Michael Davidsen Martin Jacobsen

# Weak Convergence of Twosided Stochastic Integrals, with an Application to Models for Left Truncated Survival Data

Preprint
January
1989
2

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Michael Davidsen* and Martin Jacobsen

# WEAK CONVERGENCE OF TWOSIDED STOCHASTIC INTEGRALS, WITH AN APPLICATION TO MODELS FOR LEFT TRUNCATED SURVIVAL DATA

Preprint 1989 No. 2

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

January 1989

*) Copenhagen County Hospital, Herlev
Herlev Ringvej 75
DK-2730 Herlev, Denmark

Summary

A class of additive processes is studied that typically arise as stochastic integrals from $s$ to $t$ when both arguments are allowed to vary. A Skorohod-type topology is introduced on the space of paths for such twosided integrals and the matching theory of weak convergence is developed and related to usual weak convergence in Skorohod spaces. As an application some asymptotic results, due to M. Woodroofe, for estimators based on left truncated survival data, are reformulated and rederived.

---

This work is based on the thesis written by M. Davidsen for the degree of cand.scient at the University of Copenhagen, with M. Jacobsen as supervisor.

## 0. Introduction

The main purpose of the paper is to discuss convergence in distribution of twosided additive stochastic processes. Such processes arise for instance as stochastic integrals from $s$ to $t$ say, when both $s$ and $t$ are allowed to vary. Of course the twosided integral is completely determined by the onesided integral obtained by keeping one argument fixed and allowing the other to vary. Typically this onesided process has sample paths belonging to some Skorohod space - right continuous with left limits. However, as will be shown below, the topology on the space of twosided integral paths inherited from the Skorohod space, will in general depend on the value of the fixed argument. Hence the need for a new approach, to be developed below in the case where $s \leq t$ vary in an open interval so that there is no outstanding candidate, such as an interval endpoint, for a fixed argument value.

In Section 1 we introduce the topology on the space of paths for additive processes. Section 2 treats the matching weak convergence of probabilities on this space. Finally, Section 3 contains an application involving non-parametric estimators of a survival distribution based on left truncated survival data.

## 1. The space $D(\Lambda)$

Let $\Lambda = \{(s,t): 0 < s \leq t\}$ and let $D(\Lambda)$ denote the space of all functions $w: \Lambda \to \mathbb{R}$ with the following two properties:

(i) $w(s,u) = w(s,t) + w(t,u)$ $\qquad (0 < s \leq t \leq u)$

(ii) $w(s,t)$ is right continuous with left limits in either variable $s$ or $t$.

Taking $t = u$ in (i) shows that

(iii) $w(t,t) = 0$ $\qquad (t > 0)$.

It is critical that only strictly positive arguments $s$ and $t$ are allowed. The applications we have in mind involve cases, where expressions like $w(0,t)$ are meaningless.

For an obvious example of $D(\Lambda)$-functions, let $\mu$ be a $\sigma$-finite measure on $\mathbb{R}_{++} = (0,\infty)$ and let $f: \mathbb{R}_{++} \to \mathbb{R}$ be a function which is locally $\mu$-integrable in the sense that $\int_K |f| d\mu < \infty$ for any compact $K \subset \mathbb{R}_{++}$. Then

$$(1.1) \qquad w(s,t) = \int_{(s,t]} f(u)\mu(du)$$

belongs to $D(\Lambda)$. Note that since we only consider $s,t > 0$, we allow for $\int_{(0,t]} |f| d\mu$ to diverge.

If $w \in D(\Lambda)$ it is possible to define $w(s,t)$ also if $s > t > 0$ and still retain properties (i) - (iii), viz. as (1.1) suggests, define

$$w(s,t) = - w(t,s), \tag{1.2}$$

if $0 < t < s$. So from now on we assume $w(s,t)$ to be defined for all $0 < s,t$ and use (i) - (iii) and the antisymmetry (1.2).

Our task in this section is to define a Skorohod type topology on $D(\Lambda)$. First the reader is reminded about the following standard facts and notation.

Let $I \subset \mathbb{R}$ be an arbitrary (bounded or unbounded) interval and let $D(I)$ denote the space of all functions $v: I \to \mathbb{R}$, right continuous with left limits everywhere. The time deformation group for $I$ is the collection $\Lambda_I$ of bijections $\lambda: I \to I$ which are strictly increasing and continuous. Then $v_n \to v$ in the Skorohod $D(I)$-topology (Skorohod [5], Billingsley [1], Lindvall [4], Whitt [6]) if there exists a sequence $\lambda_n \in \Lambda_I$ such that

(1.3)
(a) $\|\lambda_n - e_I\|_I \to 0$
(b) $\|v_n \circ \lambda_n - v\|_K \to 0$ for all compact intervals $K \subset I$.

Here $e_I$ denotes the identity $e_I(t) = t$ on $I$, and if $J$ is an interval and $f$ is a real valued function, defined on a domain containing $J$, we write

$$\|f\|_J = \sup_{t \in J} |f(t)|.$$

If $I = \mathbb{R}_{++}$, we omit index I. In particular $\Lambda$ is the time deformation group on $\mathbb{R}_{++}$. Also for $I = [t_0,\infty)$, we write index $t_0$ instead of $[t_0,\infty)$.

Recall that (1.3a) may be replaced by

(1.3a') $\quad \|\lambda_n - e_I\|_K \to 0$ for all compact intervals $K \subset I$.

Recall also that $D(I)$ with the Skorohod topology is metrizable as a Polish space.

