_{(-\infty,x]} (1 - G(y-))F(dy),$$

where $\alpha = \mathbb{P}(X \leq Z)$ is unknown but assumed to be positive. Here and in the following, for any function h, we denote with

$$h^-(y) \equiv h(y-) = \lim_{z \uparrow y} h(z) \qquad h^+(y) \equiv h(y+) = \lim_{z \downarrow y} h(z)$$

left and right hand limits and with

$$h\{y\} = h(y^+) - h(y^-)$$

ACCEPTED MANUSCRIPT

the jump size at y. Also, for any distribution function H, we set

$$a_H = \inf\{x : H(x) > 0\} \text{ and } b_H = \inf\{x : H(x) = 1\}.$$

Given a sample (X_i, Z_i), $1 \leq i \leq n$, of truncated replicates of (X, Z), the goal then is to reconstruct F from the observed data. Write

$$\Lambda(x) = \int\limits_{[x,\infty)} \frac{dF}{F}$$

for the cumulative hazard function associated with F, and set

$$C(x) = \mathbb{P}(X \leq x \leq Z | X \leq Z).$$

The function C is crucial when analyzing truncated data since, when $b_F \leq b_G$,

$$\Lambda(x) = \int\limits_{[x,\infty)} \frac{dF^*}{C} \tag{1}$$

and F^* and C are readily estimable through

$$F_n^*(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq x\}} \text{ and } C_n(x) = n^{-1} \sum_{i=1}^{n} 1_{\{X_i \leq x \leq Z_i\}}.$$

Plugging these into (1) yields the estimator of Λ,

$$\Lambda_n(x) = \int\limits_{[x,\infty)} \frac{dF_n^*}{C_n} = \sum_{i=1}^{n} \frac{1_{\{X_i \geq x\}}}{nC_n(X_i)}.$$

The product-limit formula finally leads to the time honoured Lynden-Bell (1971) estimator of F which, if there are no ties among the X's, equals

$$F_n(t) = \prod_{y>t} [1 + \Lambda_n\{y\}] = \prod_{X_i > t} \left[1 - \frac{1}{nC_n(X_i)}\right].$$

Stute and Wang (2008) showed how to break ties without destroying the product limit structure. Therefore, in this paper, we shall assume w.l.o.g. that there are no ties among the X's. Note that F_n reduces to the classical

ACCEPTED MANUSCRIPT

empirical d.f. when there is no truncation. In such a situation F_n^-/F^- is a martingale in reverse time for $t > a_F$.

It is the purpose of this paper to prove an analog and discuss some consequences, when the data are truncated.

A basic role in the analysis of F_n will be played by the process

$$H_n^1(t) = n^{-1} \sum_{i=1}^{n} 1_{\{t \le X_i \le Z_i\}}.$$

H_n^1 and C_n are adapted to the filtration

$$\mathcal{G}_n(t) = \sigma(\{X_i < s \le Z_i\}, \{s \le X_i \le Z_i\} : t \le s, 1 \le i \le n),$$

which is nondecreasing in reverse time. The process C_n is neither left- nor right-continuous. If we consider C_n^+, the right-continuous version, we obtain a function which is predictable in reverse time. The martingale in the Doob-Meyer decomposition of H_n^1 becomes, when F and G have no jumps in common,

$$M_n(t) = H_n^1(t) - \int_{[t,\infty)} \frac{C_n(u+)}{F(u)} F(du).$$

See, e.g., Mandrekar and Thelen (1990) and Keiding and Gill (1990). The "no-common jump" condition will be assumed throughout this paper without further mentioning. Separate discontinuities will, however, be allowed.

Now, on the set $\{t : C_n(t+) > 0\}$, we obtain

$$\frac{dM_n}{C_n^+} = \frac{dH_n^1}{C_n^+} + \frac{dF}{F} = d\Lambda_n - d\Lambda,$$

upon noting that the function H_n^1 has jumps of size $-\frac{1}{n}$ at the X_i and the function C_n satisfies $C_n(X_i+) = C_n(X_i)$.

