*Sankhyā : The Indian Journal of Statistics*
2011, Volume 73-A, Part 2, pp. 303-328


# Maximum Likelihood Estimator for Cumulative Incidence Functions under Proportionality Constraint

Ségolen Geffray
*Universite de Strasbourg, Strasbourg, France*
Agathe Guilloux
*Université Pierre et Marie Curie - Paris 6, Paris, France*

**Abstract**

This paper deals with a model of possibly dependent competing risks in the presence of additional independent censoring. Under the assumption that the cumulative incidence functions are proportional or equivalently that the cause-specific cumulative hazard functions are proportional, we derive a maximum likelihood estimator for the cumulative incidence functions. Asymptotic results are derived for our estimator namely strong consistency, convergence rate, weak convergence and strong approximation. Pointwise confidence bands are then constructed. Simulation results are carried out to assess the accuracy of the pointwise confidence bands and to investigate the effect of model misspecification. We also briefly consider the case of independent competing risks with proportional net cumulative hazard functions in the presence of independent censoring.

*AMS* (2000) *subject classification.* Primary 62-02,62N01,62N02,62G05,62G20.
*Keywords and phrases.* Competing risks, censoring, proportionality constraint, cumulative incidence functions, asymptotic theory.

## 1 Introduction and Background

This article is concerned with inference on failure times from different competing risks which have proportional cumulative incidence functions. The competing risks are kept as possibly dependent i.e. we do not require any assumption concerning potential dependency. Interest is centered on the cause-specific cumulative incidence functions representing the evolution with time of the probability of failure from a specific risk in the presence of the other competing risks taken into account in the model. In addition, we consider the presence of additional independent non-informative right-censoring. In opposition to Cox's proportional hazards model, no covariate information is accounted for.

For $j = 1, \ldots, \mathcal{J}$, we introduce a non-negative random variable (r.v.) $X_j$ which stands for the failure time of an individual from the $j$-th risk. An important point is that no assumption is made on the competing risks dependency structure, that is, the r.v. $X_j$ for $j = 1, \ldots, \mathcal{J}$ are allowed to be possibly dependent but they don't have to. Let $X = \min(X_1, \ldots, X_{\mathcal{J}})$ be the r.v. which stands for the overall failure time of the individual with distribution function (d.f.) $F$. Let $\mathcal{C}$ indicate the cause of failure meaning that $\mathcal{C}$ takes value $j$ when the failure is due to risk $j$ for a $j$ in $\{1, \ldots, \mathcal{J}\}$ i.e. when $X = X_j$. The cumulative incidence function pertaining to the $j$-th risk is given for $t \geq 0$ by

$$F^{(j)}(t) = \mathbb{P}[X \leq t, \mathcal{C} = j].$$

We assume that $X$ is at risk of being independently right-censored by a non-negative r.v. $C$ with d.f. $G$. Consequently, the observable r.v. consist of

$$\big(T = \min(X, C),\ J = \mathcal{C}\, I(X \leq C)\big)\,. \tag{1.1}$$

Throughout the sequel, we assume that $\mathbb{P}[J = j] \neq 0$ for $j = 1, \ldots, \mathcal{J}$. This entails that $\mathbb{P}[J \neq 0] \neq 0$.

Since $X$ and $C$ are independent, the r.v. $T$ has d.f. $H$ given by the relation $1 - H = (1 - F)(1 - G)$. We assume the proportionality of the different cumulative incidence functions i.e. we assume that, for $j = 1, \ldots, \mathcal{J}$, there exists a $\beta^{(j)} > 0$ such that

$$F^{(j)} = \beta^{(j)}\, F^{(1)}\,. \tag{1.2}$$

Of course, we have $\beta^{(1)} = 1$. This assumption is equivalent to the assumption of proportionality of the cause-specific cumulative hazards which are defined for $j = 1, \ldots, \mathcal{J}$ by

$$\Lambda^{(j)}(t) = \int_0^t \frac{dF^{(j)}}{1 - F^{-}}.$$

Namely, an equivalent assumption to (1.2) is the existence, for $j = 1, \ldots, \mathcal{J}$, of positive constants $\beta^{(j)}$ such that

$$\Lambda^{(j)} = \beta^{(j)}\, \Lambda^{(1)}\,. \tag{1.3}$$

Note that Assumption (1.2) together with the fact that $F = \sum_{j=1}^{\mathcal{J}} F^{(j)}$ entail that

$$F = \left(\sum_{j=1}^{\mathcal{J}} \beta^{(j)}\right) F^{(1)}\,.$$

Consequently, letting

$$\alpha^{(j)} = \frac{\beta^{(j)}}{1 + \beta^{(2)} + \cdots + \beta^{(\mathcal{J})}}\,, \tag{1.4}$$

the cumulative incidence functions $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$ are proportional to the overall d.f. $F$

$$F^{(j)} = \alpha^{(j)} F\,. \tag{1.5}$$

To avoid confusion with Cox's proportional hazards model which involves covariates, this model is referred to as the proportional cumulative incidence functions model. Dauxois and Kirmani (2003) developed an adequacy test for the proportional cumulative incidence functions model but did not propose an estimator tailored for the proportionality situation.

It is possible to see our work as a generalization of the partial Koziol-Green model of informativeness of Gather and Pawlitschko (1998). This generalization is obtained by allowing dependency between the variables which are informatively linked. In addition to widening the scope of applications, an important technical consequence is that interest is not centered on the same functions.

To be specific, let us come back for a moment to the Koziol-Green model which highlighted the concept of informativeness in the setup of independent right-censoring. When inferring from censored failure times, a crucial assumption from the efficiency viewpoint is the non-informativeness of censoring. Let $X_1$ be a non-negative r.v. with d.f. $F_1$ which stands for the failure time of an individual. Let $X_2$ be a non-negative r.v. with d.f. $F_2$ which stands for the censoring time of an individual. Instead of the failure time $X_1$, we observe

$$\big(X = \min(X_1, X_2), \delta = I(X_1 \le X_2)\big) .$$

We assume that $X_1$ and $X_2$ are independent. The r.v. $X$ has d.f. $F$ given by $1-F = (1-F_1)(1-F_2)$. The most commonly used estimator of the lifetime distribution function $F$ is the Kaplan-Meier estimator which is asymptotically efficient when no additional assumptions are made (see e.g. Gill and van der Vaart, 1993). This efficiency is lost when informative censoring is present i.e. when the censoring distribution $F_2$ carries additional information about $F_1$. In the absence of covariates, a simple technique to tackle the informative censoring is the Koziol-Green model. Formally, the Koziol-Green model is defined by the existence of a positive constant $\beta$ such that for $t \ge 0$:

$$\int_0^t \frac{dF_2}{1-F_2^-} = \beta \int_0^t \frac{dF_1}{1-F_1^-} \tag{1.6}$$

where, for any d.f. $F$, the function $F^-$ denotes the left-continuous modification of $F$. This proportional hazards condition is equivalent to the existence of a constant $\beta > 0$ such that $1-F_2 = (1-F_1)^\beta$. In the Koziol-Green model, Cheng and Lin (1987) have pointed out that the maximum likelihood estimator of the lifetime distribution function $F_1$ denoted by $\widetilde{F}_{1,n}$ is of a much simpler form than the usual product-limit estimator $\widehat{F}_{1,n}$. Indeed, the relation $1-F_1 = (1-F)^\alpha$ with $\alpha = 1/(1+\beta) = \mathbb{P}[X_1 \le X_2]$ allows easy calculations. Moreover, the estimator $\widetilde{F}_{1,n}$ is asymptotically more efficient than $\widehat{F}_{1,n}$ when the Koziol-Green assumption is fulfilled. Since this model is easily interpretable and intuitively appealing, it has been extensively studied. We refer to Csörgő and Horváth (1981, 1983), Csörgő (1988), Stute (1992), Dikta (1995) and Ghorai and Schmitter (1999) for an extensive review of the properties of the estimator $\widetilde{F}_{1,n}$.

The Koziol-Green model has been found to be interesting in practice. It turns out that Equation (1.6) holds when the $X$'s are independent of the $\delta$'s (and only in this case). This fact has been used to develop goodness-of-fit tests for the model and several clinical data sets have been shown to fit it fairly well. For instance, Csörgő (1988) has clearly demonstrated the validity of the Koziol-Green model for the Channing House data. We also refer to de Uña Álvarez, González-Manteiga and Cadarso-Suárez (1997) and de Uña Álvarez (1998).

To widen the scope of applications of the Koziol-Green model, Gather and Pawlitschko (1998) proposed a generalization of the Koziol-Green model which they called the partial Koziol-Green model. The partial Koziol-Green model relies on the awareness of different censoring mechanisms and assumes the presence of non-informative censoring in addition to the informative one. For $j = 1, 2$, we introduce a non-negative r.v. $X_j$ with d.f. $F_j$ which stands for the failure time of an individual from the $j$-th cause. We assume here that $X_1$ and $X_2$ are independent. Let $C$ be the non-negative r.v. with d.f. $G$ which stands for the censoring time of an individual. We assume that $C$ is independent from both $X_1$ and $X_2$. We only observe

$$\big(T = \min(X_1, X_2, C), I(\min(X_1, X_2) \leq C), I(X_1 \leq X_2, \min(X_1, X_2) \leq C)\big)$$

where $T$ has d.f. $H$ given by $1-H = (1-F_1)(1-F_2)(1-G)$. The r.v. $\min(X_1, X_2)$ has d.f. $F$ given by $1 - F = (1 - F_1)(1 - F_2)$. Gather and Pawlitschko (1998) assumed the existence of positive constants $\alpha_j$ for $j = 1, 2$ such that

$$1 - F_j = (1 - F)^{\alpha_j}\,.$$

Their assumption is equivalent to assume that the net cumulative hazard functions $\Lambda_j$ for $j = 1, 2$ which are defined for $t \geq 0$ by

$$\Lambda_j(t) = \int_0^t \frac{dF_j}{1 - F_j}$$

are proportional to the overall hazard function $\Lambda$ pertaining to the r.v. $\min(X_1, X_2)$ which is defined for $t \geq 0$ by

$$\Lambda(t) = \int_0^t \frac{dF}{1 - F}\,.$$

Namely, their assumption is equivalent to assume the existence of positive constants $\alpha_j$ for $j = 1, 2$ such that

$$\Lambda_j = \alpha_j \Lambda.$$

In the presence of independent non-informative right-censoring, Gather and Pawlitschko (1998) proposed a semi-parametric estimator of $F_j$ and investigated its asymptotic properties while Geffray and Guilloux (2005) proposed an estimator for $\Lambda_j$ and derived its asymptotic properties. Gather and Pawlitschko (1998) also applied successfully their model to bone marrow transplantation data.

As outlined above, the technical difference between the partial Koziol-Green model and our model of proportional cumulative incidence functions is that Gather and Pawlitschko (1998) focused on the marginal functions $F_j$ while our interest is centered on the cumulative incidence functions $F^{(j)}$. The cumulative incidence functions are much helpful in practice since, in the competing risks framework, the independence of the risks cannot be tested and when the independence cannot be reasonably assumed, the functions $F^{(j)}$ are the only estimable and pertinent quantities. Consequently, they have been much reported, see e.g. Gray (1988), Lin (1997), Fine (1999), Dauxois and Kirmani (2003), Jeong and Fine (2006), Geffray (2008) or Zhang and Fine (2008). Fine and Gray (1999) and Fine (2001) also adapted Cox's proportional hazards model to the cumulative incidence function and proposed inferences for treatment effects and other continuous prognostic factors. Regression modeling on cumulative incidence functions has also been considered in Jeong and Fine (2007) via a parametric approach.