Returning now to the space $D(\Lambda)$, fix $s_0 > 0$ and consider the map $\psi_{s_0}: D(\Lambda) \to D(\mathbb{R}_{++})$ given by

$$(\psi_{s_0} w)(t) = w(s_0,t) \qquad (t \in \mathbb{R}_{++}).$$

It is immediately checked that $\psi_{s_0}$ is a bijection from $D(\Lambda)$ onto $D_{s_0}(\mathbb{R}_{++}) = \{v \in D(\mathbb{R}_{++}): v(s_0) = 0\}$ with inverse

$$(\psi_{s_0}^{-1} v)(s,t) = v(t) - v(s) \qquad (s,t \in \mathbb{R}_{++}).$$

At a first glance it would appear natural to equip $D(\Lambda)$ with the topology obtained by using on $D_{s_0}(\mathbb{R}_{++})$ the topology it inherits as a subspace of the Skorohod $D(\mathbb{R}_{++})$-space, and then demanding that $\psi_{s_0}$ be a homeomorphism. A quick check reveals however, that this topology depends on the choice of $s_0$. We therefore take a different approach and present the following basic

1.4 Definition A sequence $(w_n)_{n\geq 1}$ of $D(\Lambda)$-functions converges to $w \in D(\Lambda)$ in the Skorohod-$D(\Lambda)$ topology if and only if there exists a sequence $(\lambda_n)$ of time deformations in $\Lambda$ such that

(1.5)
(a) $\|\lambda_n - e\| \to 0$

(b) $\sup_{s\in K,\, t\in L} |w_n(\lambda_n s, \lambda_n t) - w(s,t)| \to 0$
for all compact intervals $K, L \subset \mathbb{R}_{++}$. □

Various equivalent forms are available: in (b) it suffices to take $K = L$ and (a) may be replaced by (1.3a') (with $I = \mathbb{R}_{++}$).

For convenience we shall write $\|w_n \circ \lambda - w\|_{K,L}$ for the supremum in (1.5b).

It should be clear that Definition 1.4 defines a topology on $D(\Lambda)$: indeed, a subbase of neighborhoods of $w \in D(\Lambda)$ is given by the collection of sets of the form

$$U(\epsilon,K,L,w) = \{w' \in D(\Lambda)\colon \exists\, \lambda \in \Lambda \text{ such that } \|\lambda - e\| < \epsilon,\ \|w' \circ \lambda - w\|_{K,L} < \epsilon\}$$

for arbitrary $\epsilon > 0$ and $K, L \subset \mathbb{R}_{++}$ compact intervals.

Note that $s,t$ are treated symmetrically in (1.5b). This is not the case with the topology on $D(\Lambda)$ described above, which was derived from the Skorohod topology on $D_{s_0}(\mathbb{R}_{++})$.

Before listing some properties of the Skorohod $D(\Lambda)$-topology, we need to discuss the discontinuities of $D(\Lambda)$-functions.

Let $w \in D(\Lambda)$. For $s_0 > 0$ fixed, $\psi_{s_0} w$ has at most countably many points of discontinuity. Furthermore, since

$$(\psi_{s_0} w)(t) - (\psi_{t_0} w)(t) = w(s_0, t_0),$$

the discontinuity set does not depend on $s_0$, and we are allowed to define the set of <u>continuity points</u> for $w \in D(\Lambda)$ as

(1.6) $$C(w) = \{t \in \mathbb{R}_{++} : \psi_{s_0} w \text{ is continuous at } t\}$$

for any $s_0 > 0$.

It is immediate that $w \in D(\Lambda)$ is continuous at $(s,t) \in \Lambda$ iff $s,t \in C(w)$.

1.7 <u>Proposition</u> (a) The following three conditions are equivalent:

(i) $w_n \to w$ in $D(\Lambda)$;

(ii) for some, and then automatically for all $t_0 > 0$, there exists a sequence of $\Lambda$-functions such that

(1.8) $$\|\lambda_n - e\| \to 0,$$
$$\|(\psi_{\lambda_n t_0} w_n) \circ \lambda_n - \psi_{t_0} w\|_L \to 0 \text{ for all compact}$$
intervals $L \subset \mathbb{R}_{++}$;

(iii) for some, and then automatically for all $t_0 \in C(w)$, $\psi_{t_0} w_n \to \psi_{t_0} w$ in $D(\mathbb{R}_{++})$.

(b) For any $t_0 > 0$, the mapping $\psi_{t_0} : D(\Lambda) \to D_{t_0}(\mathbb{R}_{++})$ is continuous at $w \in D(\Lambda)$ provided $t_0 \in C(w)$. The inverse mapping $\psi_{t_0}^{-1}: D_{t_0}(\mathbb{R}_{++}) \to D(\Lambda)$ is everywhere continuous.

Proof (a) Taking $K = \{t_0\}$ in (1.5b) shows that (i) $\Rightarrow$ (ii) for all $t_0$. Conversely, if (ii) holds for some $t_0 > 0$, find $(\lambda_n)$ such that (1.8) is true. Since by additivity and antisymmetry

$$\begin{aligned} &|w_n(\lambda_n s,\lambda_n t) - w(s,t)| \\ &\leq |w_n(\lambda_n t_0,\lambda_n s) - w(t_0,s)| + |w_n(\lambda_n t_0,\lambda_n t) - w(t_0,t)|, \end{aligned}$$

(1.5b) and (i) follows.