Since on the support of H_n^1 the function C_n^+ is positive we therefore obtain

$$\frac{1_{\{C_n^+>0\}}}{C_n^+} dM_n = d\Lambda_n - 1_{\{C_n^+>0\}} d\Lambda \equiv d\Lambda_n - d\hat{\Lambda}_0.$$

ACCEPTED MANUSCRIPT

The hazard measure

$$\hat{\Lambda}_0(dt) = 1_{\{C_n(t+)>0\}}\Lambda(dt)$$

is random and has distribution function

$$\hat{F}_0(t) = \prod_{s>t}[1 + \hat{\Lambda}_0\{s\}]e^{-\hat{\Lambda}_0^c(t)}.$$

Here $\hat{\Lambda}_0^c$ is the continuous part of $\hat{\Lambda}_0$. The function $\hat{F}_0$ will be required for a first martingale representation of F_n.

For this, since

$$\begin{aligned}\hat{\Lambda}_0(ds) = 1_{\{C_n(s+)>0\}}\Lambda(ds) &= -1_{\{C_n(s+)>0\}}\frac{F(ds)}{F(s)} \\ &\geq -\frac{F(ds)}{F(s)} = \Lambda(ds)\end{aligned}$$

we have

$$\begin{aligned}\hat{F}_0(t-) &= \prod_{s\geq t}[1 + \hat{\Lambda}_0\{s\}]e^{-\hat{\Lambda}_0^c(t)} \\ &= \prod_{s\geq t}[1 + 1_{\{C_n(s+)>0\}}\Lambda\{s\}]e^{-\hat{\Lambda}_0^c(t)} \qquad (2) \\ &= \prod_{s\geq t, C_n(s+)>0}[1 + \Lambda\{s\}]e^{-\hat{\Lambda}_0^c(t)} \geq \prod_{s\geq t}[1 + \Lambda\{s\}]e^{-\Lambda^c(t)} = F(t-),\end{aligned}$$

where the second but last inequality follows from $0 \leq 1 + \Lambda\{s\} \leq 1$. Hence $a_{\hat{F}_0} \leq a_F$ and, consequently, $\hat{F}_0(t-) > 0$ for every $t > a_F$. Hence the process

$$t \to \frac{F_n(t-)}{\hat{F}_0(t-)} \qquad (3)$$

is well-defined on $t > a_F$.

From Gill's lemma, see Lemma 3, we obtain

$$\frac{F_n(t-)}{\hat{F}_0(t-)} = 1 + \int_{[t,\infty)} \frac{F_n(s)}{[1 + \hat{\Lambda}_0\{s\}]}\frac{1_{\{C_n(s+)>0\}}}{\hat{F}_0(s)C_n(s+)}M_n(ds).$$

ACCEPTED MANUSCRIPT

Since M_n is a martingale and the integrand is continuous from the right and hence predictable in reverse time, the process in (3) is a martingale.

When the X-data are not at risk of being truncated, $\hat{F}_0^-$ coincides with F^- on the set $t \geq \min X_i$ so that the ratio becomes F_n^-/F^-. It is interesting to note that for censored data, i.e., for the Kaplan-Meier estimator, the indicator defining the corresponding $\hat{\Lambda}_0$ is also non-vanishing there so that $\hat{F}_0$ and F coincide. As a conclusion we obtain the martingale property for the Kaplan-Meier estimator with the true F in the denominator. See Gill (1980) and Shorack and Wellner (1986). For truncated data, due to the non-monotonicity of C_n, a simple replacement of $\hat{F}_0$ by F is not possible. Rather, we need to introduce

$$T = T_n = \max\{X_i : nC_n(X_i) = 1\}.$$

Note that $nC_n(X) = 1$ when X is the smallest among the X_i's and $F_n(t)$ equals zero for $t < T$. For further analysis, the following property of T turns out to be useful.

Lemma 1. *T is a stopping time w.r.t. the filtration $\mathcal{G}_n$.*

We shall also need the following representation of T which connects T with $C_n(t+)$ but with t not necessarily belonging to the X-sample.

Lemma 2. *We have*

$$T = \sup\{t < \max_{1\leq i\leq n} Z_i : C_n(t+) = 0\}. \tag{4}$$

The proofs of Lemma 1 and 2 will be postponed to the Appendix.

If we stop the reverse martingale at T, Lemma 1 may be applied to show that the process

$$Z(t) = \frac{F_n(t \vee T-)}{\hat{F}_0(t \vee T-)}$$

is also a reverse martingale. The process Z is part of the following representation of F_n^-/F^-

Theorem 1. *We have, for $t > a_F$,*

$$\frac{F_n(t-)}{F(t-)} = Z(t)\frac{1}{F(\tilde{T} \vee t-)},$$

where $\tilde{T} = \max_{1\leq i\leq n} Z_i$.

