A new method for proving weak convergence results applied to nonparametric estimators in survival analysis

Jean-Yves Dauxois

CREST-ENSAI, Campus de Ker Lann, 35170-Bruz, France

Received 7 April 1999; received in revised form 16 May 2000; accepted 19 May 2000

Abstract

Using the limit theorem for stochastic integral obtained by Jakubowski et al. (Probab. Theory Related Fields 81 (1989) 111–137), we introduce in this paper a new method for proving weak convergence results of empirical processes by a martingale method which allows discontinuities for the underlying distribution. This is applied to Nelson–Aalen and Kaplan–Meier processes. We also prove that the same conclusion can be drawn for Hjort’s nonparametric Bayes estimators of the cumulative distribution function and cumulative hazard rate. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Stochastic integral; Counting process; Martingale; Weak convergence; Censored data; Product integral; Gaussian process

1. Introduction

In the Martingale framework, the famous Rebolledo’s theorem is usually used to prove weak convergence results on empirical processes. Let us refer to the works of Karr (1991), Fleming and Harrington (1991) or Andersen et al. (1993). For instance, we find there weak convergence results for the processes of Nelson–Aalen (Nelson, 1972; Aalen, 1978) and Kaplan–Meier (Kaplan and Meier, 1958). However, the lifetime distribution is always supposed to be absolutely continuous. Indeed, the theorem of Rebolledo is difficult to use without this hypothesis.

But this may be too restricting. This is particularly true for Hjort’s non-parametric Bayes estimators which, by construction, relates to nonabsolutely continuous lifetime distributions.

Therefore, we introduce a new method for proving weak convergence results. In this order, we use a very interesting limit theorem for stochastic integral obtained by Jakubowski et al. (1989). The discontinuities of the lifetime cumulative distribution function (c.d.f.) is then allowed. We apply it to the estimators of Nelson–Aalen, Kaplan–Meier and Hjort (1990).

E-mail address: dauxois@ensai.fr (J.-Y. Dauxois).
2. Model and Hjort's estimators

Let $X_1,\ldots,X_n$ be independent random variables with common discontinuous c.d.f. $F$. In the sequel, $\tilde{F}$ denotes $1 - F$ and assume that $\tilde{F}(t) > 0$ for all $t \geq 0$. Recall that the cumulative hazard rate $A$ associated with $F$ is defined by

$$ A(t) = \int_0^t \frac{dF(s)}{F(s-)} . $$

Let us suppose that we observe the random censored data

$$ T_i = X_i \wedge C_i, \quad \delta_i = I\{X_i \leq C_i\}, \quad i = 1,\ldots,n, $$

where $I\{A\}$ denotes the indicator function of the set $A$ and $C_1,\ldots,C_n$ are random censoring times with c.d.f. $G_1,\ldots,G_n$, respectively. We suppose furthermore that the random variables $X_1,\ldots,X_n, C_1,\ldots,C_n$ are independent.

Define for all $t \geq 0$

$$ N_n(t) = \sum_{i=1}^n I\{T_i \leq t, \delta_i = 1\} . $$

Note that the counting process $N_n$ is not a simple counting process, since $F$ is discontinuous. Now, consider the number at risk process defined by

$$ Y_n(t) = \sum_{i=1}^n I\{T_i \geq t\}, $$

for all $t \geq 0$ and note $J_n = I\{Y_n(t) > 0\}$. Recall that the product integral $\pi_{[0,t]}f(s)$ of a cadlag function $f$ with finite variations on $\mathbb{R}^+$ is defined by

$$ \pi_{[0,t]}f(s) = \lim_{|\tau| \rightarrow 0} \prod_{\tau}(1 + f(t_i) - f(t_{i-1})) , $$

where $\tau = \{(t_0,\ldots,t_n) \in [0,t]^{n+1} / t_0 = 0 < t_1 < \cdots < t_n = t\}$ is a finite partition of $[0,t]$ and $|\tau|$ is its mesh size (see e.g. Gill and Johansen 1990). Note also that this product integral, always used in the area of survival analysis, is the Doileans–Dade exponential of $f$, well known in stochastic calculus and usually written $\mathcal{E}(f)$.