Assume now that (ii) holds, and take an arbitrary $t_0 \in C(w)$. Find $(\lambda_n)$ such that (1.8) holds for this $t_0$. To establish (iii) we must show that for all $L$ compact,

(1.9)
$$\|(\psi_{t_0} w_n) \circ \lambda_n - \psi_{t_0} w\|_L \to 0.$$

But for all $t$,

$$\begin{aligned} &((\psi_{t_0} w_n) \circ \lambda_n)(t) - (\psi_{t_0} w)(t) \\ &= (\psi_{\lambda_n t_0} w_n(\lambda_n t) - (\psi_{t_0} w)(t)) + w_n(t_0,\lambda_n t_0), \end{aligned}$$

so it suffices to show that $w_n(t_0,\lambda_n t_0) \to 0$. But since $\lambda_n^{-1} t_0 \to t_0$, using (1.8) and the assumption that $t_0 \in C(w)$, it follows easily that $w_n(t_0,\lambda_n t_0) \to w(t_0,t_0) = 0$.

That (iii) for one $t_0 \in C(w)$ implies (i), follows trivially from the last assertion in (b) proved below. The proof of part (a) is complete.

(b). The first assertion is just the implication (i) $\Rightarrow$ (iii). For the second, let $t_0 > 0$ and assume $v_n$, $v$ are $D_{t_0}(\mathbb{R}_{++})$-functions such that $v_n \to v$ in $D(\mathbb{R}_{++})$. Find $\lambda_n \in \Lambda$ such that $\|\lambda_n - e\| \to 0$ and $\|v_n \circ \lambda_n - v\|_K \to 0$ for all compact $K \subset \mathbb{R}_{++}$. Then for $K,L \subset \mathbb{R}_{++}$ compact

$$\|(\psi_{t_0}^{-1} v_n) \circ \lambda_n - \psi_{t_0}^{-1} v\|_{K,L}$$

$$= \sup_{s \in K,\, t \in L} |(v_n(\lambda_n t) - v_n(\lambda_n s)) - (v(t) - v(s))| \to 0. \qquad \square$$

From the proposition it follows in particular that if $w_n \to w$, then $w_n(s,t) \to w(s,t)$ if $s,t \in C(w)$.

1.10 <u>Proposition</u> The Skorohod space $D(\Lambda)$ is metrizable as a Polish space. $\square$

We shall only outline the proof. For $0 < s < t$ define a distance between the restrictions to $[s,t]$ of two $D(\Lambda)$-functions $w_1, w_2$ by

$$d_{st}(w_1, w_2) = \inf_{\lambda \in \Lambda_{st}} (\|\lambda - e\|_{st} \vee \|w_1 - w_2 \circ \lambda\|_{st}),$$

where $st$ is short for $[s,t]$ and

$$\|w_1 - w_2 \circ \lambda\|_{st} = \sup_{u,u' \in [s,t]} |w_1(u,u') - w_2(\lambda u, \lambda u')|.$$

Then (cf. [6, Section 2]),

$$d(w_1,w_2) = \int_0^1 \int_1^\infty e^{s-t}(d_{st}(w_1,w_2) \wedge 1)\, dt\, ds$$

defines a metric for the Skorohod-$D(\Lambda)$-topology. Thus $D(\Lambda)$ is metrizable and it is then readily seen, that it is separable: convert a countable, dense subset A of $D(\mathbb{R}_{++})$ to a countable, dense subset A' of $D_{t_0}(\mathbb{R}_{++})$ by subtracting from any $v \in A$ the constant $v(t_0)$, and note that since $\psi_{t_0}^{-1}$ is continuous (Proposition 1.7b) and onto, $\psi_{t_0}^{-1}(A')$ is countable and dense in $D(\Lambda)$.

Finally, to prove completeness, one must modify the metric $d$ by considering only timedeformations that are not too steep, exactly as in [1, p.113].

## 2. Weak convergence

Let $P_n$ for $n \geq 1$ and $P$ be probabilities on a metric space. We write $P_n \Rightarrow P$ if $P_n$ converges weakly to $P$ as $n \to \infty$, i.e.

$$\int f dP_n \to \int f dP$$

for all bounded and continuous $f$.

We shall now discuss weak convergence of probabilities on $D(\Lambda)$. Since $D(\Lambda)$ is Polish, Prohorov's theorem applies and consequently $P_n \Rightarrow P$ iff there is weak convergence of all finite-dimensional distributions corresponding to continuity points for $P$ (see below) and the family $(P_n)$ is tight.

We shall not here discuss conditions for tightness, but instead relate weak convergence of probabilities on $D(\Lambda)$ to the standard case of weak convergence of probabilities on $D(\mathbb{R}_{++})$ (or rather $D[t_0,\infty)$ for almost all $t_0 > 0$).

Let $t > 0$, let $r_t$ denote the map that restricts a function with a domain containing $[t,\infty)$ to $[t,\infty)$, e.g. $r_t v = (v(u))_{v \geq t}$ for $v \in D(\mathbb{R}_{++})$. Also, write $\eta_t = r_t \circ \psi_t$, so that $\eta_t : D(\Lambda) \to D[t,\infty)$ and

$$(2.1) \qquad (\eta_t w)(u) = w(t,u) \qquad (u \geq t).$$

If $P$ is a probability on $D(\Lambda)$, introduce $T_P$ as the points of continuity for $P$, i.e.

$$T_P = \{t > 0 : P(C_t) = 1\},$$

where

$$C_t = \{w \in D(\Lambda) : t \in C(w)\}.$$

Recalling that $t \in C(w)$ iff $t$ is a continuity point for one (and then all) $\psi_{s_0} w$ (see (1.6)), and using standard properties of probabilities on $D(\mathbb{R}_{++})$, it follows that $T_P$ is dense in $\mathbb{R}_{++}$.