ACCEPTED MANUSCRIPT

Proof. For $t \leq T$, both sides vanish. For $T < t < \tilde{T}$, we may proceed as for (2) to get

$$\frac{F_n(t-)}{F(t-)} = Z(t)\frac{\hat{F}_0(t-)}{F(t-)} = \frac{Z(t)}{F(t-)} \prod_{s \geq t, C_n(s+)>0} [1 + \Lambda\{s\}]e^{-\hat{\Lambda}_0^c(t)}$$
$$= \frac{Z(t)}{F(t-)} \prod_{\tilde{T} > s \geq t} [1 + \Lambda\{s\}]e^{\int_{[t,\tilde{T})} d\Lambda^c} = \frac{Z(t)}{F(t-)}\frac{F(t-)}{F(\tilde{T})} = \frac{Z(t)}{F(\tilde{T})},$$

where the third equality follows from Lemma 2. Finally, for $t \geq \tilde{T}$, $F_n(t-) = \hat{F}_0(t-) = 1$ so that the conclusion follows. □

The case of the simple empirical d.f. may be recovered from Theorem 1 if we formally set $Z_i = \infty$. Hence $F(\tilde{T} \vee t-) = 1$ for all t and $T = \min X_i$. Since $\hat{F}_0 = F$ on $t \geq T$ we obtain the aforementioned fact that for the empirical d.f. the process F_n^-/F^- is a reverse martingale. Under truncation the additional term $F(\tilde{T} \vee t-)$ may be less than one which destroys the martingale structure. Interestingly enough the next Corollary shows that some weaker martingale structure is still obtained.

Corollary 1. *F_n^-/F^- is a nonnegative reverse sub-martingale.*

Proof. Fix $t < s$. Since the process $t \to F(\tilde{T} \vee t-)$ is adapted we have, by monotonicity of the denominator,

$$\mathbb{E}\left[\frac{F_n(t-)}{F(t-)}|\mathcal{G}_n(s)\right] = \mathbb{E}\left[\frac{Z(t)}{F(\tilde{T} \vee t-)}|\mathcal{G}_n(s)\right]$$
$$\geq \frac{\mathbb{E}[Z(t)|\mathcal{G}_n(s)]}{F(\tilde{T} \vee s-)} = \frac{Z(s)}{F(\tilde{T} \vee s-)}.$$

□

2. Linear Bounds For The Lynden-Bell Estimator

In this section we derive so-called linear bounds for F_n. Such bounds have found a lot of interest, for the classical empirical d.f. and the Kaplan-Meier estimator, see Shorack and Wellner (1986).

Theorem 2. *We have, for any $\lambda > 0$,*

$$\mathbb{P}\left(\sup_{t > a_F} \frac{F_n(t-)}{F(t-)} \geq \lambda\right) \leq \left(\frac{1}{\lambda\alpha}\right)^{\frac{n}{n+1}} \left[n^{\frac{1}{n+1}} + n^{-\frac{n}{n+1}}\right] \leq 2\left(\frac{1}{\lambda\alpha}\right)^{\frac{n}{n+1}}.$$

ACCEPTED MANUSCRIPT

For the classical empirical d.f., under continuity of F and for $\lambda \geq 1$, the left-hand side equals λ^{-1}. See Daniels (1945). For truncated data the factor α needs to be included to take care of truncation effects. Finally, we may write

$$\left(\frac{1}{\lambda\alpha}\right)^{\frac{n}{n+1}} = \frac{1}{\lambda\alpha}(\lambda\alpha)^{\frac{1}{n+1}}$$

so that the right side is also $O(\lambda^{-1})$ uniformly in λ for λ varying in a bounded set or tending to infinity such that $\lambda^{\frac{1}{n+1}}$ remains bounded, as $n \to \infty$. In particular, we obtain that the ratio is uniformly bounded in probability:

$$\sup_{t>a_F} \frac{F_n(t-)}{F(t-)} = O_{\mathbb{P}}(1). \tag{5}$$

Proof. For any $a > a_F$ and positive c we have from Theorem 1

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t>a} \frac{F_n(t-)}{F(t-)} \geq \lambda\right) \\
&\leq \mathbb{P}\left(\sup_{t>a} Z(t) \geq c\lambda\right) + \mathbb{P}(F(\tilde{T}) \leq c) \leq \frac{1}{c\lambda} + \mathbb{P}^n(F(Z_1) \leq c),
\end{aligned} \tag{6}$$

where the first part of the last inequality follows from the Doob-maximal inequality for $Z(t)$ and the second is a consequence of the monotonicity of F and the independence of the Z_i's.

