Independent competing risks versus a general semi-Markov model. Application to heart transplant data.

Catherine Huber [*], O. Pons [†], N. Heutte [‡]

May 5, 2004

Abstract

In the general setting of semi-Markov models, the independent competing risks model may be viewed as a special case where the smallest of potential independent sojourn times leading to different possible final states is observed. Inference for the general case is valid in this special case, but estimators specifically derived for this particular case may be used in order to test whether it takes place or not. The proposed estimators and test are proved to work well in a simulation study. They are also used to analyze the well known heart transplant data.

1 Introduction

When analyzing a cohort of patients, one is often interested not only in their survival time but also in the quality of life they are experiencing, as there may happen toxicity, disability, etc ...(Heutte and Huber (2002)). Usually, there is only a finite number of such states. The process is assumed to be semi-Markov in order to weaken the often too restrictive Markov assumption. The behavior of such a process is defined through the initial probabilities on the set of possible states, and the transition functions defined as the probabilities, starting from any specified state, to reach another state within a certain amount of time. The set of the transition functions may be replaced by two sets. The first one is the set of direct transition probabilities $p_{jj'}$ from any state j to any other state j'. The second one is the set of the sojourn times distributions $F_{|jj'}$ as functions of the actual state j and the state j' reached from there at the end of the sojourn (section 2).

The most usual model in this framework is the so-called competing risks model. This model may be viewed as one where, starting in a specific state j, all states that may be reached directly from j are in competition: the state j' with the smallest random time $W_{jj'}$ to reach it from j will be the one. It

[*]MAP 5, FRE CNRS 2428, UFR Biomédicale, Université René Descartes, et U 472 INSERM, France, e-mail: catherine.huber@univ-paris5.fr.

[†]MAP 5, FRE CNRS 2428, Université René Descartes, Paris 5, and INRA, Jouy-en-Josas, France

[‡]Université de Caen, IUT, 14 000 Lisieux, France

is well known that the joint distribution and the marginal distribution of the latent sojourn times $W_{jj'}$ cannot be identified in a general competing risks model (Tsiatis (1975)). In a semi-Markov model as well as in a competing risks model, only the sub-distribution functions $F_{j'|j} = p_{jj'} F_{|jj'}$ are identifiable and it is always possible to define an independent competing risk (ICR) model by assuming that the variables $W_{jj'}$, $j' = 1, \ldots, m$, are independent with distributions $F_{|jj'} = F_{j'|j}/F_{j'|j}(\infty)$. Without an assumption about their dependence, their joint distribution is not identifiable. Thus a test of an ICR model against an alternative of a general model is not possible. So we shall always assume an ICR (Independent Competing Risk) model.

For a general right-censored semi-Markov process, Lagakos, Sommer and Zelen (1978) proposed a maximum likelihood estimator for the direct transition probabilities and the distribution functions of the sojourn times, under the assumption of a discrete function with a finite number of jumps. In non-parametric models for censored counting processes, Gill (1980), Voelkel and Crowley (1986) considered estimators of the sub-distribution functions $F_{j'|j} = p_{jj'} F_{|jj'}$ and they studied their asymptotic behavior. Here, we consider maximum likelihood estimation for the general semi-parametric model defined by the probabilities $p_{jj'}$ and the hazard functions related to the distribution functions $F_{|jj'}$ (section 4). If the mean number of transitions by an individual tends to infinity, then, the maximum likelihood estimators are asymptotically equivalent to those of the uncensored case. Conversely, when the number of transitions is assumed to be bounded, we define, in section 5, new estimators for a right-censored process. In that case, the effect of the censoring does not become asymptotically negligible.

Under the ICR assumption, specific estimators of the distribution functions $F_{|jj'}$ and of the direct transition probabilities $p_{jj'}$ are deduced from Gill's estimator of the transition functions $F_{j'|j}$. A comparison of those estimators to the estimators for a general semi-Markov process leads to tests for an ICR model against the semi-Markov alternative (section 6). Simulations of the estimators and of the test based on the comparison of the estimated transition probabilities are performed in section 7. The method is used on the Stanford heart transplant data from Kalbfleisch and Prentice (1980).