The main result we shall prove is the following.

2.2 Theorem For $(P_n)_{n\geq 1}$, $P$ probabilities on $D(\Lambda)$, $P_n \Rightarrow P$ if and only if $\eta_t(P_n) \Rightarrow \eta_t(P)$ for all $t \in T_P$.

Notation $\eta_t(P_n)$, $\eta_t(P)$ are $P_n$, $P$ transformed by $\eta_t$.

Proof Whitt [6, Theorem 2.8] showed that weak convergence on $D(\mathbb{R}_{++})$ amounts (essentially) to weak convergence on $D[s,t]$ for all $0 < s < t$. We follow his proof in order to obtain the non-trivial half of the theorem.

If $P_n \Rightarrow P$, trivially $\eta_t(P_n) \Rightarrow \eta_t(P)$ for $t \in T_P$ because $\eta_t = r_t \circ \psi_t$ is P-a.s continuous, since $\psi_t$ is a.s. continuous by Proposition 1.7 b, and $r_t$ is continuous at $v \in D(\mathbb{R}_{++})$, whenever $v$ is continuous at $t$.

Suppose conversely that $\eta_t(P_n) \Rightarrow \eta_t(P)$ for all $t \in T_P$. In order that $P_n \Rightarrow P$ it is necessary and sufficient that

$$\limsup_{n\to\infty} P_n(F) \leq P(F) \tag{2.3}$$

for any closed set $F \subset D(\Lambda)$.

For $t > 0$, write $H_t = \eta_t^{-1}(\overline{\eta_t F})$ where $\overline{\eta_t F}$ is the closure in $D[t,\infty)$ of the image of $F$ under $\eta_t$. Since $T_P$ is dense, we can choose a sequence $t_k \downarrow 0$ of points in $T_P$. Write $H_k = H_{t_k}$. Now $H_k \supset F$ and since $\eta_{t_k}(P_n) \Rightarrow \eta_{t_k}(P)$ as $n \to \infty$, for all $k$

$$\limsup_{n\to\infty} P_n(F) \leq \limsup_{n\to\infty} P_n(H_k) = \limsup_{n\to\infty} \eta_{t_k}(P_n)(\overline{\eta_{t_k} F})$$
$$\leq \eta_{t_k}(P)(\overline{\eta_{t_k} F}) = P(H_k).$$

Thus (2.3) will follow, if we show that

(2.4) $$H_k \supset H_{k+1} \text{ P-a.s. for all } k,$$

(2.5) $$\bigcap_{k=1}^{\infty} H_k \subset F.$$

For (2.4), because $t_k$, $t_{k+1} \in T_P$, it is enough to show that if $w \in H_{k+1}$ is such that $t_k$, $t_{k+1} \in C(w)$, then $w \in H_k$. First note that for any $w' \in D(\Lambda)$,

(2.6) $$\eta_{t_k} w' = r_{t_k}(\eta_{t_{k+1}} w') - w'(t_{k+1}, t_k).$$

Since $\eta_{t_{k+1}} w \in \overline{\eta_{t_{k+1}} F}$ we can find a sequence $(w_n)$ from $F$ such that $\eta_{t_{k+1}} w_n \to \eta_{t_{k+1}} w$ in $D[t_{k+1}, \infty)$ as $n \to \infty$. Because $t_k \in C(w)$, in particular $\eta_{t_{k+1}} w$ is continuous at $t_k$, and therefore $r_{t_k}$ is continuous at $\eta_{t_{k+1}} w$. From (2.6) it follows that as $n \to \infty$,

$$\eta_{t_k} w_n \to r_{t_k}(\eta_{t_{k+1}} w) - w(t_{k+1}, t_k) = \eta_{t_k} w,$$

i.e. $\eta_{t_k} w \in \overline{\eta_{t_k} F}$, which is exactly to say that $w \in H_k$.

To show (2.5), assume that $w \in \cap H_k$. We shall exhibit a sequence $(w_k)$ from F such that $w_k \to w$. Since F is closed, $w \in F$ then follows. By assumption, for any k, $\eta_{t_k} w \in \overline{\eta_{t_k} F}$, and we can find a sequence $(w_{kn})_{n\geq 1}$ from F such that $\eta_{t_k} w_{kn} \to \eta_{t_k} w$ as $n \to \infty$. By the definition of convergence in $D[t_k,\infty)$, this implies that there exists integers $n_k$ and $\lambda_k' \in \Lambda_{t_k}$ such that

(2.7) $$\|\lambda_k' - e_{t_k}\|_{t_k} \leq \frac{1}{k}$$

(2.8) $$\sup_{t\in[t_k,k]} |w_{kn_k}(t_k,\lambda_k' t) - w(t_k,t)| \leq \frac{1}{k}.$$

Using Proposition 1.7a, we show that $w_{kn_k} \to w$ in $D(\Lambda)$ by showing that $\psi_{t_1} w_{kn_k} \to \psi_{t_1} w$ in $D(\mathbb{R}_{++})$ as $k \to \infty$. To this end, define for $k = 1,2,\ldots$

$$\lambda_k t = \begin{cases} \lambda_k' t & t \geq t_k \\ t & 0 < t \leq t_k. \end{cases}$$