Recall that the nonparametric estimators of Nelson–Aalen and Kaplan–Meier, respectively are

$$ \hat{A}_n(t) = \int_0^t \frac{J_n(s)dN_n(s)}{Y_n(s)} . $$
and
\[
\hat{F}_n(t) = 1 - \pi_{s \in [0,t]} \left( 1 - \frac{dN_n(s)}{Y_n(s)} \right).
\]

Results concerning these estimators can be found, for instance, in Karr (1991), Fleming and Harrington (1991) or Andersen et al. (1993).

In 1990, Hjort introduced nonparametric Bayes estimators, based on Beta processes (which are special Lévy processes), of the cumulative hazard rate \( A \) and c.d.f. \( F \) defined by
\[
\tilde{A}_n(t) = \int_0^t v(s) dA_0(s) + J_n(s) dN_n(s)
\]
and
\[
\tilde{F}_n(t) = 1 - \pi_{s \in [0,t]} \left( 1 - \frac{v(s)dA_0(s) + dN_n(s)}{v(s) + Y_n(s)} \right),
\]

where \( v \) and \( A_0 \) are the two parameters of the prior Beta process. Here \( A_0 \) is a cumulative hazard with a finite number of jumps such that \( A_0(t) < +\infty \) for all \( t \geq 0 \). It represents the prior guess in Hjort’s Bayesian approach. The function \( v \) is supposed to be piecewise continuous and nonnegative on \([0, +\infty[\). It measures the strength of belief in the prior guess \( A_0 \).

### 3. Martingale structure

Let \( (\tilde{\mathcal{T}}_t)_t \) denote the filtration defined by
\[
\tilde{\mathcal{T}}_t = \sigma\{N_n(u), Y_n(u); 0 \leq u \leq t, i = 1, \ldots, n\}.
\]

Although, \( N_n \) is not a simple counting process, it is easy to check that
\[
M_n(t) = N_n(t) - \int_0^t Y_n(s) dA(s)
\]
is a square-integrable martingale with respect to \( \tilde{\mathcal{T}}_t \). Its predictable variation process \( \langle M_n \rangle \) is given by
\[
\langle M_n \rangle(t) = \int_0^t (1 - \Delta A(s)) Y_n(s) dA(s).
\]

Now, let us introduce a martingale structure associated with Hjort’s estimators. Define
\[
A_n^\mathbb{E}(t) = \int_0^t \frac{v(s) dA_0(s) + J_n(s) Y_n(s) dA(s)}{v(s) + Y_n(s)}.
\]

By (1) and the definition of \( \tilde{A}_n \), we have
\[
\tilde{A}_n(t) - A_n^\mathbb{E}(t) = \int_0^t J_n(s) dM_n(s)
\]
for all \( t \geq 0 \). Since the integrand on the right-hand side is bounded and \( \tilde{\mathcal{T}}_t \)-predictable, \( \tilde{A}_n - A_n^\mathbb{E} \) is a local square-integrable \( \tilde{\mathcal{T}}_t \)-martingale.
We now turn to the cumulative distribution function estimator. Let

\[ \tilde{F}_n^z(t) = \pi_{s \in [0,t]}(1 - d \tilde{A}_n(s)) = \pi_{s \in [0,t]} \left( 1 - \frac{Y_n(s) dA(s) + v(s) dA_0(s)}{v(s) + Y_n(s)} \right). \]