2 Framework

For each subject i, $i = 1, \cdots, n$, we observe, during a period of time t_i, $K(i) + 1$ successive states $J(i) = (J_0(i), J_1(i), \cdots, J_{K(i)}(i))$, where $J_0(i)$ is the initial state, $J_{K(i)}(i)$ the final state after $K(i)$ transitions. From now on, let i be fixed so that we skip it in the notation.

The total number of possible states is assumed to be finite and equal to m. The successive observed sojourn times are denoted $X = (X_1, X_2, \cdots, X_K)$, where X_k is the sojourn time i spent in state J_{k-1} after $(k-1)$ transitions, and the cumulative sojourn times are $T_k = \Sigma_{\ell=1}^{k} X_\ell$.
One must notice that, if i changes state K times, the sojourn time i spent in his last state J_K is generally

right censored by $t_i - T_K$, where t_i is the total period of observation for subject i. We simplify the rather heavy notation for this last duration, and the last state J_K as

$$X^* \equiv t_i - T_K, \quad J^* \equiv J_K.$$

The subjects are assumed independent and the probability distribution of the sojourn times absolutely continuous. The two models we propose for the process describing the states of the patient are renewal semi-Markov processes. Their behavior is defined through the following quantities:

1. The probabilities of the initial state $\rho = (\rho_1, \rho_2, \cdots, \rho_m)$:

$$\begin{aligned} \rho_j &= P(J_0 = j), \quad j \in \{1, 2, \cdots, m\}, \\ \sum_{j \in \{1,2,\cdots,m\}} \rho_j &= 1. \end{aligned} \tag{1}$$

2. The transition functions $F_{j'|j}(t)$:

$$F_{j'|j}(t) = P(J_k = j', X_k \leq t | J_{k-1} = j) \ , \ j, j' \in \{1, 2, \cdots, m\}. \tag{2}$$

Equivalent to the set of the transition functions $F_{j'|j}$, is the set of the transition probabilities, $p = \{p_{jj'} \ , \ j, j' \in \{1, 2, \cdots, m\}$, together with the set of the distribution functions $F_{|jj'}$ of the sojourn times in each state conditional on the final state as defined below:

1. The direct transition probabilities from a state j to another state j' :

$$p_{jj'} = P(J_k = j' | J_{k-1} = j), \tag{3}$$

2. The law of the sojourn time between two states j and j' defined by its distribution function:

$$F_{|jj'}(t) = P(X_k \leq t | J_{k-1} = j, J_k = j'), \tag{4}$$

$$\text{where} \sum_{j'=1}^{m} p_{jj'} = 1 \ , \ \ p_{jj'} \geq 0 \ , \ \ j, j' \in \{1, 2, \cdots, m\}. \tag{5}$$

The distribution functions $F_{|jj'}$ conditional on states (j, j') do not depend on the value of k, the rank of the state reached by the patient along the process, which is a characteristic of a renewal process. Let us define now the hazard rate conditional on the present state and the next one:

$$\lambda_{|jj'}(t) \ = \ \lim_{dt \longrightarrow 0} \frac{P(t \leq X_k \leq t + dt | X_k \geq t, J_{k-1} = j, J_k = j')}{dt}, \tag{6}$$

as well as the cumulative conditional hazard:

$$\Lambda_{|jj'}(t) \; = \; \int_0^t \lambda_{|jj'}(u)du. \qquad (7)$$

Let W_j be a sojourn time in state j when no censoring is involved, F_j its distribution function, and $\overline{F}_j \equiv 1 - F_j$ its survival function, such that

$$\overline{F}_j(x) \equiv P(W_j > x) = \sum_{j'=1}^{m} p_{jj'}\overline{F}_{|jj'}(x). \qquad (8)$$

The potential sojourn time in state j may be right censored by a random variable C_j having distribution function G_j, density g_j and survival function $\overline{G}_j$. The observed sojourn time in state j is $W_j \wedge C_j$.
A general notation will be $\overline{F}$ for the survival function corresponding to a distribution function F, so that, for example, $\overline{F}_{|jj'} = 1 - F_{|jj'}$ and similarly, for the transition functions, $\overline{F}_{j'|_j} = p_{jj'} - F_{j'|j}$.