Clearly $\lambda_k \in \Lambda$ and $\|\lambda_k - e\| = \|\lambda_k' - e_{t_k}\|_{t_k} \to 0$ by (2.7). It remains to show uniform convergence on compacta of $\psi_{t_1} w_{kn_k}$ to $\psi_{t_1} w$. Since any compact set is contained in $[t_k,k]$ for k large enough, for this it clearly suffices to show that

$$\sup_{t\in[t_k,k]} |w_{kn_k}(t_1,\lambda_k t) - w(t_1,t)| \to 0.$$

But the supremum is

$$\leq \sup_{t\in[t_k,k]} |w_{kn_k}(t_k,\lambda_k t) - w(t_k,t)| + |w_{kn_k}(t_k,t_1) - w(t_k,t_1)|.$$

The first term is $\leq \frac{1}{k}$ by (2.8). The second equals

$$|w_{kn_k}(t_k,\lambda_k(\lambda_k^{-1}t_1)) - w(t_k,t_1)|$$
$$\leq |w_{kn_k}(t_k,\lambda_k(\lambda_k^{-1}t_1)) - w(t_k,\lambda_k^{-1}t_1)| + |w(\lambda_k^{-1}t_1,t_1)|.$$

Because of (2.7), $\lambda_k^{-1}t_1 \to t_1$, so by (2.8) the first term is $\leq \frac{1}{k}$ for $k$ large. The second $\to 0$ because $t_1 \in C(w)$. □

Suppose now that $(X_n)_{n\geq 1}$, $X$ are real valued stochastic processes, $X_n = (X_n(s,t))_{s,t>0}$, $X = (X(s,t))_{s,t>0}$ with sample paths in $D(\Lambda)$. Write $P_n$, $P$ for the distribution of $X_n$, $X$ respectively and $X_n \overset{d}{\to} X$ if $X_n$ converges in distribution to $X$, i.e. if $P_n \Rightarrow P$. With $T_X = T_P$ the theorem may be restated as follows: $X_n \overset{d}{\to} X$ (in $D(\Lambda)$) iff $\eta_{t_0} X_n \overset{d}{\to} \eta_{t_0} X$ (in $D[t_0,\infty)$) for every $t_0 \in T_X$.

We shall conclude this section with a discussion of when it is possible to include $t_0 = 0$ so as to deduce weak convergence on $D[0,\infty)$ from weak convergence on $D(\Lambda)$.

For this problem to make sense at all, it is necessary that the $X_n$ and $X$ in a natural fashion extend to processes defined also at time $0$. More specifically, assume that for some (and then as seen below, automatically for all) $t_0 > 0$, the limits

(2.9) $$X_{+n}(t_0) = \lim_{h\downarrow 0} X_n(h,t_0) \qquad (n \geq 1),$$

(2.10) $$X_{+}(t_0) = \lim_{h\downarrow 0} X(h,t_0)$$

exist almost surely.

Using additivity it follows immediately that if (2.9) holds, then a.s. $X_{+n}(t) = \lim_{h\downarrow 0} X_n(h,t)$ exists simultaneously for all $t$. Also

$$X_{+n}(t) - X_{+n}(s) = X_n(s,t),$$
$$\lim_{h\downarrow 0} X_{+n}(h) = 0 \text{ a.s.}$$

Thus we may define processes $X_{+n} = (X_{+n}(t))_{t\geq 0}$, $X_+ = (X_+(t))_{t\geq 0}$ with paths in $D[0,\infty)$, always taking the value 0 for $t = 0$, and may then ask when $X_{+n} \overset{d}{\to} X_+$ (in $D[0,\infty)$), assuming that $X_n \overset{d}{\to} X$.

First, consider the following simple example: $P_n$, $P$ are degenerate with unit mass at $w_n$, $w \in D(\Lambda)$ respectively, where

$$w_n(s,t) = \int_s^t a_n(1_{\left[\frac{1}{n},\frac{2}{n}\right]}(u) - 1_{\left[\frac{2}{n},\frac{3}{n}\right]}(u))du, \qquad w \equiv 0.$$

No matter what are the constants $a_n$, $w_n \to w$ in $D(\Lambda)$ and thus $P_n \Rightarrow P$. Also, (2.8), (2.9) hold and almost surely

$$X_{+n}(t) = \begin{cases} 0 & (t \leq \frac{1}{n}) \\ a_n(t - \frac{1}{n}) & (\frac{1}{n} < t \leq \frac{2}{n}) \\ a_n(\frac{3}{n} - t) & (\frac{2}{n} < t \leq \frac{3}{n}) \\ 0 & (t > \frac{2}{n}) \end{cases}; \quad X_+(t) = 0.$$

From this it is clear that the finite-dimensional distributions of $X_{+n}$ converge to those of $X_+$, but it is also clear that $X_{+n}$ need not converge in distribution to $X_+$, indeed $X_{+n} \overset{d}{\to} X_+$ iff $a_n/n \to 0$.

Thus, to obtain the desired $D[0,\infty)$-convergence, one must be able to control all process values close to 0. The precise formulation as given in condition (2.12) below, may be referred to as "tightness close to 0".