Note that we have \( \tilde{F}_n^z(t) = 1 - \tilde{F}_n(t) = \pi_{s \in [0,t]}(1 - d \tilde{A}_n) \). Thus, by Duhamel’s equation (see Gill and Johansen, 1990) and (3) we get

\[ \frac{\tilde{F}_n(t) - F_n^z(t)}{\tilde{F}_n^z(t)} = \int_0^t \frac{\tilde{F}_n(s^-)}{\tilde{F}_n^z(s)} dM_n(s) \] \( v(s) + Y_n(s) \)

\[ = \int_0^t \frac{\tilde{F}_n(s^-)}{\tilde{F}_n^z(s)} dM_n(s) \] \( v(s) + Y_n(s) \)

Once again the integrand on the right-hand side is locally bounded on \( R^+ \) and \( \tilde{F}_n^z \)-predictable. Thus, \( (\tilde{F}_n - F_n^z)/\tilde{F}_n^z \) is also a local square-integrable martingale.

4. Asymptotic equivalence between the Bayes and frequentist estimators

In order to get this asymptotic equivalence, let us make the following assumption on the censoring times.

Suppose that \( H \) there exists a function \( G \) with support \( R^+ \) such that

\[ \sup_{s > 0} \left| \frac{1}{n} \sum_{i=1}^n G_i(s) - G(s) \right| \to 0 \quad \text{as} \quad n \to +\infty. \]

Using the Glivenko–Cantelli Theorem for independent, but not necessarily identically distributed, random variables (see. e.g. Shorack and Wellner 1986, Theorem 3.2.1) we deduce from assumption \( (H) \) that

\[ \sup_{s > 0} \left| \frac{Y_n(s)}{n} - \tilde{F}_n(s^-) \cdot \tilde{G}_n(s^-) \right| \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to +\infty. \] (5)

Theorem 1. Let assumption \( (H) \) hold and suppose that \( v \) is bounded by \( V < +\infty \). Then, for all \( t \geq 0 \), Hjort’s nonparametric Bayes estimators are a.s. asymptotically equivalent in the Skorohod space \( D[0,t] \) to the frequentist estimators of Nelson–Aalen and Kaplan–Meier, i.e.,

\[ \sup_{s \in [0,t]} \sqrt{n} |\hat{A}_n(s) - \tilde{A}_n(s)| \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to +\infty, \]

\[ \sup_{s \in [0,t]} \sqrt{n} |\hat{F}_n(s) - \tilde{F}_n(s)| \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to +\infty. \]

Proof. Let \( t \geq 0 \). Write

\[ \sqrt{n} |\hat{A}_n(s) - \tilde{A}_n(s)| = \sqrt{n} \int_0^s \frac{v(u)}{Y_n(u)(v(u) + Y_n(u))} J_n(u) \, dN_n(u) \]

\[ - \sqrt{n} \int_0^s \frac{v(u)}{v(u) + Y_n(u)} \, dA_0(u), \]

\[ = \sqrt{n} \int_0^s \frac{v(u)}{Y_n(u)} \, dJ_n(u). \]
for all \( s \leq t \). Thus, since \( Y_n \) is nonincreasing, we have

\[
\sup_{s \in [0,t]} \sqrt{n} |\hat{A}_n(s) - \tilde{A}_n(s)| \leq \sqrt{n} \frac{V}{Y_n(t)} A_0(t) + \sqrt{n} \frac{V}{Y_n^2(t)} N_n(t)
\]

\[
\leq V \frac{n^2}{Y_n^2(t)} \frac{1}{\sqrt{n}}.
\]

Note that the second term on the right-hand side converges to \(1/(\hat{F}(t^-)\hat{G}(t^-))^2\), by (5). Therefore, under the hypotheses on \( F, F_0 \) and \( v \), this implies that the right-hand side of the last inequality converges to zero almost surely and the first assertion of the theorem follows.