3 Independent Competing Risks Model

We assume now that, starting from a state j, the potential sojourn times $W_{jj'}$ until reaching each of the states j' directly reachable from j are independent random variables having distribution functions defined through (4). The final state is the one for which the duration is the smallest. One can thus say that all other durations are right censored by this one. Without restriction of the generality, we assume that the subject is experiencing his k^{th} transition. The competing risks model is defined by

$$\begin{aligned} X_k &= \min_{j'=1,\ldots,m} W_{jj'}, \\ J_k &= j' \text{ such that } W_{jj'} < W_{jj"}, j" \neq j', \end{aligned} \qquad (9)$$

where $W_{jj'}$ has the distribution function $F_{|jj'}$.

In this simple case, independence, both of the subjects and of the potential sojourn times in a given state, allows us to write down the likelihood as a product of factors dealing separately with the time elapsed between two specific states (j, j'). For the Independent Competing Risk model, one derives from (6), (8) and(9) that

$$\begin{aligned} F_{j'|j}(t) &= P(J_k = j', X_k \leq t | J_{k-1} = j) = \int_0^t \{ \prod_{j" \neq j'} \overline{F}_{|jj"}(u) \} \, dF_{|jj'}(u) \qquad (10) \\ &= \int_0^t \lambda_{|jj'}(u) e^{-\sum_{j"} \Lambda_{|jj"}(u)} du. \end{aligned}$$

A consequence is that the direct transition probabilities $p_{jj'}$ defined in (3) may be derived from the probabilities defined in (4),

$$p_{jj'} = P(J_{k+1} = j' | J_k = j) = \int_0^\infty \lambda_{|jj'}(u) e^{-\sum_{j"} \Lambda_{|jj"}(u)} du. \qquad (11)$$

In this special case, the likelihood is fully determined by the initial ρ_j and the functions $\lambda_{|jj'}$ defined in (6).

This problem may be treated as m parallel and independent problems of right censored survival analysis. The only link between them is the derivation of the direct transition probabilities using (11).

4 General Model

The patients are still assumed to be independent, but the potential times for a given subject are no longer assumed to be independent. We model separately the hazard rate and the transition functions ρ_j, $p_{jj'}$ and $\lambda_{|jj'}$ defined as in (1), (3) and (6). The direct transition probabilities $p_{jj'}$ can no longer be derived from the hazard rates.They are now free, except for the constraints (5).
The likelihood may be written as a product of terms each of which implies sojourn times exclusively in one specific state j, $L_n = \prod_{j=1}^{m} L_n(j)$.

For each subject i, and for each $k \in \{1, 2, \cdots, K(i)\}$, we denote $1 - \delta_k(i)$ the censoring indicator of its sojourn time in the k^{th} visited state, $J_{k-1}(i)$, with the convention that $\delta_0(i) \equiv 1$ for every i. If j' is an absorbing state, and if $J_k(i) = j'$, then j' is he last state observed for subject i, $k \equiv K(i)$, and we denote it $X^*(i) = 0$ and $\delta_{K(i)+1}(i) = 1$.
Another convention is that subject i is censored, when the last visited state $J^*(i)$ is not absorbing and the sojourn time in this state $X^*(i)$ is strictly positive and we denote $1 - \delta_i$ the censoring indicator. In all other cases, in particular if the last visited state is absorbing or if the sojourn time there is equal to 0, we say that the subject is not censored and we thus have $\delta_i = 1$. We can then write

$$\delta_k(i) = \prod_{k'=1}^{k} \delta_{k'}(i), \quad \delta_i = 1\{X^*(i) = 0\}.$$