2.11 <u>Theorem</u> Let $(X_n)_{n\geq 1}$, $X$ be processes with sample paths in $D(\Lambda)$ such that $X_n \overset{d}{\to} X$ and (2.9), (2.10) hold almost surely. In order that $X_{+n} \overset{d}{\to} X_+$ (in $D[0,\infty)$) it is necessary and sufficient that the following condition holds:

(2.12)
$$\forall \epsilon, \eta > 0 \; \exists \delta > 0, n_0 \in \mathbb{N} \; \forall n \geq n_0$$
$$\Pr(\sup_{0<s\leq t\leq \delta} |X_n(s,t)| > \epsilon) < \eta.$$

<u>Notation</u> The processes $X_n$, $X$ may be defined on different probability spaces, but we always write Pr for the relevant probability and E for the relevant expectation.

<u>Proof</u> Assume first that $X_{+n} \overset{d}{\to} X_+$ (in $D[0,\infty)$). In particular, for a

given $0 < t_0 \in T_X$, there is weak convergence (in $D[0,t_0]$) of the restrictions of the $X_{+n}$ to $[0,t_0]$, towards the restriction of $X_+$. In particular, by (15.8) in [1, Theorem 15.3],

$$\forall \epsilon,\ \eta > 0 \quad \exists\, 0 < \delta < t_0,\ n_0 \in \mathbb{N} \quad \forall\, n \geq n_0$$

$$\Pr\Big( \sup_{0 \leq s \leq t \leq \delta} |X_{+n}(t) - X_{+n}(s)| > \epsilon \Big) < \eta.$$

Since $X_{+n}(t) - X_{+n}(s) = X_n(s,t)$, this is precisely (2.12).

Suppose now that we have tightness close to 0. In order to show $X_{+n} \overset{d}{\to} X_+$, we must show that for all $t_0 \in T_X$, $(X_{+n}(t))_{0 \leq t \leq t_0} \overset{d}{\to} (X_+(t))_{0 \leq t \leq t_0}$ (in $D[0,t_0]$). For this, by [1, Theorem 15.4], it is enough to show that (i) the finite-dimensional distributions of $X_{+n}$ converge to those of $X_+$ when all timepoints involved belong to $T_X$; and (ii) the following condition holds for any $t_0 \in T_X$:

$$\forall\ \epsilon,\eta > 0\ \exists\, 0 < \delta < t_0,\ n_0 \in \mathbb{N}\ \forall\ n \geq n_0$$

(2.13)

$$\Pr\Big( \sup_{\substack{0 \leq t_1 \leq t \leq t_2 \leq t_0 \\ |t_2 - t_1| \leq \delta}} |X_{+n}(t) - X_{+n}(t_1)| \wedge |X_{+n}(t_2) - X_{+n}(t)| > \epsilon \Big) < \eta.$$

Proof of (i). Because $X_{+n}(0)$, $X_+(0)$ both = 0 a.s., we need only show that for any choice of $k$ and $0 < t_1 \leq \ldots \leq t_k \in T_X$,

(2.14) $$(X_{+n}(t_1),\ldots,X_{+n}(t_k)) \overset{d}{\to} (X_+(t_1),\ldots,X_+(t_k)).$$

Given $\epsilon, \eta > 0$ find $0 < \delta < t_1$, $\delta \in T_X$ such that (2.12) holds for n sufficiently large. Because $X_n \overset{d}{\to} X$,

$$(2.15) \qquad (X_n(\delta,t_1),\ldots,X_n(\delta,t_k)) \overset{d}{\to} (X(\delta,t_1),\ldots,X(\delta,t_k)).$$

By (2.12) the random vector on the left hand side of (2.14) is close to the vector on the left of (2.15) with high probability. Because $X_+(t) = \lim_{\delta\downarrow 0} X(\delta,t)$, the same is true for the two right hand sides if $\delta$ is small, and (2.14) follows easily.

Proof of (ii). Since $X_n \overset{d}{\to} X$, by Theorem 2.2, given $0 < s_0 < t_0$, $s_0,t_0 \in T_X$, $(X_n(s_0,t))_{s_0\le t\le t_0} \overset{d}{\to} (X(s,t))_{s_0\le t\le t_0}$ (in $D[s_0,t_0]$), so by (15.7) in [1, Theorem 15.3],

$$\forall\, \epsilon, \eta > 0\ \exists\ 0 < \delta < t_0 - s_0,\ n_0 \in \mathbb{N}\ \forall\, n \ge n_0$$

$$(2.16) \qquad \Pr\Big(\sup_{\substack{s_0\le t_1\le t\le t_2\le t_0\\ t_2-t_1\le\delta}} |X_n(s_0,t) - X_n(s_0,t_1)| \wedge |X_n(s_0,t_2) - X_n(s_0,t)| > \tfrac{\epsilon}{2}\Big) < \tfrac{\eta}{2}.$$

Note that since e.g. $X_{+n}(t) - X_{+n}(t_1) = X_n(s_0,t) - X_n(s_0,t_1) = X_n(t_1,t)$, (2.12) formally emerges from (2.16) by replacing $s_0$ by 0.