Remark 1.1. It is very interesting to note that our initial model (1.1) of $\mathcal{J}$ competing risks with proportional cumulative incidence functions as in (1.2) can be extended into a cure model with $\mathcal{J}$ competing risks in two different ways. To this aim, we redefine our observable r.v. to account for the fact that there may be cure. We define an additional level for the variable $\mathcal{C}$ so that

$$\mathcal{C} = \begin{cases} j & \text{if the } j\text{-th competing event is observed for } j \in \{1, ..., \mathcal{J}\}, \\ \mathcal{J}+1 & \text{if there is cure.} \end{cases}$$

To give a practical example, one may think of the breast cancer clinical trial setting of Jeong and Fine (2006). They recall that the proportion of recurrences in breast cancer tends to increase for a period of time and then a plateau is reached. A patient can experience two kinds of event as first cancer recurrence, namely locoregional recurrences or distant site recurrences. The patients who do not experience any recurrence can be viewed as a cured population. We now define the variable $\mathcal{C}$ as:

$$\mathcal{C} = \begin{cases} 1 & \text{if the patient experiences a locoregional recurrence as a first event,} \\ 2 & \text{if the patient experiences a distant site recurrence as a first event,} \\ 3 & \text{if the patient never experiences any recurrence, i.e., she is cured.} \end{cases}$$

The competing risks model is well suited to the statistical analysis of first event occurrence with possible cure. The cumulative incidence functions provide meaningful information for the competing events without assumption about the dependence between the risks and the cure.

The first extended cure model can be defined by assuming that the cumulative incidence functions of the competing risks are proportional to the overall distribution function $F$ meaning that, for $j = 1, ..., \mathcal{J}$, there exists a $\alpha^{(j)} > 0$ such that

$$F^{(j)} = \alpha^{(j)} F.$$

Since the competing risks and the cure are mutually exclusive, this entails a proportionality constraint also for the cure part. Indeed, letting $F^{(+)} = \sum_{j=1}^{\mathcal{J}} F^{(j)}$, we have

$$F^{(+)} = \left( \sum_{j=1}^{\mathcal{J}} \alpha^{(j)} \right) F$$

with $0 < \sum_{j=1}^{\mathcal{J}} \alpha^{(j)} < 1$ which implies that

$$F^{(\mathcal{J}+1)} = F - F^{(+)} = \left( 1 - \sum_{j=1}^{\mathcal{J}} \alpha^{(j)} \right) F.$$

Consequently, the first extended cure model can be deduced from our initial model of $\mathcal{J}$ competing risks with proportional cumulative incidence functions just by adding a level to the variable $\mathcal{C}$.

The second extended cure model can be defined by assuming that the cumulative incidence functions of the competing risks are proportional meaning that, for $j = 1, ..., \mathcal{J}$, there exists a $\beta^{(j)} > 0$ such that

$$F^{(j)} = \beta^{(j)} F^{(1)}.$$

This assumption is less restrictive since it does not imply that the cure part satisfies the proportionality condition. Letting

$$F^{(+)} = \sum_{j=1}^{\mathcal{J}} F^{(j)} \text{ and } \alpha^{(j)} = \frac{\beta^{(j)}}{\sum_{k=1}^{\mathcal{J}} \beta^{(k)}},$$

we deduce that

$$F^{(j)} = \alpha^{(j)} F^{(+)}$$

with $\sum_{j=1}^{\mathcal{J}} \alpha^{(j)} = 1$. Consequently, the second extended cure model is slightly different from our initial model of $\mathcal{J}$ competing risks with proportional cumulative incidence functions.

In the set-up of our initial model (1.1) of $\mathcal{J}$ competing risks with independent censoring under the proportionality constraint (1.2), we work on the maximum likelihood estimator $\widetilde{F}_n^{(j)}$ of the cumulative incidence function $F^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. The paper is organized as follows. In Section 2, we derive the maximum likelihood estimator. In Section 3, we expose its main asymptotic properties. Namely, we first show its consistency uniformly on a maximal time-interval. We then present a Law-of-the-Iterated-Logarithm (LIL) type result valid up to the $(n-k_n)$-th order statistic of the sample $(T_1, \ldots, T_n)$. We also establish a strong approximation result for the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathcal{J}$ valid up to the $(n-k_n)$-th order statistic of the sample $(T_1, \ldots, T_n)$. We also deal with the weak convergence of the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathcal{J}$. In Section 4, simulation results are displayed to assess the efficiency of the proposed estimator and to investigate the effect of model misspecification. Comparison with the unconstraint nonparametric estimator of cumulative incidence functions is considered. The preceding weak convergence result enables the construction of pointwise confidence bands which are illustrated on a real dataset in Section 5. Finally, in Section 6, we focus on the case in which the r.v. $X_j$ are independent. Then the marginal d.f. $F_j$ are also of interest. We briefly outline the analogous of the preceding results for an estimator of the functions $F_j$. For the sake of clarity, all the proofs are collected in Section 7.

## 2 Derivation of the Maximum Likelihood Estimator

Let us derive our estimator of the cumulative incidence function $F^{(j)}$. In view of Equation (1.5), a semiparametric estimator for $F^{(j)}$ based on an estimator for $\alpha^{(j)}$ and on an estimator for $F$ would be intuitively appealing. Such an estimator naturally arises when applying the maximum likelihood method to our situation.

Let $(T_i, J_i)_{i=1,\ldots,n}$ be $n$ independent copies of the random vector $(T, J)$ and let $T_{1,n} \leq T_{2,n} \leq \cdots \leq T_{n,n}$ be the order statistics associated with the sample $T_1, \ldots, T_n$. Letting, for $j = 1, ..., \mathcal{J}$ and $i = 1, ..., n$,

$$\delta_i^{(j)} = I(J_i = j) \quad \text{and } \delta_i^{(+)} = \sum_{j=1}^{\mathcal{J}} \delta_i^{(j)},$$

the likelihood based on the i.i.d. sample $(T_i, J_i)_{i=1,\ldots,n}$ can be written as

$$L(\alpha^{(j)}, F) = \prod_{i=1}^{n} \left( \prod_{j=1}^{\mathcal{J}} \left[ \Delta F^{(j)}(T_i)(1 - G^{-}(T_i)) \right]^{\delta_i^{(j)}} \left[ (1 - F(T_i)) \Delta G(T_i) \right]^{1-\delta_i^{(+)}} \right)$$

where $\Delta\varphi(T_i)$ stands for the jump size of any function $\varphi$ at time $T_i$. As $G$ factors out, we can neglect it in the remainder. Then, plugging in Equation (1.5), we get the useful part of the likelihood:

$$L(\alpha^{(j)}, F) = \prod_{i=1}^{n} \left( \prod_{j=1}^{\mathcal{J}} \left[ \alpha^{(j)} \Delta F(T_i) \right]^{\delta_j^{(j)}} \left[ (1 - F(T_i)) \right]^{1-\delta_i^{(+)}} \right) .$$

Hence, the log-likelihood becomes

$$\log(L(\alpha^{(j)}, F)) = \sum_{i=1}^{n} \left( \sum_{j=1}^{\mathcal{J}} \delta_i^{(j)} \log(\alpha^{(j)}) + \delta_i^{(+)} \log(\Delta F(T_i)) + (1 - \delta_i^{(+)}) \log(1 - F(T_i)) \right) .$$

This shows that we can maximize separately over $F$ and $\alpha^{(j)}$. Maximization over $F$ gives the Kaplan-Meier estimator of $F$ denoted by $\widehat{F}_n$ and defined for $t \geq 0$ by

$$\widehat{F}_n(t) = 1 - \prod_{i=1}^{n} \left( 1 - \frac{I(T_i \leq t,\, J_i \neq 0)}{n(1 - H_n^{-}(T_i))} \right) \tag{2.1}$$

with $H_n$ standing for the empirical distribution function given for $t \geq 0$ by

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(T_i \leq t)\,. \tag{2.2}$$

Maximization of

$$\sum_{i=1}^{n}\sum_{j=1}^{\mathcal{J}} \delta_i^{(j)} \log(\alpha^{(j)})$$

over $\{\alpha^{(j)} \geq 0,\ j = 1, ..., \mathcal{J} :\ \sum_{j=1}^{\mathcal{J}} \alpha^{(j)} = 1\}$ yields

$$\widehat{\alpha}^{(j)} = \frac{\sum_{i=1}^{n} \delta_i^{(j)}}{\sum_{i=1}^{n} \delta_i^{(+)}}\,.$$

Consequently, we propose the following estimator for $F^{(j)}$ defined for $t \geq 0$ by:

$$\widetilde{F}_n^{(j)}(t) = \widehat{\alpha}^{(j)} \widehat{F}_n(t)\,.$$

Remark 2.1. It is straightforward to see that the maximum likelihood methodology leads to similar results in the cure model set-up.

In the first extended cure model, we get that, for $j = 1, ..., \mathcal{J}+1$,

$$\widetilde{F}_n^{(j)}(t) = \widehat{\alpha}^{(j)} \widehat{F}_n(t)$$

with

$$\widehat{\alpha}^{(j)} = \frac{\sum_{i=1}^{n} \delta_i^{(j)}}{\sum_{i=1}^{n} \sum_{j=1}^{\mathcal{J}+1} \delta_i^{(j)}}$$

and $\widehat{F}_n(t)$ as in (2.1).

In the second extended cure model, we get that, for $j = 1, ..., \mathcal{J}$,

$$\widetilde{F}_n^{(j)}(t) = \widehat{\alpha}^{(j)} \widehat{F}_n^{(+)}(t)$$

with

$$\widehat{\alpha}^{(j)} = \frac{\sum_{i=1}^{n} \delta_i^{(j)}}{\sum_{i=1}^{n} \sum_{j=1}^{\mathcal{J}} \delta_i^{(j)}}\,,$$

$$\widehat{F}_n^{(+)}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{1-\widehat{F}_n^{-}(T_i)}{1-H_n^{-}(T_i)} \sum_{j=1}^{\mathcal{J}} I(T_i \leq t, J_i = j)$$

with $\widehat{F}_n(t)$ and $H_n(t)$ as in (2.1) and (2.2) respectively.

We also get that

$$\widehat{F}^{(\mathcal{J}+1)} = \frac{1}{n}\sum_{i=1}^{n} \frac{1-\widehat{F}_n^{-}(T_i)}{1-H_n^{-}(T_i)} I(T_i \leq t, J_i = \mathcal{J}+1).$$

## 3 Asymptotic Results

Our first result consists in the strong consistency of the estimator uniformly on a maximal-time interval.

THEOREM 3.1. *Let $\mathcal{I} = \{t : H(t^-) < 1\}$ with $H$ being the d.f. of $T$. If Condition* (1.2) *holds, the following convergence holds almost-surely:*

$$\sup_{t \in \mathcal{I}} \left| \widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right| \overset{n \to \infty}{\longrightarrow} 0 \,.$$

Empirical processes-based inference for the product-limit estimator have drawn considerable attention from many authors. Three main type of results may be distinguished.