To proceed, let G^* be the d.f. of the actually observed Z's. Then

$$G^*(x) = \mathbb{P}(Z \leq x | X \leq Z) = \alpha^{-1}\mathbb{P}(Z \leq x, X \leq Z) \leq \alpha^{-1}F(x).$$

Hence we obtain

$$\mathbb{P}^n(F(Z_1) \leq c) \leq \mathbb{P}^n\left(G^*(Z_1) \leq \frac{c}{\alpha}\right) \leq \left(\frac{c}{\alpha}\right)^n.$$

In summary, the right side of (6) is bounded from above by

$$\frac{1}{c\lambda} + \alpha^{-n}c^n \equiv f(c).$$

Finally, the function $f(c)$ attains its minimum at

$$c^* = \left(\frac{\alpha^n}{\lambda n}\right)^{\frac{1}{n+1}}$$

ACCEPTED MANUSCRIPT

with

$$f(c^*) = \left(\frac{1}{\lambda\alpha}\right)^{\frac{n}{n+1}} \left[n^{\frac{1}{n+1}} + n^{-\frac{n}{n+1}}\right].$$

Since the bound in (6) does not depend on a, we may let a go to a_F to complete the proof. The second bound follows from the fact that [...] equals 2 for $n = 1$ and is nonincreasing in n. □

Next we study the ratio F/F_n. It is only defined for $t \geq T$ since the denominator vanishes for $t < T$. As it will turn out F/F_n is closely related to $(1 - G_n)/(1 - G)$ where

$$1 - G_n(t) = \prod_{Z_j \leq t} \left[1 - \frac{1}{nC_n(Z_j)}\right]$$

is the Lynden-Bell estimator for the survival function $1 - G$ of the left-truncated Z's. Theorem 1 appropriately modified yields a representation of $(1 - G_n)/(1 - G)$ as a forward martingale, where

$$T_n^1 = \inf\{Z_j : nC_n(Z_j) = 1\}$$

plays the same role for the Z's as T_n played for the X's. Another important quantity in this context is the unknown probability α. Note that

$$\frac{F(t)(1 - G(t-))}{C(t)} =: \alpha(t) \tag{7}$$

equals α for all $a_F < t < b_F$. This observation led He and Yang (1998) to propose

$$\hat{\alpha}_n(t) = \frac{F_n(t)(1 - G_n(t-))}{C_n(t)} \tag{8}$$

as an estimator for α. Unfortunately, the numerator of $\hat{\alpha}_n$ vanishes everywhere if $T_n^1 < T_n$. This again holds if $T_n > \min_{1 \leq i \leq n} X_i$, i.e., if there exists a so-called hole or inner risk set among the X's. In order to justify $\hat{\alpha}_n$ one therefore needs a careful study of the possibility of holes. In Strzalkowska-Kominiak and Stute (2008) it was shown that $T_n = \min_{1 \leq i \leq n} X_i$ with probability tending to one, as $n \to \infty$. Similarly, $T_n^1 = \max_{1 \leq i \leq n} Z_i$ with probability tending to one, as $n \to \infty$.

Hence, setting

$$\Omega_1^{(n)} = \left\{T_n = \min_{1 \leq i \leq n} X_i \text{ and } T_n^1 = \max_{1 \leq i \leq n} Z_i\right\}$$

ACCEPTED MANUSCRIPT

we get $\mathbb{P}(\Omega_1^{(n)}) \to 1$ as $n \to \infty$. Moreover, on $\Omega_1^{(n)}$, $\hat{\alpha}_n(t) = \hat{\alpha}$ is a (random) constant for $T_n < t < T_n^1$. See He and Yang (1998).

We are now in a position to formulate our next main result.

Theorem 3. *Assume $a_F \le a_G$ and $b_F \le b_G$. Then we have*

$$\sup_{T_n \le t} \frac{F(t)}{F_n(t)} = O_{\mathbb{P}}(1) \text{ as } n \to \infty.$$

Moreover, $T_n \to a_F$ in probability.