Write $S_n$ for the supremum appearing in (2.16) and introduce

$$R_n = \sup_{0<s\le t\le s_0} |X_n(s,t)|.$$

Then whenever $0 \leq t_1 \leq t \leq t_2 \leq t_0$, $t_2 - t_1 \leq \delta$ the minimum of differences in (2.13) equals

$$(2.17)\qquad |X_n(t_1,t)| \wedge |X_n(t,t_2)| \leq \begin{cases} R_n & \text{if } t \leq s_0 \\ R_n + S_n & \text{if } t_1 \leq s_0 < t, \\ S_n & \text{if } s_0 < t_1 \end{cases}$$

and therefore is always $\leq R_n + S_n$. Thus to prove (2.13) for $t_0 \in T_X$, given $\epsilon, \eta > 0$ first use (2.12) to choose $0 < s_0 < t_0$, $s_0 \in T_X$, so that $\Pr(R_n > \frac{\epsilon}{2}) < \frac{\eta}{2}$ for $n$ sufficiently large. Then pick $\delta$ in accordance with (2.16) and use (2.17) to arrive at (2.13). □

## 3. An application to models for left truncated survival data

In this section we review some results due to Woodroofe [7], using the theory of the preceding sections.

Let $F$, $G$ denote two distribution functions for probabilities on $\mathbb{R}_{++}$ and consider $n$ i.i.d. pairs of $\mathbb{R}_{++}$-valued random variables $(X_1,Y_1),\dots,(X_n,Y_n)$, where the distribution of $(X_i,Y_i)$ is that of a pair $(U,V)$ conditionally on the event $(V < U)$ with $U$, $V$ independent and having distribution functions $F$ and $G$ respectively. We assume of course that

$$\alpha = \Pr(V < U) > 0.$$

Thus in particular $Y_i < X_i$ a.s.

Referring to each $i$ as an item, $X_i$ is the failure time and $Y_i$ the truncation time for item $i$. We say that $i$ is at risk at time

$t > 0$ if $Y_i \leq t < X_i$, corresponding to $I_i(t) = 1$ where

$$I_i(t) = 1_{(Y_i \leq t < X_i)}.$$

Now consider the process $N_n$ counting the number of observed failures,

$$N_n(t) = \sum_{i=1}^{n} 1_{(X_i \leq t)}$$

and denote by $\mathcal{F}_t$ the $\sigma$-algebra generated by all observations before $t$, i.e. $(N_n(s))_{s \leq t}$ and the events $(Y_i \leq s)$ for $i = 1,\ldots,n$, $s \leq t$.

Assume from now on that $F$ is absolutely continuous with hazard function $\mu$, so that

$$F(t) = 1 - \exp(-\int_0^t \mu(s)ds).$$

Also, for convenience assume that $\int_0^t \mu < \infty$ and that $G(t) > 0$ for all $t \in \mathbb{R}_{++}$.

It is standard (e.g. Keiding and Gill [2]), that with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$, the increasing process $N_n$ has compensator

$$\Lambda_n(t) = \int_0^t \mu(s) 1_{(R_n(s) > 0)} ds,$$

where $R_n(s) = \sum_{i=1}^{n} I_i(s)$. Also then, the integrated hazard $\int_0^t \mu$ may be estimated by the Nelson-Aalen estimator

$$\hat{\beta}_n(t) = \int_{(0,t]} \frac{1}{R_n(s-)} N_n(ds)$$

and in particular, defining

$$\beta_n(t) = \int_0^t \mu(s) 1_{(R_n(s)>0)} ds,$$

$\hat{\beta}_n - \beta_n$ is a martingale.

Now introduce the process $X_n$,

$$X_n(s,t) = \sqrt{n}((\hat{\beta}_n(t) - \beta_n(t)) - (\hat{\beta}_n(s) - \beta_n(s)))$$

$$= \sqrt{n}\left[ \int_{(s,t]} \frac{1}{R_n(u-)} N_n(du) - \int_s^t \mu(u) 1_{(R_n(u)>0)} du \right],$$

with paths in $D(\Lambda)$. Woodroofe [7, Theorem 3] showed that for any $t_0 > 0$,

(3.1) $$\eta_{t_0} \circ X_n \overset{d}{\rightarrow} X_{t_0}$$

in $D[t_0,\infty)$, where $X_{t_0} = (X_{t_0}(t))_{t \geq t_0}$ is the mean zero Gaussian process with independent increments and variance function $EX_{t_0}^2(t) = \sigma^2(t_0,t)$, where

$$\sigma^2(t_0,t) = \alpha \int_{t_0}^t \frac{1}{G(u)(1-F(u))} \mu(u) du.$$

(Recall that

$$\alpha = \Pr(V < U) = \int_0^\infty G(u)(1 - F(u))\mu(u)du.$$

In (3.1) the asymptotics are given for the stochastic integrals $\hat{\beta}_n - \beta_n$ from $t_0$ and out. By Theorem 2.2, Woodroofe's result may immediately be restated as follows:

$$X_n \xrightarrow{d} X, \tag{3.2}$$

with $X = (X(s,t))_{0<s,t}$ the continuous mean zero Gaussian process, uniquely determined by the requirements that it has paths in $D(\Lambda)$, that $X(s,t)$, $X(u,v)$ are independent whenever $s \leq t \leq u \leq v$ and that the variance function is

$$EX^2(s,t) = \sigma^2(s,t) \qquad ((s,t) \in \Lambda).$$

(3.1) is proved easily via a suitable functional martingale central limit theorem. In particular, for $t_0 > 0$ the martingale $\eta_{t_0} \circ X_n$ on $[t_0, \infty)$ has quadratic characteristic

$$\langle \eta_{t_0} \circ X_n \rangle (t) = \int_{t_0}^t \frac{n}{R_n(u)} 1_{(R_n(u)>0)}\mu(u)du, \tag{3.3}$$

converging in probability to $\sigma^2(t_0,t)$ as $n \to \infty$. (To argue this, note that by the law of large numbers, the integrand in (3.3) converges pointwise to

$$\mu(u)/\Pr(V \leq u < U | V < U) = \alpha\mu(u)(G(u)(1 - F(u)))^{-1}.$$

Gill's [2] useful concept of convergence, boundedly in probability, now allows us to change the order of taking limits in probability and integrating).