Some authors obtained results on a fixed compact interval $[0, \sigma]$ with $\sigma < \tau_H$. For example, Földes and Rejtó (1981b) obtained a LIL-type upper bound on $[0, \sigma]$. Major and Rejtő (1988) obtained a strong approximation result for $\sqrt{n}(\widehat{F}_n - F)$ on $[0, \sigma]$. Such a choice asymptotically excludes a fixed proportion of the data.

Földes and Rejtó (1981a) got a LIL-type result on the whole real line provided $\tau_F < \tau_H$ where $\tau_F = \sup\{t : F(t) < 1\}$ is the right-endpoint of $F$. But such an assumption guarantees that, on the whole support of $F$, there is no uncontrolled increase of the bias due to censoring effect. Csörgő and Horváth (1983) found a rate of convergence of $\widehat{F}_n$ to $F$ which is given on the whole real line without assumptions on the tails. But the bound which is explicitly given in function of $F$ and $G$ may not converge to zero at all. Relying on combinatorial and analytic calculations, Stute and Wang (1993) obtained that, when $F$ and $G$ do not have jumps in common, the estimator $\widehat{F}_n$ is consistent for $F$ on $[0, \tau_H]$ if and only if either $\Delta F(\tau_H) = 0$ or $\Delta F(\tau_H) > 0$ but $G(\tau_H) < 1$. Chen and Lo (1997) obtained a LIL-type result on $[0, \tau_H]$, under assumptions on the tail distributions of $F$ and $G$. Apparently, it is not possible to get neither the strong uniform consistency nor LIL-type results on the whole support $[0, \tau_H]$ without an hypothesis on the tail distributions, hypothesis uncheckable in practice and restrictive.

Stute (1994), followed by Csörgő (1996), Giné and Guillou (1999) turned to a compromise. They got asymptotic results for the Nelson-Aalen and the Kaplan-Meier estimators on increasing intervals that may asymptotically cover each $[0, \sigma]$ for $\sigma < \tau_H$. No assumption on the distribution tails of $F$ and $G$ is needed. These increasing intervals are determined by the data. Indeed, they consists of $[0, T_{n-k_n,n}]$ where $T_{n-k_n,n}$ is the $(n-k_n)$-th order statistic of the sample $T_1, \ldots, T_n$ with $(k_n)$ being a sequence of integers such that $1 \le k_n < n$. If $k_n$ is chosen to be negligible with respect to $n$, then $T_{n-k_n,n}$ converges in probability to $\tau_H$ as $n$ goes to $\infty$. In order to get almost-sure asymptotic results, $(k_n)$ is assumed to fulfill some light growing hypothesis.

In the sequel, we shall denote by hypothesis $(\mathcal{H})$ the following conditions:

1. there exists a constant $M > 0$ such that $k_{2n} \ge M k_n$ for $n$ large enough,

2. the sequence $(k_n/n)$ is non-increasing for $n$ large enough,
3. there exists a constant $C > 0$ such that $k_n \geq C d_n \log n$ with $(d_n)$ is a non-decreasing sequence such that $\sum \frac{1}{k d_{2^k} \log k} < \infty$ (e.g. $d_n = (\log\log\log n)^{1+\varepsilon}$, $d_n = (\log\log\log n)(\log\log\log\log n)^{1+\varepsilon}$, etc).

When Hypothesis $(\mathcal{H})$ is fulfilled, then the conditions required on the sequence $(k_n)$ in Csörgő (1996) and in Giné and Guillou (2001) are also fulfilled.

The following theorem formulates a LIL type result valid up to a certain order statistic.

THEOREM 3.2. *Let $(k_n)$ is a sequence of integers such that $1 \leq k_n < n$ and satisfying Condition $(\mathcal{H})$ for the almost-sure result. If Condition* (1.2) *holds and if $F^{(j)}$ is continuous for $j = 1, \ldots, \mathfrak{J}$, then*

$$\sup_{t \leq T_{n-k_n,n}} \left| \widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right| = \begin{cases} O\left( \left( \frac{\log\log n}{k_n} \right)^{1/2} \right), \\ O_{\mathbb{P}}\left( \frac{1}{(k_n)^{1/2}} \right). \end{cases}$$

REMARK 3.1. Theorems 3.1 and 3.2 also hold in the first and second extended cure models (in the latter case in view of the results of Geffray, 2008).

We now provide a strong approximation result for the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathfrak{J}$. This result holds uniformly on the random increasing intervals $[0, T_{n-k_n,n}]$.

THEOREM 3.3. *Assume that Condition* (1.2) *holds. Let $(k_n)$ be a sequence of integers such that $1 \leq k_n < n$ for all $n$ and fulfilling Hypothesis $(\mathcal{H})$. Assume that $F^{(j)}$ is continuous for $j = 1, \ldots, \mathfrak{J}$. For $n$ large enough, there exists a suitably enlarged probability space on which the following approximation result holds almost-surely, jointly for $j = 1 \ldots, \mathfrak{J}$:*

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}\left( \widetilde{F}_n^{(j)}(t) - F^{(j)}(t) \right) - \widetilde{L}_n^{(j)}(t) \right| = O\left( \sqrt{n} \frac{\log n}{k_n} \right).$$

*For $j = 1, \ldots, \mathfrak{J}$, for fixed $n$, the processes $(\widetilde{L}_n^{(j)})$ are zero-mean Gaussian processes with covariance function given for $k, j = 1, \ldots, \mathfrak{J}$ and for $s, t \geq 0$ by:*

$$\begin{aligned} Cov\left( \widetilde{L}_n^{(j)}(s), \widetilde{L}_n^{(k)}(t) \right) &= F(s)F(t) \frac{\alpha^{(j)} \left( I(k=j) - \alpha^{(k)} \right)}{\mathbb{P}[J \neq 0]} \\ &+ \alpha^{(j)} \alpha^{(k)} (1 - F(s))(1 - F(t)) \int_0^{s \wedge t} \frac{dH^{(1)}}{(1 - H^-)^2} \end{aligned}$$

*where $H^{(1)}(.) = \mathbb{P}[T \leq ., J \neq 0]$.*

We also state the weak convergence of the processes $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})$ jointly for $j = 1, \ldots, \mathfrak{J}$. Here we cannot use the martingale theory, as did Gill (1983) for the Kaplan-Meier estimator, because of $\widehat{\alpha}^{(j)}$ which is neither a martingale nor predictable.

For $j = 1, \ldots, \mathcal{J}$, we denote by $\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})^*$ the process stopped at the $(n-k_n)$-th order statistic. It is given for $t \geq 0$ by:

$$\sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})^*(t) = \sqrt{n}(\widetilde{F}_n^{(j)} - F^{(j)})(t \wedge T_{n-k_n,n}).$$

THEOREM 3.4. *Assume that Condition* (1.2) *holds. Let* $(k_n)$ *be a sequence of integers such that* $1 \leq k_n < n$ *for all* $n$, *fulfilling Hypothesis* ($\mathcal{H}$) *and* $\sqrt{n}\log n/k_n \to 0$ *as* $n \to \infty$. *Assume that* $F^{(j)}$ *is continuous for* $j = 1, \ldots, \mathcal{J}$. *The following convergence holds in the Skorohod space* $D^{\mathcal{J}}[0, \tau_H]$ *of càdlàg functions from* $[0, \tau_H]$ *to* $\mathbb{R}^{\mathcal{J}}$:

$$\left(\sqrt{n}\left(\widetilde{F}_n^{(1)} - F^{(1)}\right)^*, \ldots, \sqrt{n}\left(\widetilde{F}_n^{(\mathcal{J})} - F^{(\mathcal{J})}\right)^*\right) \xrightarrow{\mathcal{D}} \left(\widetilde{K}^{(1)}, \ldots, \widetilde{K}^{(\mathcal{J})}\right),$$

*where the* $\widetilde{K}^{(j)}$ *are mean-zero Gaussian processes with covariance function given for* $s, t > 0$ *and for* $j = 1, \ldots, \mathcal{J}$, *by:*

$$\begin{aligned} Cov\left(\widetilde{K}^{(j)}(t), \widetilde{K}^{(k)}(s)\right) &= F(s)F(t)\,\frac{\alpha^{(j)}\left(I(k=j) - \alpha^{(k)}\right)}{\mathbb{P}[J \neq 0]} \\ &\quad + \alpha^{(j)}\alpha^{(k)}\,(1-F(s))(1-F(t)) \int_0^{s\wedge t} \frac{dH^{(1)}}{(1-H^-)^2}\,. \end{aligned} \tag{3.1}$$

REMARK 3.2. The structure of the covariance function above is very similar to the structure obtained by Gather and Pawlitschko (1998). In addition, under the assumption of continuity of $F$, the term $(1-F(s))(1-F(t)) \int_0^{s\wedge t} dH^{(1)}/(1-H^-)^2$ is exactly the variance function of the limiting process of the Kaplan-Meier estimator. Consequently, the variance is exactly the same as if $\widehat{F}_n$ and $\widehat{\alpha}^{(j)}$ were independent which corresponds to the uncensored case.

REMARK 3.3. If the model given by (1.2) is valid, then we have seen in this paper that the subdistribution function $F^{(j)}$ may be consistently estimated by the estimator $\widetilde{F}_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. This estimator has very similar properties to the purely nonparametric estimator of $F^{(j)}$ derived from Aalen and Johansen (1978) and denoted by $\widehat{F}_n^{(j)}$ for $j = 1, \ldots, \mathcal{J}$. Indeed, the estimator $\widetilde{F}_n^{(j)}$ achieves the same rates of convergence in probability and almost-surely than $\widehat{F}_n^{(j)}$, it may be approximated by a Gaussian process with the same error uniformly on $[0, T_{n-k_n,n}]$ and it converges weakly to a Gaussian process. In addition, the estimator $\widetilde{F}_n^{(j)}$ is asymptotically more efficient than $\widehat{F}_n^{(j)}$ under the model given in (1.3) since, in this case, the asymptotic covariance function of $\widehat{F}_n^{(j)}$ is

$$\begin{aligned} &\lim_{n\to\infty} \text{Cov}\,(\sqrt{n}(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)), \sqrt{n}(\widehat{F}_n^{(j)}(s) - F^{(j)}(s))) = \\ &\alpha^{(j)}\left(1-\alpha^{(j)}\right)\int_0^{s\wedge t} \frac{dF}{(1-G^-)} + (\alpha^{(j)})^2\,(1-F(s))(1-F(t)) \int_0^{s\wedge t} \frac{dH^{(1)}}{(1-H^-)^2}. \end{aligned} \tag{3.2}$$

This asymptotic covariance is greater than that in (3.1) by (7.1) and by the Jensen inequality applied to the first term of (3.2).

Remark 3.4. Theorems 3.3 and 3.4 also hold for the first extended cure model while similar results hold for the second extended cure model with a slightly different covariance structure.