Now, by Duhamel's equation (see Gill and Johansen, 1990), we can write

\[
\left| \frac{\hat{F}_n(s) - \tilde{F}_n(s)}{\hat{F}_n(s)} \right| = \left| \frac{\hat{F}_n(s)}{\hat{F}_n(s)} - 1 \right| = \left| \int_0^s \frac{\hat{F}_n(u^-)}{\hat{F}_n(u)} \frac{v(u)}{v(u) + Y_n(u)} \text{d}A_0(u) \right|
\]

\[
\leq \int_0^s \frac{\hat{F}_n(u^-)}{\hat{F}_n(u)} \frac{v(u)}{v(u) + Y_n(u)} \text{d}A_0(u)
\]

\[
+ \int_0^s \frac{\hat{F}_n(u^-)}{\hat{F}_n(u)} \frac{v(u)J_n(u)}{Y_n(u)(v(u) + Y_n(u))} \text{d}N_n(u)
\]

\[
\leq \frac{1}{\hat{F}_n(t)} \frac{V}{Y_n(t)} A_0(t) + \frac{1}{\hat{F}_n(t)} \frac{V}{Y_n^2(t)} N_n(t),
\]

for all \( s \leq t \). Thus,

\[
\sup_{s \in [0,t]} \sqrt{n} |\hat{F}_n(s) - \tilde{F}_n(s)| \leq \sup_{s \in [0,t]} \sqrt{n} \left| \frac{\hat{F}_n(s) - \tilde{F}_n(s)}{\tilde{F}_n(s)} \right| \leq \frac{V}{\hat{F}_n(t)} \frac{n^2}{Y_n^2(t)} \frac{1}{\sqrt{n}}.
\]

It is straightforward to see that the first term on the right-hand side is an \( O_p(1) \), and the rest of the proof runs as before. \( \Box \)

For instance, this theorem proves that the same limiting gaussian martingale will appear with the frequentist and Bayes estimators. Of course, it is natural and expected since the empirical part in Hjort’s estimators swamps the prior one, when \( n \) grows.

But, since trajectories of Beta processes are almost surely discontinuous, Hjort’s estimators relates to prior and posterior distributions which are exclusively concentrated on the set of discontinuous cumulative hazard rate. So, large sample results for Nelson–Aalen and Kaplan–Meier estimators are not completely satisfactory since they are, we believe, only stated with a continuous c.d.f. \( F \). The famous Rebolledo’s theorem is not easy to use when \( F \) is discontinuous. This motivates us to introduce a new method for proving weak convergence results which allows discontinuities for \( F \).

5. Weak convergence results with a possibly discontinuous c.d.f.

**Theorem 2.** Let \( \hat{A}_n \) and \( \tilde{F}_n \) be nonparametric estimators of the cumulative hazard rate and c.d.f., respectively, obtained either empirically (Nelson–Aalen and Kaplan–Meier)
or by Hjort’s bayesian method. Let \( W_1 \) and \( W_2 \) denote the functions, respectively, defined by

\[
W_1(s) = \int_0^s \frac{1 - \Delta A(u)}{\tilde{F}(u^-)\tilde{G}(u^-)} \, dF(u),
\]

\[
W_2(s) = \int_0^s \frac{dF(u)}{\tilde{F}(u^-)\tilde{G}(u^-)(1 - \Delta A(u))}
\]

and \( B \) the Brownian motion on \([0, +\infty[\). Then, under assumption (H), the following convergence results hold in the Skorohod space \( D[0, +\infty[\):

\[
\sqrt{n}(\hat{A}_n - A) \overset{D}{\to} Z_1 = B \circ W_1 \quad \text{as } n \to +\infty,
\]

\[
\sqrt{n} \left( \frac{\hat{F}_n - F}{F} \right) \overset{D}{\to} Z_2 = B \circ W_2 \quad \text{as } n \to +\infty.
\]

In order to prove this theorem, the following lemma is needed.