For each state j of $\{1, 2, \cdots, m\}$, we define the following counts where k varies, for each subject i, between 1 and $K(i)$, $i \in \{1, 2, \cdots, n\}$, and $x \geq 0$,

$$\begin{aligned} N_{i,k}(x, j, j') &= 1\{J_{k-1}(i) = j, J_k(i) = j'\}1\{X_k(i) \leq x\}, \\ Y_{i,k}(x, j, j') &= 1\{J_{k-1}(i) = j, J_k(i) = j'\}1\{X_k(i) \geq x\}, \end{aligned} \tag{12}$$

$$N_i^c(x, j) = (1 - \delta_i)1\{J^*(i) = j\}1\{X^*(i) \leq x\},$$

$$Y_i^c(x, j) = (1 - \delta_i)1\{J^*(i) = j\}1\{X^*(i) \geq x\}.$$

By summation of the counts thus defined on the indices j', i, or k, we get

$$\begin{aligned} N(x,j,j',n) &= \sum_{i=1}^{n}\sum_{k=1}^{K(i)} N_{i,k}(x,j,j'), \\ N^{nc}(x,j) &= \sum_{j'=1}^{m} N(x,j,j',n), \\ N(x,j,n) &= \sum_{i=1}^{n} N_i^{nc}(x,j) + N^{nc}(x,j), \\ Y^{nc}(x,j,j',n) &= \sum_{i=1}^{n}\sum_{k=1}^{K(i)} Y_{i,k}(x,j,j'), \\ Y(x,j,n) &= \sum_{j'=1}^{m} Y^{nc}(x,j,j',n) + \sum_{i=1}^{n} Y_i^c(x,j). \end{aligned} \tag{13}$$

By taking for x the limiting value ∞ we define $N_{i,k}(j,j') = N_{i,k}(\infty,j,j')$, $N_i^c(j) = N_i^c(\infty,j)$, $N(j,j',n) = N(\infty,j,j',n)$, $N^{nc}(j,n) = N^{nc}(\infty,j,n)$, so that $N(j,j',n)$ is the number of direct transitions from j to j' that are fully observed,$N(j,n)$ is the number of sojourn times in state j, whose $N^{nc}(j,n)$ (nc for not censored) are fully observed and $N^c(j,n)$ (c for censored) are censored. For $x = 0$, we denote $Y_i^c(j) = Y_i^c(0,j)$. The number of individuals initially in state j is $N^0(j,n) = \sum_{i=1}^{n} 1\{J_0(i) = j\}$.

The true parameter values are denoted ρ_j^0 and $p_{jj'}^0$, and the true functions of the model are $\overline{F}^0_{j'|j}$, $\overline{F}^0_{|jj'}$, $\overline{F}^0_j$, $\overline{G}^0_j$ and $\Lambda^0_{|jj'}$.

Let $l_n = \log(L_n)$ and $l_n(j) = \log(L_n(j))$. The log-likelihood relative to state j is proportional to

$$\begin{aligned} l_n(j) &= \rho_j N^0(j,n) + \sum_{j'=1}^{m} N(j,j',n)\log(p_{jj'}) \\ &\quad + \sum_{i=1}^{n}\sum_{k=1}^{K(i)}\sum_{j'=1}^{m} N_{i,k}(j,j')[\log(\lambda_{|jj'}(X_k(i))) - \Lambda_{|jj'}(X_k(i))] \\ &\quad + \sum_{i=1}^{n} N_i^c(j)[\log\{\sum_{j'=1}^{m} p_{jj'}e^{-\Lambda_{|jj'}(X^*(i))}\}] \\ &= l_n^0(j) + l_n^{nc}(j) + l_n^c(j), \end{aligned} \tag{14}$$

Among the sum of four terms giving (14), let l_n^0 be the first term relative to the initial state, l_n^{nc} (nc for non censored) the sum of the second and third terms, which involve exclusively fully observed sojourn times in state j, and finally l_n^c (c for censored) the last term which deals with censored sojourn times in state j.

We denote $K_n = \max_{i=1,2\cdots,n} K(i)$ and $n\overline{K}_n = \sum_{i=1}^{n} K(i)$ respectively the maximum number of transitions and the total number of transitions for the n subjects. We consider two different designs of experiments, whether or not observations are stopped after a fixed amount K of direct transitions.