## 4 A Simulation Study

In order to illustrate the small sample behavior of our estimator of the cumulative incidence functions, we conduct a Monte-Carlo study. For $n = 50, 100$ and 200, we compare our estimator $\widetilde{F}_n^{(j)}$ of the cumulative incidence function $F^{(j)}$ to the unconstraint Aalen-Johansen estimator defined for $t \geq 0$ by:

$$\widehat{F}_n^{(j)}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{(1 - \widehat{F}_n^-(T_i))}{(1 - H_n^-(T_i))} I(T_i \leq t, J_i = j).$$

For each replication of the sample, we compute two quantities, namely, the stopped Mean Integrated Squared Error (MISE*) and the stopped normalized Mean Integrated Squared Error (nMISE*) pertaining respectively to $\widetilde{F}_n^{(j)}$ and $\widehat{F}_n^{(j)}$ which are displayed below. Then we compute the mean MISE* and the mean nMISE* over the 1000 sample replications. The number of grid points is set as 1000. The MISE* pertaining respectively to $\widetilde{F}_n^{(j)}$ and $\widehat{F}_n^{(j)}$ are defined for $j = 1, \ldots, \mathcal{J}$ and $t \geq 0$ by:

$$MISE^*(\widetilde{F}_n^{(j)}) = \int_0^\infty \mathbb{E}\left[\left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right)^2 I(t \leq T_{n-k_n,n})\right] dt$$

and

$$MISE^*(\widehat{F}_n^{(j)}) = \int_0^\infty \mathbb{E}\left[\left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right)^2 I(t \leq T_{n-k_n,n})\right] dt.$$

The nMISE* pertaining respectively to $\widetilde{F}_n^{(j)}$ and $\widehat{F}_n^{(j)}$ are defined for $j = 1, \ldots, \mathcal{J}$ and $t \geq 0$ by:

$$nMISE^*(\widetilde{F}_n^{(j)}) = \int_0^\infty \mathbb{E}\left[\frac{1}{T_{n-k_n,n}}\left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right)^2 I(t \leq T_{n-k_n,n})\right] dt$$

and

$$nMISE^*(\widehat{F}_n^{(j)}) = \int_0^\infty \mathbb{E}\left[\frac{1}{T_{n-k_n,n}}\left(\widehat{F}_n^{(j)}(t) - F^{(j)}(t)\right)^2 I(t \leq T_{n-k_n,n})\right] dt.$$

The sequence $(k_n)$ is set as $k_n = \lfloor (\log n)^{5/4} \rfloor$ leading to $k_n = 5$ for $n = 50$, $k_n = 6$ for $n = 100$ and $k_n = 8$ for $n = 200$.

We consider a model of $\mathcal{J} = 2$ competing risks. To generate the data, we use a power-transformation of the absolutely continuous bivariate exponential model of Block and Basu (1974) denoted by ACBVW($\alpha$, $\beta$,$\lambda_0$,$\lambda_1$,$\lambda_2$) involving two positive parameters $\alpha$ and $\beta$ and three non-negative parameters $\lambda_0$, $\lambda_1$ and $\lambda_2$, see Geffray (2006, p. 91) for an explicit expression of the ACBVW model. In this study, we fix $\lambda_1 = 0.4$ and $\lambda_2 = 0.6$ while the dependence parameter $\lambda_0$ is set either to 0 or 0.5. The competing risks are dependent if and only if $\lambda_0 = 0$ and are proportional if and only if $\alpha = \beta$. The more $\alpha$ and $\beta$ are distant from each other, the more the model diverges from the proportionality assumption. The theoretical underlying cause-specific distribution functions of the model used to simulate our data in the different scenarii are represented on Figure 1. The censoring distribution is taken exponential with parameter $\lambda_C$ chosen to obtain a theoretical censoring proportion equal either to 0%, 25% or 50%.

The results concerning the MISE$^*$ are displayed in Table 1 and the results concerning the nMISE$^*$ are displayed in Table 2. All the displayed results are multiplied by $10^3$ for ease of reading.

The main point is that when Assumption (1.2) is satisfied, the estimator $\widetilde{F}_n^{(j)}$ of $F^{(j)}$ is more efficient than the estimator $\widehat{F}_n^{(j)}$ for $j = 1, 2$, even if it does not prevent from the usual variance explosion problem in the tail distribution. We also see that the estimator is robust to little departures from the proportionality assumption. Obviously, the further the model stands from the proportionality assumption, the more biased the estimator gets.

Another point is that both the MISE$^*$ and nMISE$^*$ decrease as $n$ increases illustrating the consistency of the different estimators.

A third technical point is that, for fixed $n$, the MISE$^*$ decreases as the proportion of censored observations increases while the nMISE$^*$ increases as the proportion of censored observations increases. This observation concerning the MISE$^*$ illustrates the fact that the time interval (that goes from 0 to the $(n - k_n)$-th observed percentile) on which the MISE$^*$ is computed gets smaller as the proportion of censored observations increases. This point does not appear with the nMISE$^*$ since the nMISE$^*$ is normalized by the width of the aforementioned time interval. This is explained by the fact that the censoring mechanism favors the observation of small values with the consequence that the $(n - k_n)$-th observed percentile (up to which the MISE$^*$ is computed) gets smaller when the proportion of censored observations increases. Hence the number of data points used to make inference remains the same (for fixed $n$) but the data points are located in a shorter time interval when the proportion of censored observations increases. The fact that the $(n - k_n)$-th observed percentile gets smaller when the proportion of censored observations increases can be shown analytically. For that, we make use of the law of the $k$-th order statistic of the sample $(T_1, \ldots, T_n)$ which may be found e.g. in Capéraà and Van Cutsem (1988). We have for $t \geq 0$:

$$\mathbb{P}[T_{k,n} \leq t] = \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} H(t)^i (1 - H(t))^{n-i}.$$

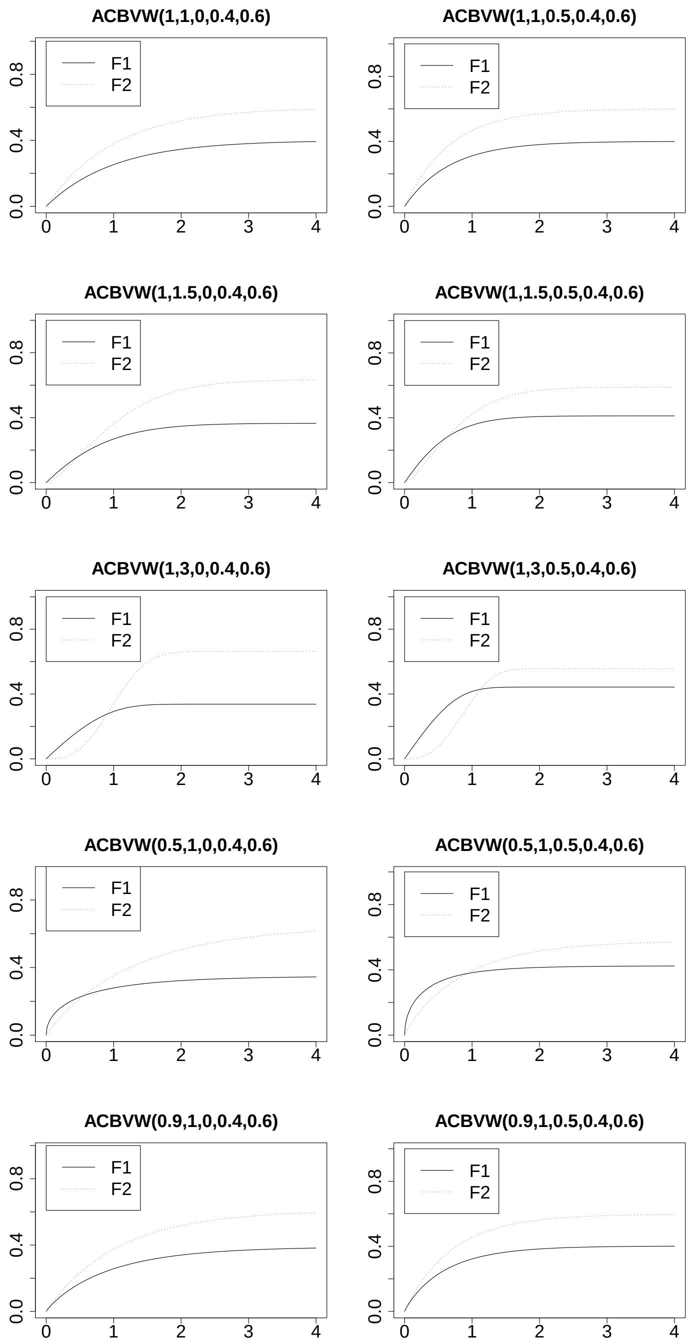


Figure 1: Theoretical underlying cumulative incidence functions of the simulated data.

Table 1: $MISE^*$ multiplied by $10^3$ (CP = censoring proportion)