Proof. In view of $\mathbb{P}(\Omega_1^{(n)}) \to 1$ it suffices to study the ratio on $\Omega_1^{(n)}$. We then have

$$\frac{F(t)}{F_n(t)} = \frac{\alpha(t)}{\hat{\alpha}(t)} \frac{C(t)}{C_n(t)} \frac{1 - G_n(t-)}{1 - G(t-)}.$$

Denote with $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ the ordered X-data. Since $F_n(t) = 1$ for $t \ge X_{n:n}$ so that $F(t)/F_n(t) \le 1$ there, it suffices to consider t's such that $X_{i:n} \le t < X_{i+1:n}$ for some $i = 1, \ldots, n-1$. In such a situation, we have, in view of (7) and (8),

$$\frac{F(t)}{F_n(t)} \le \frac{F(X_{i+1:n})}{F_n(X_{i:n})} = \frac{\alpha(X_{i+1:n})}{\hat{\alpha}(X_{i+1:n})} \frac{C(X_{i+1:n})}{C_n(X_{i+1:n})} \frac{1 - G_n(X_{i+1:n}-)}{1 - G(X_{i+1:n}-)} \frac{F_n(X_{i+1:n})}{F_n(X_{i:n})}.$$

The first ratio is the same for each i and converges to 1, see He and Yang (1998). The ratio C/C_n is also bounded in probability, see Stute and Wang (2008). The ratio $(1 - G_n^-)/(1 - G^-)$ is bounded according to Theorem 2, applied to the left-truncated Z's. Finally,

$$\frac{F_n(X_{i+1:n})}{F_n(X_{i:n})} = \frac{1}{1 - \frac{1}{nC_n(X_{i+1:n})}} = \frac{nC_n(X_{i+1:n})}{nC_n(X_{i+1:n}) - 1}. \tag{9}$$

Since on $\Omega_1^{(n)}$ we have $nC_n(X_{i+1:n}) \ge 2$, the term in (9) is bounded from above by 2. That $T_n \to a_F$ in probability, is an immediate consequence of Strzalkowska-Kominiak and Stute (2008). The proof is complete. □

Appendix

Proof of Lemma 1. We shall proof the assertion by showing that T may be approximated by a sequence $T^{(m)}$, as $m \to \infty$, where for each $m \ge 1$ the variable $T^{(m)}$ is a stopping time. For this, set

$$X_i^{(m)} = \sum_{k=0}^{\infty} k2^{-m} 1_{\{k2^{-m} \le X_i < (k+1)2^{-m}\}}$$

ACCEPTED MANUSCRIPT

and

$$T^{(m)} = \max_{1\le i\le n}\left\{X_i^{(m)} : 1_{\{X_j<X_i^{(m)}\le Z_j\}} = 0 \text{ for } 1\le j\le n\right\}.$$

Then we obtain

$$\left\{T^{(m)}\ge t\right\} = \bigcup_{i=1}^{n}\left\{X_i^{(m)}\ge t : 1_{\{X_j<X_i^{(m)}\le Z_j\}} = 0 \text{ for } 1\le j\le n\right\}.$$

If t satisfies $k2^{-m} < t \le (k+1)2^{-m}$, we have

$$\left\{X_i^{(m)}\ge t\right\} = \left\{X_i^{(m)}\ge (k+1)2^{-m}\right\} = \bigcup_{l=k+1}^{\infty}\{X_i^{(m)} = l2^{-m}\}$$

whence

$$\left\{X_j < X_i^{(m)}\le Z_j\right\}\cap\left\{X_i^{(m)}\ge t\right\} = \bigcup_{l=k+1}^{\infty}\left\{X_j < l2^{-m}\le Z_j, X_i^{(m)} = l2^{-m}\right\}.$$

Since by definition of $\mathcal{G}_n$

$$\left\{X_j < l2^{-m}\le Z_j\right\}\in\mathcal{G}_n(l2^{-m})$$

and

$$\left\{X_i^{(m)} = l2^{-m}\right\} = \left\{l2^{-m}\le X_i < (l+1)2^{-m}\right\}\in\mathcal{G}_n(l2^{-m}),$$

monotonicity of $\mathcal{G}_n(t)$ in reverse time yields

$$\left\{X_j < X_i^{(m)}\le Z_j, X_i^{(m)}\ge t\right\}\in\mathcal{G}_n((k+1)2^{-m})\in\mathcal{G}_n(t).$$