**Lemma 3.** The (resp. local) square-integrable martingale

\[
\mathcal{M}_n = \frac{M_n}{\sqrt{n}} \quad (\text{resp. } \mathcal{M}'_n = \int_0^t (1 - \Delta A(u))^{-1} d\mathcal{M}_n(u))
\]

converges weakly in \( D[0, +\infty[ \) to a gaussian martingale \( M \) (resp. \( M' \)) with zero mean and covariance function:

\[
\text{Cov}(M(s_1), M(s_2)) = \int_0^{s_1 \wedge s_2} (1 - \Delta A(u))\tilde{F}(u^-)\tilde{G}(u^-) \, dA(u)
\]

\[
\left( \text{resp. } \text{Cov}(M'(s_1), M'(s_2)) = \int_0^{s_1 \wedge s_2} \frac{\tilde{F}(u^-)\tilde{G}(u^-)}{(1 - \Delta A(u))} \, dA(u) \right).
\]

**Proof.** Here we only give the proof for \( \mathcal{M}_n \). The local square-integrable martingale \( \mathcal{M}'_n \) may be handled in much the same way.

Now, it suffices to prove that \( \mathcal{M}_n \) converges weakly in \( D[0, t] \), for all \( t \geq 0 \). Fix \( t \geq 0 \). Since \( \mathcal{M}_n \) is a sum of bounded independent random variables, its finite-dimensional distributions converge weakly, by the central limit theorem.

Tightness will be proved by an application of the theorem VI.5.17 of Jacod and Shiryaev (1987). Condition (i) is trivial. In order to prove (ii), let us note that \( (\mathcal{M}_n) \) is a predictable process. Furthermore,

\[
\sup_{s \in [0, t]} \left| \int_0^s (1 - \Delta A(u))\tilde{F}(u^-)\tilde{G}(u^-) \, dA(u) \right|
\]

\[
= \sup_{s \in [0, t]} \left| \int_0^s (1 - \Delta A(u)) \left( \frac{Y_n(u)}{n} - \tilde{F}(u^-)\tilde{G}(u^-) \right) \, dA(u) \right|
\]

\[
\leq \sup_{s \in [0, t]} \left| \frac{Y_n(s)}{n} - \tilde{F}(s^-)\tilde{G}(s^-) \right| \times \int_0^t (1 - \Delta A(u)) \, dA(u).
\]
From this and (5), it follows that \(\langle \hat{M}_n \rangle\) converges in probability in \(D[0,t]\) to the deterministic process \(\int_0^t (1 - \Delta A(u)) F(u^-) G(u^-) \, dA(u)\), since the last integral is finite by assumption. Condition (C1) of Jacod and Shiryaev (1987) is satisfied, which completes the proof.

**Proof of Theorem 2.** By Theorem 1, we only need to prove the weak convergence results for Nelson–Aalen and Kaplan–Meier estimators. Note that the results may also be obtained directly using the following method (see Dauxois, 1998). The main idea of the proof for each result is to use Proposition 2.9 of Jakubowski et al. (1989). We first look at the cumulative hazard rate estimator.

Let \(t \geq 0\) and \(Z_n\) denote \(\sqrt{n}(\hat{A}_n - A_n^*)\), where \(A_n^* = \int_0^t J_n(u) \, dA(u)\). Note that \(Z_n\) is a locally square-integrable martingale on \(\mathbb{R}^+\) and that we can write

\[
Z_n(s) = n \int_0^s \frac{J_n(u)}{Y_n(u)} \, d\hat{M}_n(u) = \int_0^s H_n(u^-) \, d\hat{M}_n(u),
\]

for all \(s \in [0,t]\) and where \(H_n(u) = n J_n(u^-)/Y_n(u^-)\). From (5), we have the convergence in probability in \(D[0,t]\) of the process \(H_n\) to the deterministic process \(H\) defined by \(H(s) = 1/(\hat{F}(s)\hat{G}(s))\), for all \(s \in [0,t]\).

Since all jump times of \(H_n\) are also jump times of \(\hat{M}_n\), this result and Lemma 3 yield the joint weak convergence of \((H_n, \hat{M}_n)\) to \((H, M)\) as \(n \to +\infty\).