| | | | | $MISE^*(\widehat{F}_n^{(1)})$ | | | $MISE^*(\widehat{F}_n^{(2)})$ | | | $MISE^*(\widetilde{F}_n^{(1)})$ | | | $MISE^*(\widetilde{F}_n^{(2)})$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $b$ | $\lambda_0$ | CP | $n=$ 50 | $n=$ 100 | $n=$ 200 | $n=$ 50 | $n=$ 100 | $n=$ 200 | $n=$ 50 | $n=$ 100 | $n=$ 200 | $n=$ 50 | $n=$ 100 | $n=$ 200 |
| 1 | 1 | 0 | 0 | 7.47 | 4.87 | 2.98 | 8.85 | 5.87 | 3.41 | 5.38 | 3.89 | 2.44 | 6.98 | 4.61 | 2.83 |
| 1 | 1 | 0 | 0.25 | 5.37 | 3.86 | 2.34 | 6.54 | 4.89 | 3.18 | 4.22 | 3.00 | 1.79 | 5.44 | 4.00 | 2.42 |
| 1 | 1 | 0 | 0.50 | 2.97 | 2.32 | 1.57 | 4.15 | 3.25 | 2.22 | 2.35 | 1.78 | 1.27 | 3.54 | 2.72 | 1.86 |
| 1 | 1 | 0.5 | 0 | 4.86 | 3.86 | 2.31 | 5.66 | 4.81 | 3.25 | 3.53 | 2.71 | 1.61 | 4.61 | 3.29 | 1.91 |
| 1 | 1 | 0.5 | 0.25 | 4.15 | 2.94 | 1.79 | 5.48 | 3.64 | 2.16 | 3.15 | 2.23 | 1.51 | 4.18 | 2.87 | 1.87 |
| 1 | 1 | 0.5 | 0.50 | 3.10 | 2.43 | 1.56 | 4.25 | 3.05 | 1.84 | 2.50 | 2.00 | 1.28 | 3.58 | 2.61 | 1.54 |
| 1 | 3 | 0 | 0 | 6.21 | 3.48 | 2.30 | 6.97 | 3.85 | 2.51 | 10.10 | 8.03 | 5.79 | 10.16 | 7.28 | 5.00 |
| 1 | 3 | 0 | 0.25 | 2.38 | 2.45 | 2.79 | 2.02 | 2.25 | 2.55 | 9.40 | 6.69 | 4.96 | 9.43 | 7.00 | 4.78 |
| 1 | 3 | 0 | 0.50 | 1.26 | 1.37 | 1.45 | 1.11 | 1.26 | 1.47 | 8.68 | 6.33 | 4.36 | 8.78 | 6.47 | 4.19 |
| 1 | 3 | 0.5 | 0 | 4.06 | 4.54 | 4.07 | 2.58 | 2.62 | 2.47 | 11.96 | 9.22 | 6.27 | 10.60 | 7.13 | 4.09 |
| 1 | 3 | 0.5 | 0.25 | 2.38 | 2.70 | 2.71 | 1.55 | 1.84 | 1.86 | 11.19 | 7.60 | 5.05 | 10.27 | 6.90 | 3.85 |
| 1 | 3 | 0.5 | 0.50 | 1.31 | 1.39 | 1.50 | 0.86 | 1.04 | 1.18 | 10.23 | 7.02 | 3.97 | 9.94 | 5.38 | 3.61 |
| 1 | 1.5 | 0 | 0 | 6.21 | 3.40 | 2.30 | 6.97 | 3.85 | 2.51 | 6.51 | 4.73 | 3.53 | 6.57 | 4.88 | 3.97 |
| 1 | 1.5 | 0 | 0.25 | 5.08 | 3.36 | 2.02 | 6.10 | 4.08 | 2.44 | 4.88 | 3.63 | 2.64 | 5.13 | 4.31 | 3.00 |
| 1 | 1.5 | 0 | 0.50 | 4.34 | 2.73 | 1.72 | 5.19 | 3.38 | 2.25 | 3.73 | 2.58 | 1.87 | 4.38 | 3.15 | 2.38 |
| 1 | 1.5 | 0.5 | 0 | 4.79 | 2.87 | 1.64 | 4.45 | 2.86 | 1.65 | 5.55 | 4.19 | 3.53 | 4.99 | 3.94 | 3.38 |
| 1 | 1.5 | 0.5 | 0.25 | 4.18 | 2.57 | 1.70 | 4.14 | 2.65 | 1.65 | 4.21 | 2.92 | 2.38 | 3.88 | 2.94 | 2.24 |
| 1 | 1.5 | 0.5 | 0.50 | 3.29 | 2.19 | 1.51 | 3.13 | 2.39 | 1.58 | 3.05 | 2.11 | 1.62 | 2.65 | 2.19 | 1.56 |
| 0.5 | 1 | 0 | 0 | 9.83 | 6.62 | 3.89 | 11.11 | 7.18 | 4.29 | 13.58 | 12.03 | 10.02 | 13.81 | 11.74 | 10.40 |
| 0.5 | 1 | 0 | 0.25 | 7.17 | 5.19 | 2.82 | 8.35 | 5.60 | 3.78 | 11.21 | 6.68 | 5.00 | 8.86 | 6.91 | 5.68 |
| 0.5 | 1 | 0 | 0.50 | 4.30 | 3.36 | 2.08 | 4.77 | 3.84 | 2.41 | 6.09 | 4.14 | 2.59 | 6.04 | 4.59 | 3.28 |
| 0.5 | 1 | 0.5 | 0 | 6.74 | 4.45 | 2.96 | 5.99 | 4.76 | 2.47 | 12.07 | 10.15 | 7.92 | 9.97 | 9.10 | 8.13 |
| 0.5 | 1 | 0.5 | 0.25 | 4.33 | 3.44 | 2.08 | 4.66 | 3.20 | 2.06 | 6.00 | 4.64 | 3.68 | 5.13 | 4.19 | 3.67 |
| 0.5 | 1 | 0.5 | 0.50 | 2.61 | 2.17 | 1.21 | 2.63 | 2.10 | 1.34 | 2.92 | 2.26 | 1.68 | 2.31 | 2.06 | 1.64 |
| 0.9 | 1 | 0 | 0 | 7.65 | 5.38 | 3.93 | 9.84 | 6.32 | 3.56 | 6.67 | 4.24 | 2.68 | 7.62 | 4.81 | 3.31 |
| 0.9 | 1 | 0 | 0.25 | 6.19 | 4.63 | 2.85 | 7.51 | 5.28 | 3.16 | 4.85 | 3.79 | 2.39 | 6.00 | 4.45 | 2.86 |
| 0.9 | 1 | 0 | 0.50 | 4.02 | 3.06 | 1.93 | 5.20 | 4.09 | 2.62 | 3.27 | 2.44 | 1.62 | 3.11 | 2.83 | 2.28 |
| 0.9 | 1 | 0.5 | 0 | 5.42 | 3.63 | 2.06 | 6.16 | 4.04 | 3.39 | 4.45 | 3.15 | 1.97 | 4.57 | 3.36 | 2.09 |
| 0.9 | 1 | 0.5 | 0.25 | 3.64 | 2.94 | 1.70 | 4.20 | 3.51 | 1.95 | 3.32 | 2.46 | 1.47 | 3.77 | 2.89 | 1.71 |
| 0.9 | 1 | 0.5 | 0.50 | 3.02 | 2.19 | 1.34 | 3.47 | 2.66 | 1.72 | 2.40 | 1.81 | 1.15 | 3.17 | 2.24 | 1.58 |

Table 2: $nMISE^*$ multiplied by $10^3$ (CP = censoring proportion)

| | | | | $nMISE^*(\widehat{F}_n^{(1)})$ | | | $nMISE^*(\widehat{F}_n^{(2)})$ | | | $nMISE^*(\widetilde{F}_n^{(1)})$ | | | $nMISE^*(\widetilde{F}_n^{(2)})$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $b$ | $\lambda_0$ | CP | $n$= 50 | $n$= 100 | $n$= 200 | $n$= 50 | $n$= 100 | $n$= 200 | $n$= 50 | $n$= 100 | $n$= 200 | $n$= 50 | $n$= 100 | $n$= 200 |
| 1 | 1 | 0 | 0 | 3.34 | 1.95 | 0.98 | 3.98 | 2.05 | 1.10 | 2.39 | 1.41 | 0.75 | 3.12 | 1.69 | 0.89 |
| 1 | 1 | 0 | 0.25 | 3.63 | 2.10 | 1.10 | 4.68 | 2.70 | 1.35 | 2.80 | 1.74 | 0.89 | 3.65 | 2.21 | 1.13 |
| 1 | 1 | 0 | 0.50 | 3.28 | 2.09 | 1.22 | 4.48 | 2.85 | 1.70 | 2.58 | 1.61 | 1.04 | 3.73 | 2.27 | 1.47 |
| 1 | 1 | 0.5 | 0 | 3.29 | 1.79 | 0.95 | 4.36 | 2.11 | 1.15 | 2.38 | 1.37 | 0.77 | 3.35 | 1.70 | 0.93 |
| 1 | 1 | 0.5 | 0.25 | 3.51 | 1.86 | 1.08 | 4.37 | 2.39 | 1.29 | 2.59 | 1.49 | 0.86 | 3.57 | 1.96 | 1.11 |
| 1 | 1 | 0.5 | 0.50 | 3.49 | 2.21 | 1.22 | 4.81 | 2.73 | 1.47 | 2.70 | 1.75 | 0.95 | 3.77 | 2.24 | 1.23 |
| 1 | 3 | 0 | 0 | 3.24 | 1.62 | 0.79 | 2.54 | 1.36 | 0.68 | 7.33 | 6.02 | 5.55 | 7.24 | 6.30 | 5.46 |
| 1 | 3 | 0 | 0.25 | 3.30 | 1.79 | 0.89 | 2.82 | 1.66 | 0.87 | 6.18 | 4.99 | 4.16 | 6.06 | 4.93 | 4.32 |
| 1 | 3 | 0 | 0.50 | 3.70 | 2.08 | 1.12 | 3.36 | 1.98 | 1.11 | 5.25 | 3.76 | 3.10 | 5.03 | 3.85 | 3.28 |
| 1 | 3 | 0.5 | 0 | 3.61 | 1.94 | 0.96 | 2.22 | 1.27 | 0.66 | 10.09 | 8.85 | 7.54 | 9.19 | 8.20 | 7.43 |
| 1 | 3 | 0.5 | 0.25 | 4.04 | 2.15 | 1.15 | 2.60 | 1.40 | 0.80 | 8.27 | 6.40 | 5.36 | 6.98 | 5.75 | 5.20 |
| 1 | 3 | 0.5 | 0.50 | 4.62 | 2.56 | 1.34 | 2.59 | 1.72 | 1.03 | 6.57 | 4.55 | 3.51 | 4.68 | 3.79 | 3.27 |
| 1 | 1.5 | 0 | 0 | 3.41 | 1.58 | 0.96 | 3.80 | 1.80 | 1.05 | 3.47 | 2.22 | 1.62 | 3.86 | 2.32 | 1.68 |
| 1 | 1.5 | 0 | 0.25 | 3.35 | 1.78 | 1.01 | 3.99 | 2.20 | 1.22 | 3.15 | 1.95 | 1.35 | 3.66 | 2.36 | 1.51 |
| 1 | 1.5 | 0 | 0.50 | 3.84 | 2.05 | 1.14 | 4.60 | 2.51 | 1.48 | 3.24 | 1.90 | 1.23 | 3.95 | 2.35 | 1.56 |
| 1 | 1.5 | 0.5 | 0 | 3.59 | 1.83 | 0.93 | 3.34 | 1.81 | 0.94 | 4.06 | 2.63 | 1.98 | 3.82 | 2.60 | 1.95 |
| 1 | 1.5 | 0.5 | 0.25 | 3.83 | 1.99 | 1.16 | 3.74 | 2.03 | 1.13 | 3.77 | 2.22 | 1.60 | 3.61 | 2.31 | 1.57 |
| 1 | 1.5 | 0.5 | 0.50 | 4.02 | 2.19 | 1.35 | 3.80 | 2.40 | 1.41 | 3.66 | 2.11 | 1.44 | 3.26 | 2.21 | 1.50 |
| 0.5 | 1 | 0 | 0 | 3.74 | 1.98 | 0.99 | 4.36 | 2.11 | 1.09 | 4.93 | 3.51 | 2.53 | 5.56 | 3.63 | 2.73 |
| 0.5 | 1 | 0 | 0.25 | 3.98 | 2.23 | 1.06 | 4.57 | 2.41 | 1.34 | 4.33 | 2.92 | 1.86 | 4.77 | 3.03 | 2.14 |
| 0.5 | 1 | 0 | 0.50 | 4.06 | 2.52 | 1.32 | 4.38 | 2.81 | 1.56 | 3.92 | 2.65 | 1.81 | 4.14 | 2.97 | 2.07 |
| 0.5 | 1 | 0.5 | 0 | 3.99 | 2.15 | 1.12 | 3.73 | 2.19 | 1.04 | 7.25 | 4.83 | 3.44 | 6.35 | 4.74 | 3.44 |
| 0.5 | 1 | 0.5 | 0.25 | 4.47 | 2.29 | 1.27 | 4.21 | 2.28 | 1.25 | 5.27 | 3.25 | 2.23 | 4.94 | 3.07 | 2.24 |
| 0.5 | 1 | 0.5 | 0.50 | 4.51 | 2.71 | 1.41 | 3.95 | 2.44 | 1.36 | 4.60 | 2.87 | 1.83 | 3.76 | 2.55 | 1.74 |
| 0.9 | 1 | 0 | 0 | 3.74 | 1.87 | 0.91 | 4.25 | 2.13 | 1.07 | 2.82 | 1.48 | 0.80 | 3.32 | 1.71 | 0.96 |
| 0.9 | 1 | 0 | 0.25 | 3.63 | 2.17 | 1.15 | 4.47 | 2.51 | 1.28 | 2.78 | 1.80 | 0.99 | 3.67 | 2.12 | 1.14 |
| 0.9 | 1 | 0 | 0.50 | 3.63 | 2.23 | 1.22 | 4.65 | 2.93 | 1.64 | 2.93 | 1.77 | 1.02 | 3.86 | 2.49 | 1.43 |
| 0.9 | 1 | 0.5 | 0 | 3.78 | 1.96 | 0.94 | 4.18 | 2.15 | 1.09 | 2.93 | 1.68 | 0.89 | 2.28 | 1.81 | 0.99 |
| 0.9 | 1 | 0.5 | 0.25 | 3.72 | 2.11 | 1.06 | 4.14 | 2.46 | 1.22 | 2.91 | 1.74 | 0.91 | 3.41 | 2.07 | 1.07 |
| 0.9 | 1 | 0.5 | 0.50 | 3.70 | 2.19 | 1.22 | 4.70 | 2.68 | 1.53 | 2.93 | 1.80 | 1.02 | 3.80 | 2.25 | 1.34 |