Assuming that

$$(3.4) \qquad \int_0^\infty \frac{1}{G}\, dF < \infty,$$

Woodroofe [7, Theorem 5] showed that (3.1) holds, even for $t_0 = 0$. Here we shall derive the same result by verifying the condition for tightness close to 0 from Theorem 2.11.

Clearly (2.9) is satisfied for all $n$, while (2.10) holds iff $\lim_{s\downarrow 0} \sigma^2(s,t)$ exists and is finite for some $t > 0$. It is immediately verified that this condition is equivalent to (3.4). So assume that (3.4) holds and let $\epsilon, \eta > 0$ be given. Our aim is to find $\delta$, $n_0$ such that (2.12) holds for $n \geq n_0$. But

$$(3.5) \quad \Pr(\sup_{s,t:0<s\leq t\leq\delta} |X_n(s,t)| > \epsilon) = \lim_{s\downarrow 0} \Pr(\sup_{t:s\leq t\leq\delta} |X_n(s,t)| > \epsilon),$$

and by Doob's inequality applied to the martingale $\eta_s \circ X_n$, the probability on the right is dominated by

$$\frac{1}{\epsilon^2} E(\sup_{t:s\le t\le\delta} X_n^2(s,t)) \le \frac{4}{\epsilon^2} E\, X_n^2(s,\delta)$$

$$= \frac{4}{\epsilon^2} E < \eta_s \circ X_n > (\delta)$$

$$= \frac{4}{\epsilon^2} \int_s^\delta E\left[\frac{n}{R_n(u)} 1_{(R_n(u)>0)}\right]\mu(u)du,$$

cf. (3.3). Now $R_n(u)$ is binomial with probability parameter $p = G(u)(1 - F(u))/\alpha$ and therefore the expectation in the integral above becomes

$$\sum_{k=1}^{n} n \binom{n}{k} \frac{1}{k} p^k(1 - p)^{n-k} \le \frac{2}{p}$$

yielding via (3.5) the estimate

$$\Pr(\sup_{s,t:\ 0<s\le t\le\delta} |X_n(s,t)| > \epsilon) \le \frac{8}{\epsilon^2} \alpha \int_0^\delta \frac{\mu(u)}{G(u)(1-F(u))} du,$$

which is a bound that applies uniformly in $n$ and by (3.4) tends to 0 as $\delta \downarrow 0$. We have established tightness close to 0.

Woodroofe's result (3.1) is local in the sense that it applies only to a restricted time domain. The global formulation (3.2) appears more satisfactory and in principle at least allows one to study the behaviour of estimators of $\mu$, even at timepoints close to 0.

References

[1] Billingsley, P. Convergence of Probability Measures. Wiley, New York, 1968.

[2] Gill, R.D. (1984). A Note on two Papers in Central Limit Theory. Proc. 44th Session ISI, Madrid; Bull. Inst. Internat. Statist. 50, Vol. L, Book 3, 239-243.

[3] Keiding, N. and Gill, R.D. (1988). Random Truncation Models and Markov Processes. To appear.

[4] Lindvall, T. (1973). Weak Convergence of Probability Measures and Random Functions in the Function Space $D[0,\infty]$. J. Appl. Probab. 10, 109-121.

[5] Skorohod, A.V. (1956). Limit Theorems for Stochastic Processes. Theory Probab. Appl. 1, 261-290.

[6] Whitt, W. (1980). Some useful Functions for Functional Limit Theorems. Math. Oper. Res. 5, 67-85.

[7] Woodroofe, M. (1985). Estimating a Distribution Function with Truncated Data. Ann. Statist. 13, 163-177.

PREPRINTS 1988

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33.

No. 1 Jacobsen, Martin: Discrete Exponential Families: Deciding when the Maximum Likelihood Estimator Exists and Is Unique.

No. 2 Johansen, Søren and Juselius, Katarina: Hypothesis Testing for Cointegration Vectors - with an Application to the Demand for Money in Denmark and Finland.

No. 3 Jensen, Søren Tolver, Johansen, Søren and Lauritzen, Steffen L.: An Algorithm for Maximizing a Likelihood Function.

No. 4 Bertelsen, Aksel: On Non-Null Distributions Connected with Testing that a Real Normal Distribution Is Complex.

No. 5 Tjur, Tue: Statistical Tables for Personal Computer Users.

No. 6 Tjur, Tue: A New Upper Bound for the Efficiency of a Block Design.

No. 7 Bunzel, Henning, Høst, Viggo and Johansen, Søren: Some Simple Non-Parametric Tests for Misspecification of Regression Models Using Sign Changes of Residuals.

No. 8 Brøns, Hans and Jensen, Søren Tolver: Maximum Likelihood Estimation in the Negative Binomial Distribution.

No. 9 Andersson, S.A. and Perlman, M.D.: Lattice Models for Conditional Independence in a Multivariate Normal Distribution.

# PREPRINTS 1989

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 1 35 31 33 .

No. 1 Bertelsen, Aksel: Asymptotic Expansion of a Complex Hypergeometric Function.

No. 2 Davidsen, Michael and Jacobsen, Martin: Weak Convergence of Twosided Stochastic Integrals, with an Application to Models for Left Truncated Survival Data.