On the other hand, by Lemma 3 and Proposition 3.2 of Jakubowski et al. (1989), the square-integrable martingale \(\hat{M}_n\) is uniformly tight (UT) if

\[
\sup_n \mathbb{E} \left[ \sup_{s \in [0,t]} |\Delta \hat{M}_n(s)| \right] < +\infty.
\]

But, we have

\[
\left( \mathbb{E} \left( \sup_{s \in [0,t]} |\Delta \hat{M}_n(s)| \right) \right)^2 \leq \mathbb{E} \left( \sup_{s \in [0,t]} |\Delta \hat{M}_n(s)| \right)^2 \leq \mathbb{E} \left( \sum_{s \in \mathcal{I}_n} (\Delta \hat{M}_n(s))^2 \right) \leq \mathbb{E}(|\hat{M}_n|)(t) \leq A(t) < +\infty
\]

and the uniformly tightness of \(\hat{M}_n\) is proved. From this and the joint convergence of \((H_n, \hat{M}_n)\), we are in a position to apply Proposition 2.9 of Jakubowski et al. (1989). Thus, we have the weak convergence of \(Z_n\) in \(D[0,t]\) to the process \(Z_1\) defined by \(Z_1(s) = \int_0^s H(u^-) \, dM(u)\).

It is straightforward to see that \(Z_1\) is a mean zero gaussian process with covariance function, for all \(s_1\) and \(s_2\) in \([0,t]\)

\[
\text{cov}(Z_1(s_1), Z_1(s_2)) = \int_0^{s_1 \wedge s_2} \frac{1 - \Delta A(u)}{F(u^-) G(u^-)} \, dA(u) = \int_0^{s_1 \wedge s_2} \frac{1 - \Delta A(u)}{F^2(u^-) G(u^-)} \, dF(u) = W_1(s_1 \wedge s_2).
\]

Now, note that

\[
\sqrt{n}(\hat{A}_n - A) = Z_n + \sqrt{n} \int_0^t (J_n(u) - 1) \, dA(u).
\]
From (5), we see that the set $\Omega_n^d = \{ J_n(s) = 1, \text{ for all } s \in [0,t] \}$ is such that $P(\Omega_n^d) \to 1$ as $n \to +\infty$. Since $\sqrt{n}(\hat{A}_n - A) = Z_n$ on $\Omega_n^d$, the weak convergence of $\sqrt{n}(\hat{A}_n - A)$ is proved.

We can now proceed analogously for the proof of the second weak convergence of Theorem 2. Using once again Duhamel’s equation, we get

$$\sqrt{n} \left( \hat{F}_n(s) - F(s) \right) = \int_0^s H_n'(u^-) d\hat{M}_n'(u) + \sqrt{n} \int_0^s (J_n(u) - 1) \frac{\hat{F}_n(u^-)}{\hat{F}(u)} dA(u),$$

for all $s$ in $[0,t]$ and where $H_n'(u) = n \hat{F}_n(s) J_n(s^+)/\hat{F}(s) Y_n(s^+))$. Here also, we only have to prove that the first term converges to the process

$$Z_2 = \int_0^s H(u^-) dM'(u)$$

in $D[0,t]$. But, it is easily seen that the local square integrable martingale $\hat{M}_n'$ is UT. Furthermore, since $\hat{F}_n$ and $\hat{F}$ are both solutions of stochastic differential equations (recall that $\hat{F}_n = \delta(-\hat{A}_n)$ and $\hat{F} = \delta(-A)$) and since $(\hat{A}_n)$ is an UT sequence which converges to $A$ in probability, we can use the results on stability of stochastic differential equations to prove that $\hat{F}_n/\hat{F}$ converges to 1 in probability in $D[0,t]$ (see Kurtz and Protter, 1991). Thus, $H_n' \to H$ in probability and the end of the proof runs as before.

\[\square\]

Acknowledgements

The author wishes to express his thanks to both Professor Jean MÉMIN for his significant contribution in proving Theorem 2 and also to a referee for his valuable comments which largely improved the final version of this paper.

References