In our set-up, the d.f. $H$ is given by the relation $H = 1-(1-F)(1-G)$ where $G$ is the exponential distribution with parameter $a$. This leads to:

$$\mathbb{P}[T_{k,n} \le t] = \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} \left(1-(1-F(t))(1-G(t))\right)^i (1-F(t))^{n-i}(1-G(t))^{n-i}$$

$$= \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} \left(1-(1-F(t))e^{-at}\right)^{i} (1-F(t))^{n-i} e^{(n-i)at}.$$

Consider a sample $(T_1', \ldots, T_n')$ distributed according to a distribution $1-(1-F)(1-G')$ where $G'$ is an exponential distribution with parameter $a'$ and let $T_{k,n}'$ be the $k$-th order statistic of this sample. >From the above formula, we see that if $a' \geq a$, then we have $\mathbb{P}[T_{k,n} \leq t] \leq \mathbb{P}[T_{k,n}' \leq t]$ for $t \geq 0$.

Remark 4.1. The direct parametrisation approach of Jeong and Fine (2006) for cumulative incidence functions is another way of gaining efficiency over non-parametric estimates. The parametrization of the non-parametric part of our model could help in gaining additional efficiency which can be especially useful for small sample sizes and for long-term event probabilities extrapolation. The main limitation lies in the fact that adjusting a potential parametric model would require goodness-of-fit testing as a necessary preliminary. Unfortunately, up to our knowledge, tests for the fit of parametric models for the cumulative incidence function are not available.

## 5 A Real Data Case

We can use the limiting distribution of $\sqrt{n}\left(\widetilde{F}_n^{(j)} - F^{(j)}\right)$ to construct an asymptotic pointwise confidence band for $F^{(j)}$ via a normal approximation with a plug-in estimation of the limiting variance. Indeed, when Assumption (1.2) is fulfilled, we get for $n$ large enough that $F^{(j)}(t)$ lies between

$$\widetilde{F}_n^{(j)}(t) \pm q_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{\alpha}^{(j)}(1-\widehat{\alpha}^{(j)})\left(\widehat{F}_n(t)\right)^2}{\sum_{i=1}^{n} I(J_i \neq 0)} + \left(\widehat{\alpha}^{(j)}(1-\widehat{F}_n(t))\right)^2 \sum_{i=1}^{n} \frac{I(T_i \leq t, J_i \neq 0)}{n^2(1-H_n^-(T_i))^2}}$$

with probability $(1-\alpha)$ where $q_{1-\alpha/2}$ stands for the $(1-\alpha/2)$-fractile of the standard Gaussian distribution.

We now illustrate these pointwise confidence bands on a real dataset. For that, we use the data and the competing risks model presented as Example I.3.9 in Andersen et al. (1993). In the period 1972-1977, 205 patients with malignant melanoma were operated to remove their skin tumor surgically. All the patients were followed until the end of 1977, that is, it was noted if and when any of these patients died. The survival time is known only for those patients who died before the end of 1977. The rest of the patients are censored at the duration in the study obtained then. Among the deaths observed during the study, 14 are unrelated to melanoma while 57 are directly linked to it. When interest is centered on both event types, we have a competing risks model with two causes of failure and both cumulative incidence functions are then of interest. Before plotting the pointwise confidence bands for these data, we carry out the proportionality test of Dauxois and Kirmani (2003) with the bilateral alternative. This leads to a p-value of 0.653 meaning that the proportionality assumption cannot be rejected. The obtained bands are exposed on

Figure 2. They are plotted on the time interval $[0, 2000]$ to avoid the usual problem of variance explosion.

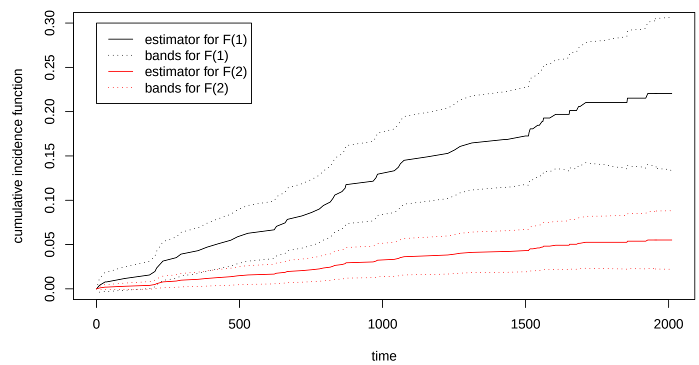


Figure 2: Pointwise confidence bands for the Andersen et al. (1993) data.

## 6 Independent Case

In this section, we assume that the r.v. $X_j$ pertaining to the different causes of failure are mutually independent for $j = 1, \ldots, \mathcal{J}$. As outlined in the introduction, the marginal distribution functions $F_j$ and the net cumulative hazard functions $\Lambda_j$ may be of interest, instead of the cumulative incidence functions $F^{(j)}$ and the crude cumulative hazard functions $\Lambda^{(j)}$. We assume now, for $j = 1, \ldots, \mathcal{J}$, the existence of positive constants $\alpha_j$ such that:

$$1 - F_j = (1 - F)^{\alpha_j} .$$

or equivalently:

$$\Lambda_j = \alpha_j \Lambda. \tag{6.1}$$

We can derive the maximum likelihood estimate of $\Lambda_j$ as follows. The useful part of the likelihood based on the i.i.d. sample $(T_i, J_i)_{i=1,\ldots,n}$ can be written as

$$L(\alpha^{(j)}, \Lambda) = \prod_{i=1}^{n} \left( \prod_{j=1}^{\mathcal{J}} \left[ \alpha_j \Delta\Lambda(T_i) \prod_{k:T_k<T_i} (1 - \Delta\Lambda(T_k)) \right]^{\delta_i^{(j)}} \left[ \prod_{k:T_k \le T_i} (1 - \Delta\Lambda(T_k)) \right]^{1-\delta_i^{(+)}} \right).$$

Maximisation in $\Lambda$ and $\alpha_j$ leads to the following estimator of $\Lambda_j$ defined for $t \ge 0$ by:

$$\Lambda_{j,n}(t) = \widehat{\alpha}_j \Lambda_n(t)$$

where $\Lambda_n$ is the Nelson-Aalen estimator of $\Lambda$ defined for $t \ge 0$ by:

$$\Lambda_n(t) = \sum_{i=1}^{n} \frac{I(T_i \le t, J_i \ne 0)}{n(1 - H_n^-(T_i))}$$

and where

$$\widehat{\alpha}_j = \frac{\sum_{i=1}^n \delta_i^{(j)}}{\sum_{i=1}^n \delta_i^{(+)}} .$$

To estimate the marginal distribution functions $F^{(j)}$, we propose to use Yang's (1977) estimator which is slightly different from that of Gather and Pawlitschko (1998). This estimator is denoted by $\widetilde{F}_{j,n}$ and defined for $t \geq 0$ by:

$$\widetilde{F}_{j,n}(t) = \exp\left(-\Lambda_{j,n}(t)\right).$$

Assume that $F_j$ is continuous for $j = 1, \ldots, \mathfrak{J}$. Then, carrying out a Taylor expansion, we see that the asymptotic behavior of $\widetilde{F}_{j,n} - F_j$ and $-(1-F_j)(\Lambda_{j,n} - \Lambda_j)$ are equivalent. Indeed, we have for $t \geq 0$:

$$\widetilde{F}_{j,n}(t) - F_j(t) = -(1 - F_j(t))\left(\Lambda_{j,n}(t) - \Lambda_j(t)\right) + \frac{1}{2}\left(\Lambda_{j,n}(t) - \Lambda_j(t)\right)^2 \exp\left(-\Lambda^*_{j,n,t}\right)$$

where $\Lambda^*_{j,n,t}$ is between $\Lambda_j(t)$ and $\Lambda_{j,n}(t)$. Consequently, we only need to study the asymptotic behavior of $\Lambda_{j,n} - \Lambda_j$.

It is easily seen that the almost-sure (resp. in probability) rate of convergence of $\Lambda_{j,n}$ to $\Lambda_j$ is the same that the almost-sure (resp. in probability) rate of convergence of $\widetilde{F}_n^{(j)}$ to $F^{(j)}$. On the other hand, by decomposing $\Lambda_{j,n} - \Lambda_j$ into empirical processes plus remainder terms and using Lemma 7.2 (see Section 7), we can get straightforwardly the following result which improves Geffray and Guilloux (2005).

THEOREM 6.1. *Assume that Condition* (6.1) *holds. Let* $(k_n)$ *be a sequence of integers such that* $1 \leq k_n < n$ *for all* $n$ *and fulfilling Hypothesis* $(\mathcal{H})$. *For* $n$ *large enough, there exists a suitably enlarged probability space on which the following approximation result holds almost-surely, jointly for* $j = 1 \ldots, \mathfrak{J}$:

$$\sup_{t \leq T_{n-k_n,n}} \left| \sqrt{n}\left(\Lambda_{j,n}(t) - \Lambda_j(t)\right) - \widetilde{L}_{j,n}(t) \right| = O\left(\sqrt{n}\frac{\log n}{k_n}\right).$$

*For* $j = 1, \ldots, \mathfrak{J}$, *for fixed* $n$, *the processes* $(\widetilde{L}_{j,n})$ *zero-mean independent Gaussian processes with covariance functions given for* $j = 1, \ldots, \mathfrak{J}$ *and for* $s, t \geq 0$ *by:*

$$Cov\left(\widetilde{L}_{j,n}(s), \widetilde{L}_{j,n}(t)\right) = \Lambda_j(s)\Lambda_j(t)\frac{\alpha_j\left(1-\alpha_j\right)}{p} + \alpha_j^2 \int_0^{s\wedge t} \frac{dH^{(1)}}{(1-H^-)^2}. \qquad (6.2)$$

As in the dependent case, the stopped processes $\sqrt{n}(\Lambda_{j,n} - \Lambda_j)^*$ converge weakly in $D[0, \tau_H]$ jointly for $j = 1, \ldots, \mathfrak{J}$. The limiting processes are independent Gaussian processes with the same covariance functions as in (6.2).