This implies that $T^{(m)}$ is a stopping time. To prove the lemma it suffices to show that $T^{(m)}\to T$ as $n\to\infty$. For this, consider any $X_i^{(m)}$ such that for each $j = 1,\dots,n$ either $X_j\ge X_i^{(m)}$ or $X_i^{(m)} > Z_j$. Since $X_i^{(m)}\le X_i < X_i^{(m)} + 2^{-m}$, the inequality $X_i^{(m)} > Z_j$ yields $X_i > Z_j$. Furthermore, since there are no ties among the X_i's, there exists $M_0 = M_0(n,w)$ such that for $i\ne j = 1,\dots,n$ and $m\ge M$ we have $|X_i(w) - X_j(w)| > 2^{-m}$. Hence from $X_j\ge X_i^{(m)}$ we also get $X_j\ge X_i$. Since $X_i^{(m)}\le X_i$, we finally obtain

$$T^{(m)}(w)\le T(w) \text{ for } m\ge M_0.$$

ACCEPTED MANUSCRIPT

With similar arguments one can show that for $m \geq M_1(n, w)$

$$T^{(m)}(w) \geq T(w) - 2^{-m}.$$

This shows $T^{(m)} \to T$ pointwise and completes the proof. □

Proof of Lemma 2. Let $t < \max Z_i$ be such that $C_n(t+) = 0$. Then $1_{\{X_j \leq t < Z_j\}} = 0$ for all $j = 1, \ldots, n$. Since $t < \max Z_i$, there exists at least one $X_j > t$. For the smallest among such X_j's we have $nC_n(X_j) = 1$. Conclude that the right-hand side of (4) is less than or equal to T. Conversely, let X_{i_0} be the maximum of the X_i's for which $nC_n(X_i) = 1$. This set is nonempty since the smallest of the X's always satisfies this equation. Let Z_{j_0} be the largest among the Z_j's which is strictly smaller than X_{i_0}, if there is any. For any $Z_{j_0} \leq t < X_{i_0}$, we have $C_n(t+) = 0$. If not such Z exists, any $t < X_{i_0}$ will do the job. This shows that T is less than or equal to the right-hand side of (4). The proof is complete. □

The following lemma is an adaptation of a result due to Gill (1980) for the survival function to the left tails of a d.f.

Lemma 3. *Let A and B be two nonincreasing, left-continuous functions satisfying*

$$A\{x\} \geq -1 \text{ and } B\{x\} > -1 \text{ for all } x \in \mathbb{R}.$$

The function

$$Z(t) = 1 - \frac{\prod_{s \geq t}(1 + A\{s\}) \exp(-A^c(t))}{\prod_{s \geq t}(1 + B\{s\}) \exp(-B^c(t))}$$

satisfies the integral equation

$$\int_{[t,\infty)} \frac{1 - Z(s^+)}{1 + B\{s\}}(B(ds) - A(ds)) = Z(t).$$

In this paper $A(t)$ and $B(t)$ are the left-continuous cumulative hazard functions of F_n and $\hat{F}_0$, respectively.

References

[1] Daniels, H. E. (1945). The statistical theory of the strength of bundles of thread. Proc. Roy. Soc., London Ser. A 183, 405-435.

ACCEPTED MANUSCRIPT

[2] Gill, R. D. (1980). Censoring and Stochastic Integrals. MC Tracts 124. Amsterdam, Centre for Mathematics and Computer Science.

[3] He, S. and Yang, G. L. (1998). Estimation of the truncation probability in the random truncation model. Ann. Statist. 26, 1011-1027.

[4] Keiding, N. and Gill, R.D. (1990). Random truncation models and Markov processes. Ann. Statist. 18, 582-602.

[5] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. Mon. Not. R. Astr. Soc. 155, 95-118.

[6] Mandrekar, V. and Thelen, B. (1990). Joint weak convergence on the whole line in the truncation model. R. C. Bose Symposium on Probability, Statistics and Design of Experiments. Wiley Eastern, New Dehli, 495-515.

[7] Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

[8] Strzalkowska-Kominiak, E. and Stute, W. (2008). On the effect of holes in truncated samples. Submitted.

[9] Stute, W. and Wang, J.-L. (2008). The central limit theorem under random truncation. To appear in J. Bernoulli.

ACCEPTED MANUSCRIPT