## 7 Proofs

PROOF OF THEOREM 3.1. The strong law of large numbers leads to the following almost-sure convergence

$$\widehat{\alpha}^{(j)} = \frac{\frac{1}{n}\sum_{i=1}^{n} I(J_i = j)}{\frac{1}{n}\sum_{i=1}^{n} I(J_i \neq 0)} \overset{n\to\infty}{\longrightarrow} \frac{p^{(j)}}{p}$$

where we let $p^{(j)} = \mathbb{P}[J = j]$ and $p = \mathbb{P}[J \neq 0]$. Some additional work is needed to see why the limit is actually $\alpha^{(j)}$. In view of (1.2), the quantity $\beta^{(j)}$ may be written for $j = 1, \ldots, \mathcal{J}$ as:

$$\beta^{(j)} = \frac{\mathbb{P}[\mathcal{C} = j]}{\mathbb{P}[\mathcal{C} = 1]}.$$

This fact combined with Equality (1.4) entails that:

$$\alpha^{(j)} = \frac{\mathbb{P}[\mathcal{C} = j]}{\sum_{k=1}^{\mathcal{J}} \mathbb{P}[\mathcal{C} = k]} = \mathbb{P}[\mathcal{C} = j]\,.$$

This means that $\alpha^{(j)}$ is the theoretical proportion of observations of $X$ which are due to cause $j$. We then derive another expression for $\alpha^{(j)}$. For that, in view of Equation (1.1) and of the independence between $(X, \mathcal{C})$ and $C$, we can write:

$$\begin{aligned} p^{(j)} &= \mathbb{P}\left[X \leq C, \mathcal{C} = j\right] \\ &= \int_0^\infty (1 - G^-) dF^{(j)} \\ &= \alpha^{(j)} \int_0^\infty (1 - G^-) dF. \end{aligned}$$

On the other hand, we have

$$p = \mathbb{P}\left[X \leq C\right] = \int_0^\infty (1 - G^-) dF. \tag{7.1}$$

Consequently, $\alpha^{(j)}$ can be rewritten:

$$\alpha^{(j)} = \frac{p^{(j)}}{p}.$$

leading to the almost-sure convergence

$$\widehat{\alpha}^{(j)} \overset{n\to\infty}{\longrightarrow} \alpha^{(j)}\,.$$

We carry out the following decomposition for $t > 0$

$$\begin{aligned} \widetilde{F}_n^{(j)}(t) - F^{(j)}(t) = \alpha^{(j)} \left(\widehat{F}_n(t) - F(t)\right) + \left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right) F(t) \\ + \left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right) \left(\widehat{F}_n(t) - F(t)\right). \end{aligned} \tag{7.2}$$

Since Gill (1994) obtained the strong convergence result for the Kaplan-Meier estimator in the following form

$$\sup_{t\in\mathcal{I}}\left|\widehat{F}_n(t)-F(t)\right|\overset{n\to\infty}{\longrightarrow}0\,,$$

this leads to the desired result. □

The proof of Theorem 3.2 requires LIL-type results on the mentioned increasing intervals $[0, T_{n-k_n,n}]$ for the Kaplan-Meier process and for $\widehat{\alpha}^{(j)}$. A LIL-type result has been obtained almost-surely and in probability by Csörgő (1996) and the almost-sure part has been refined later by Giné and Guillou (1999). The result concerning $\widehat{\alpha}^{(j)}$ is exposed below.

LEMMA 7.1. *The following equality holds almost-surely.*

$$\left|\widehat{\alpha}^{(j)}-\alpha^{(j)}\right|=O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right).$$

PROOF. We introduce the empirical counterparts of the probabilities $p$ and $p^{(j)}$ for $j=1,\ldots,\mathfrak{J}$:

$$p_n=\frac{1}{n}\sum_{i=1}^{n}I(J_i\neq 0)\,,$$

and for $j=1,\ldots,\mathfrak{J}$

$$p_n^{(j)}=\frac{1}{n}\sum_{i=1}^{n}I(J_i=j)\,.$$

We write $\widehat{\alpha}^{(j)}-\alpha^{(j)}$ as:

$$\widehat{\alpha}^{(j)}-\alpha^{(j)}=p_n^{(j)}\frac{p-p_n}{p\,p_n}+\frac{p_n^{(j)}-p^{(j)}}{p}\,.$$

Applying the strong law of large number, we get, for $n$ large enough, that the inequalities $1/p_n\leq 2/p$ and $p_n^{(j)}\leq 2\,p^{(j)}$ hold almost-surely. Then, the classical law of the iterated logarithm for i.i.d. r.v. of Hartman and Wintner (1941) entails that we have almost-surely

$$|p-p_n|=O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right)$$

and for $j=1,\ldots,\mathfrak{J}$

$$\left|p^{(j)}-p_n^{(j)}\right|=O\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right).$$

This leads to the desired result. □

PROOF OF THEOREM 3.2. Using the decomposition in (7.2) and applying Theorem 7 of Giné and Guillou (1999) for the almost-sure part and Theorem 1 of Csörgő (1996) for the in-probability part, we obtain straightly the result. □

The following decomposition will be of use both for the weak convergence and strong approximation results.

$$\begin{aligned}
&\widehat{\alpha}^{(j)} - \alpha^{(j)} &(7.3)\\
&= -\,\alpha^{(j)}\,\frac{p_n - p}{p} + \frac{p_n^{(j)} - p^{(j)}}{p} - \left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)\frac{p_n - p}{p}\\
&= -\int \frac{\alpha^{(j)}}{p}\, d\left(H_n^{(1)} - H^{(1)}\right) + \int \frac{1}{p}\, d\left(H_n^{(1,j)} - H^{(1,j)}\right) + O\left(\frac{\log\log n}{n}\right). &(7.4)
\end{aligned}$$

We now state an approximation result valid for the random increasing intervals $[0, T_{n-k_n,n}]$ valid jointly for different empirical processes. For $t$ in $\mathbb{R}$ and for $\mathbf{t} = (t_0, t_1, \ldots, t_{\mathcal{J}+2})$ in $\mathbb{R}^{\mathcal{J}+3}$, we introduce the following empirical processes:

$$\begin{aligned}
E_n(t) &= \sqrt{n}\left(H_n(t) - H(t)\right),\\
E_n^{(0)}(t) &= \sqrt{n}\left(H_n^{(0)}(t) - H^{(0)}(t)\right),\\
E_n^{(1)}(t) &= \sqrt{n}\left(H_n^{(1)}(t) - H^{(1)}(t)\right),\\
E_n^{(1,j)}(t) &= \sqrt{n}\left(H_n^{(1,j)}(t) - H^{(1,j)}(t)\right) \quad \text{for } j = 1, \ldots, \mathcal{J},\\
\boldsymbol{E}_n(\mathbf{t}) &= \left(E_n^{(0)}(t_0), E_n^{(1,1)}(t_1), \ldots, E_n^{(1,\mathcal{J})}(t_{\mathcal{J}}), E_n^{(1)}(t_{\mathcal{J}+1}), E_n(t_{\mathcal{J}+2})\right).
\end{aligned}$$

In addition, we define the norm of a vector $\mathbf{x} = (x_1, \ldots, x_k)$ in $\mathbb{R}^k$ by $\|\mathbf{x}\|_k = \max_{i=1,\ldots,k} |x_i|$.

LEMMA 7.2 (Horvàth, 1980). *On a suitably enlarged probability space, the process $\boldsymbol{E}_n(\mathbf{t})$ may be strongly approximated by a multivariate Gaussian process $\mathbf{B}_n(\mathbf{t})$. Namely,*

$$\sup_{\mathbf{t}\in[0,\tau_H]^{\mathcal{J}+3}} \|\boldsymbol{E}_n(\mathbf{t}) - \mathbf{B}_n(\mathbf{t})\|_{\mathcal{J}+3} = O\left(\frac{\log n}{\sqrt{n}}\right)$$

*where we set for $t \in \mathbb{R}$ and for $\mathbf{t} = (t_0, t_1, \ldots, t_{\mathcal{J}+2}) \in \mathbb{R}^{\mathcal{J}+3}$:*

$$\mathbf{B}_n(\mathbf{t}) = \left(B_n^{(0)}(t_0), B_n^{(1,1)}(t_1), \ldots, B_n^{(1,\mathcal{J})}(t_{\mathcal{J}}), B_n^{(1)}(t_{\mathcal{J}+1}), B_n(t_{\mathcal{J}+2})\right).$$

*The Gaussian processes $(B_n^{(0)})$, $(B_n^{(1,1)})$,..., $(B_n^{(1,\mathcal{J})})$, $(B_n^{(1)})$ and $(B_n)$ have mean 0 and covariance structure defined for $k, j = 1, \ldots, \mathcal{J}$ such that $k \neq j$ and for $s, t \geq 0$ by:*

$$\begin{aligned}
Cov\left(B_n^{(1,j)}(t), B_n^{(1,j)}(s)\right) &= H^{(1,j)}(s \wedge t) - H^{(1,j)}(t)H^{(1,j)}(s),\\
Cov\left(B_n^{(1,j)}(t), B_n^{(1,k)}(s)\right) &= -H^{(1,j)}(t)H^{(1,k)}(s),
\end{aligned}$$

$$\begin{aligned}
Cov\left(B_n^{(1,j)}(t), B_n^{(1)}(s)\right) &= H^{(1,j)}(s\wedge t) - H^{(1,j)}(t)H^{(1)}(s)\,,\\
Cov\left(B_n^{(1,j)}(t), B_n^{(0)}(s)\right) &= -H^{(1,j)}(t)H^{(0)}(s)\,,\\
Cov\left(B_n^{(1,j)}(t), B_n(s)\right) &= H^{(1,j)}(s\wedge t) - H^{(1,j)}(t)H(s)\,,\\
Cov\left(B_n^{(1)}(t), B_n^{(1)}(s)\right) &= H^{(1)}(s\wedge t) - H^{(1)}(t)H^{(1)}(s)\,,\\
Cov\left(B_n^{(0)}(t), B_n^{(0)}(s)\right) &= H^{(0)}(s\wedge t) - H^{(0)}(t)H^{(0)}(s)\,,\\
Cov\left(B_n^{(1)}(t), B_n^{(0)}(s)\right) &= -H^{(1)}(t)H^{(0)}(s)\,,\\
Cov\left(B_n^{(1)}(t), B_n(s)\right) &= H^{(1)}(s\wedge t) - H^{(1)}(t)H(s)\,,\\
Cov\left(B_n(t), B_n(s)\right) &= H(s\wedge t) - H(t)H(s).
\end{aligned}$$

Proof of Theorem 3.3. We write:

$$\begin{aligned}
\sqrt{n}\left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right) = \alpha^{(j)}\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)F(t)\\
+ \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)\left(\widehat{F}_n(t) - F(t)\right).
\end{aligned}$$

In view of Theorem 7 of Giné and Guillou (1999) and Lemma 7.1, we have almost-surely:

$$\sup_{t\le T_{n-k_n,n}}\left|\sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)\left(\widehat{F}_n(t) - F(t)\right)\right| = O\left(\frac{\log\log n}{(k_n)^{1/2}}\right).$$

Consequently, it suffices to consider:

$$I_n^{(j)}(t) = \alpha^{(j)}\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)F(t)\,.$$

The next step consists in linearizing $\widehat{F}_n - F$ by means of Theorem 8 of Giné and Guillou (1999). We get uniformly for $t \ge 0$ that:

$$\begin{aligned}
I_n^{(j)}(t) = \alpha^{(j)}\sqrt{n}(1 - F(t))\left(\Lambda_n(t) - \Lambda(t)\right)\\
+ \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)F(t) + O\left(\sqrt{n}\frac{\log\log n}{k_n}\right).
\end{aligned}$$

We make use of the decomposition of $\widehat{\alpha}^{(j)} - \alpha^{(j)}$ obtained in (7.4) and of the decomposition of $\Lambda_n - \Lambda$ as derived in Proposition 1 of Csörgő (1996). Using the construction of Lemma 7.2, we define the approximating processes for $t \ge 0$ by:

$$II_n^{(j)}(t) = \alpha^{(j)}\left(1 - F(t)\right)\left(\int_0^t \frac{dB_n^{(1)}}{(1 - H^-)} + \int_0^t \frac{B_n^-}{(1 - H^-)^2}dH^{(1)}\right)$$

$$+ F(t)\left(-\int \frac{\alpha^{(j)}}{p}\, dB_n^{(1)} + \int \frac{1}{p}\, dB_n^{(j)}\right) ,$$

where $B_n$, $B_n^{(1)}$ and $B_n^{(1,j)}$ for $j = 1, \dots, \mathcal{J}$ are the correlated Brownian bridges of Lemma 7.2. The construction of Lemma 7.2 guaranties that the approximations hold jointly for $j = 1, \dots, \mathcal{J}$ on a suitably enlarged probability space. □

Proof of Theorem 3.4. We write:

$$\sqrt{n}\left(\widetilde{F}_n^{(j)}(t) - F^{(j)}(t)\right) = \alpha^{(j)}\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right) F(t) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)\left(\widehat{F}_n(t) - F(t)\right).$$

In view of Theorem 7 of Giné and Guillou (1999) and Lemma 7.1, we have almost-surely:

$$\sup_{t \le T_{n-k_n,n}} \left|\sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right)\left(\widehat{F}_n(t) - F(t)\right)\right| = o(1)\,.$$

Consequently, it suffices to consider:

$$I_n^{(j)}(t) = \alpha^{(j)}\sqrt{n}\left(\widehat{F}_n(t) - F(t)\right) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right) F(t)\,.$$

The next step consists in linearizing $\widehat{F}_n - F$ by means of Theorem 8 of Giné and Guillou (1999). We get uniformly in $t$ in $[0, T_{n-k_n,n}]$ that:

$$I_n^{(j)}(t) = \alpha^{(j)}\,(1 - F(t))\sqrt{n}\,(\Lambda_n(t) - \Lambda(t)) + \sqrt{n}\left(\widehat{\alpha}^{(j)} - \alpha^{(j)}\right) F(t) + o(1)\,.$$

Then we make use of the decomposition of $\widehat{\alpha}^{(j)} - \alpha^{(j)}$ presented in (7.4) and of the decomposition of $\Lambda_n - \Lambda$ obtained in Proposition 1 of Csörgő (1996). We get that the limiting process of $I_n^{(j)}(t)$ is the same than the limiting process of

$$II_n^{(j)}(t) = \alpha^{(j)}\,(1 - F(t))\left(\int_0^t \frac{dE_n^{(1)}}{(1 - H^-)} + \int_0^t \frac{E_n^-}{(1 - H^-)^2}\, dH^{(1)}\right) + F(t)\left(-\int \frac{\alpha^{(j)}}{p}\, dE_n^{(1)} + \int \frac{1}{p}\, dE_n^{(j)}\right).$$

The fact that $E_n^{(1)}$ and the $E_n^{(1,j)}$ for $j = 1, \dots, \mathcal{J}$ are subempirical processes of the same empirical process $E_n$ guaranties that the weak convergence of these processes holds jointly in $D^{\mathcal{J}+2}[0, \tau_H]$. It also guaranties that the stopped processes $II_n^{(j)*}$ for $j = 1, \dots, \mathcal{J}$ converge jointly in $D^{\mathcal{J}}[0, \tau_H]$ as $n$ goes to infinity to the limiting processes $II^{(j)}$ defined for $t \ge 0$ and $j = 1, \dots, \mathcal{J}$ by:

$$II^{(j)}(t) = \alpha^{(j)}\,(1 - F(t))\left(\int_0^t \frac{dB^{(1)}}{(1 - H^-)} + \int_0^t \frac{B^-}{(1 - H^-)^2}\, dH^{(1)}\right) + F(t)\left(-\int \frac{\alpha^{(j)}}{p}\, dB^{(1)} + \int \frac{1}{p}\, dB^{(1,j)}\right) ,$$

where $B$, $B^{(1)}$ and $B^{(1,j)}$ for $j = 1, \dots, \mathcal{J}$ are correlated Brownian bridges with the same covariance structure as in Lemma 7.2. □

*Acknowledgement.* The authors would like to thank the Co-editor and the two referees for insightful comments.

## References

AALEN, O. and JOHANSEN, S. (1978), An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.*, **5**, 141–150.

ANDERSEN, P.K., BORGAN, O., GILL, R.D. and KEIDING, N. (1993), *Statistical models based on counting processes.* Springer, New York.

BLOCK, H.W. and BASU, A.P. (1974), A continuous bivariate exponential extension. *J. Amer. Statist. Assoc.*, **69**, 1031–1037.

CAPÉRAÀ, P. and VAN CUSTEM, B. (1988), *Méthodes et modèles en statistique non-paramétrique.* Dunod, Paris.

CHEN, K. and LO, S.-H. (1997), On the rate of uniform convergence of the product-limit estimator: Strong and weak laws. *Ann. Statist.*, **25**, 1050–1087.

CHENG, P.E. and LIN, G.D. (1987). Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Statist. Probab. Letters*, **5**, 75–80.

CSÖRGŐ, S. (1988). Estimation in the proportional hazards model of random censorship. *Statistics*, **19**, 437–463.

CSÖRGŐ, S. (1996). Universal Gaussian approximations under random censorship. *Ann. Statist.*, **24**, 2744–2778.

CSÖRGŐ, S. and HORVÁTH, L. (1981). On the Koziol-Green model for random censorship. *Biometrika*, **68**, 391–401.

CSÖRGŐ, S. and HORVÁTH, L. (1983). On the Koziol-Green model for random censorship II. *Studia Sci. Math. Hungar.*, **18**, 195–203.

DAUXOIS, J.-Y. and KIRMANI, S.N.U.A. (2004). On testing the proportionality of two cumulative incidence functions in a competing risks setup. *J. Nonparametr. Statist.*, **16**, 479–492.

DE UÑA ÁLVAREZ, J., GONZÁLEZ-MANTEIGA, W. and CADARSO-SUÁREZ, C. (1997). Bootstrap selection of the smoothing parameter in density estimation under the Koziol-Green model. In *$L_1$-Statistical Procedures and Related Topics*, (Y. Dodge, ed.), 385–398. IMS Lecture Notes Monogr. Ser., **31**. Institute of Mathematical Statistics, Hayward.

DE UÑA ÁLVAREZ, J. (1998). Inferencia estadística en modelos de censura proporcional. Ph.D. Thesis, Department of Statistics and Operation Research, University of Santiago de Compostela, Spain.

DIKTA, G. (1995). Asymptotic normality under the Koziol-Green model. *Comm. Stat. Theory Meth.*, **24**, 1537–1549.

FINE, J.P. (1999). Analysing Competing Risks Data with Transformation Models. *J. R. Statist. Soc. B*, **61**, 817–830.

FINE, J.P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics*, **2**, 85–97.

FINE, J.P. and GRAY, R.J. (1999). A Proportional Hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.*, **94**, 496–509.

FÖLDES, A. and REJTŐ, L. (1981a). A LIL-type result for the product-limit estimator. *Z. Wahrsch. Verw. Gebiete*, **56**, 75–86.

FÖLDES, A. and REJTÓ, L. (1981b) Strong uniform consistency for nonparametric survival curve estimators from randomly censored data. *Ann. Statist.*, **9**, 122–129.

GATHER, U. and PAWLITSCHKO, J. (1998). Estimating the survival function under a generalized Koziol-Green model with partially informative censoring. *Metrika*, **48**, 189–207.

Geffray, S. (2006). Nonparametric estimation from censored data in a multi-state framework. Ph.D. Thesis, Université Pierre et Marie Curie - Paris VI, France. Available at `http://tel.archives-ouvertes.fr/docs/00/13/82/80/PDF/these-Geffray.pdf`.

Geffray, S. (2009). Strong approximations for dependent competing risks with independent censorship. *TEST*, **18**, 76–95.

Geffray, S. and Guilloux, A. (2005). Estimation dans un modèle de Koziol-Green généralisé, *C. R. Math. Acad. Sci. Paris*, **341**, 49–52.

Ghorai, J.K. and Schmitter, J. (1999). The asymptotic distribution of the suprema of the standardized empirical processes under the Koziol-Green model. *Statist. Probab. Letters*, **41**, 303–313.

Gill, R.D. (1983). Large sample behavior of the product-limit estimator in the whole line. *Ann. Statist.*, **11**, 49–58.

Gill, R.D. (1994). Glivenko-Cantelli for Kaplan-Meier. *Math. Methods Statist.*. **3**, 76–87.

Gill, R.D. and van der Vaart, A.W. (1993). Non- and semi-parametric maximum likelihood estimators and the Von-Mises method II. *Scand. J. Statist.*, **20**, 271–288.

Giné, E. and Guillou, A. (1999). Laws of the iterated logarithm for censored data. *Ann. Probab.*, **27**, 2042–2067.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptative intervals. *Ann. Inst. Henri Poincaré Probab. Stat.*, **37**, 503–522.

Gray, R.J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Statist.*, **16**, 1141–1154.

Hartman, P. and Wintner, A. (1941). On the law of the iterated logarithm. *Amer. J. Math.*, **63**, 169–176.

Horváth, L. (1980). Dropping continuity and independence assumptions in random censorship models. *Studia Sci. Math. Hungar.*, **15**, 381–389.

Fine, J.P. and Jeong, J.-H. (2006). Direct parametric inference for the cumulative incidence function. *Appl. Statist.*, **55**, 187–200.

Fine, J.P. and Jeong, J.-H. (2007). Parametric regression on cumulative incidence function. *Biostatistics*, **8**, 184–196.

Lin, D.Y. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Stat. Med.*, **16**, 901–910.

Major, P. and Rejtő, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.*, **16**, 1113–1132.

Stute, W. (1992). Strong consistency under the Koziol-Green model. *Statist. Probab. Letters*, **14**, 313–320.

Stute, W. (1994). Strong and weak representations of cumulative hazard function and Kaplan-Meier estimators on increasing sets. *J. Statist. Plann. Inference*, **42**, 315–329.

Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.*, **21**, 1591–1607.

Yang, G. (1977). Life expectancy under random censorship. *Stoch. Proc. Appl.*, **6**, 33–39.

Zhang, M.-J. and Fine, J. (2008). Summarizing differences in cumulative incidence functions. *Stoch. Proc. Appl.*, **27**, 4939–4949.

Ségolen Geffray
IRMA, UMR 7501
Université de Strasbourg
7 rue René Descartes
67084 Strasbourg Cedex
France
E-mail: geffray@math.unistra.fr

Agathe Guilloux
LSTA, EA 3124
Université Pierre et Marie Curie - Paris 6
4 place Jussieu
75252 Paris cedex 05
France
E-mail: agathe.guilloux@upmc.fr

Paper received July 2009; revised March 2011.