It is obvious that if the densities f_j of the sojourn times, without censoring, for every state j, are strictly positive on $]0;t_0[$ for some $t_0 > 0$, and if the distribution functions G_j of the censoring times are such that $G_j(t) < 1$ for all $t > 0$, the maximal number $K_n = \max_i K(i)$ of transitions experienced by a subject tends to infinity when n grows. If moreover the mean number of transitions $\overline{K}_n$ goes also to infinity, then the term relative to censored times $l_n^c(j)$ is the sum of terms of order n while the term $l_n^{nc}(j)$ is a sum of terms of order $n\overline{K}_n$. Therefore we have

Proposition 1 *Under the assumptions $\overline{K}_n \to \infty$, and*

$$\frac{N^{nc}(j,n)}{n\overline{K}_n} \quad \longrightarrow \quad q_j^0 > 0, \quad j \in \{1, 2, \cdots, m\},$$

then

$$\lim_{n \longrightarrow \infty} \frac{l_n(j)}{n\overline{K}_n} \quad = \quad \lim_{n \longrightarrow \infty} \frac{l_n^{nc}(j)}{n\overline{K}_n}.$$

and the maximum likelihood estimators are asymptotically equivalent to

$$\begin{aligned} \widehat{p}_{jj'} &= \frac{N(j,j',n)}{N^{nc}(j,n)}, \\ \widehat{\Lambda}_{|jj'}(x) &= \int_0^x \frac{dN(s,j,j',n)}{Y^{nc}(s,j,j',n)}, \\ \widehat{\overline{F}}_{|jj'}(x) &= \prod_{s \leq x} \Big\{ 1 - \frac{dN(s,j,j')}{Y^{nc}(s,j,j',n)} \Big\}. \end{aligned}$$

5 Case of a bounded number of transitions

We now assume that the number of transitions is bounded by a finite number K fixed in advance. For each subject $i = 1, \cdots, n$, the observation ends at time $t_i = \sum_{k=1}^{K(i)} X_k(i)$ if $K(i) = K$ or if $J_{K(i)}$ is an absorbing state, and at time t_i where there is a right censoring in the $K(i)^{\text{th}}$ visited state, $K(i) < K$.

Using notations in (13), the likelihood term relative to the initial state j may be written

$$l_n^0(j) \quad = \quad N^0(j,n)\log(\rho_j),$$

the terms relative to the fully observed sojourn times in state j is

$$\begin{aligned} l_n^{nc}(j) \quad = \quad & \sum_{j'=1}^m \Big\{ \; N(j,j',n)\log(p_{jj'}) \\ & + \sum_{i=1}^n \sum_{k=1}^K N_{i,k}(j,j')[\log(\lambda_{|jj'}(X_k(i))) - \Lambda_{|jj'}(X_k(i))] \; \Big\}, \end{aligned}$$

and the term relative to the censored sojourn times in state j is

$$l_n^c(j) \quad = \quad \sum_{i=1}^n N_i^c(j)[\log\{\sum_{j'=1}^m p_{jj'} e^{-\Lambda_{|jj'}(X^*(i))}\}].$$

The score equations for $p_{jj'}$ and $\Lambda_{jj'}$ do not lead to explicit solutions because they involve the survival function $\overline{F}_j$ and the transition function $\overline{F}_{j'|j}$. We define estimators $\widehat{p}_{n,jj'}$ and $\widehat{\Lambda}_{n,|jj'}$ by plugging in the score equations the Kaplan-Meier estimator of $\overline{F}_j$ and the estimator of $F_{j'|j}$ given by Gill (1980),

$$\begin{aligned}\widehat{\overline{F}}_{n,j}(x) &= \prod_{y\le x}\left\{1-\frac{dN(y,j,n)}{Y(y,j,n)}\right\}, &(15)\\ \widehat{F}_{n,j'|j}(x) &= \int_0^x \widehat{\overline{F}}_{n,j}(y^-)\frac{dN(y,j,j',n)}{Y(y,j,n)}. &(16)\end{aligned}$$

We obtain the estimators

$$\begin{aligned}\widehat{\rho}_{n,j} &= \frac{N^0(j,n)}{n},\\ \widehat{p}_{n,jj'} &= \frac{N(j,j',n)+\widehat{N}^c(j,j',n)}{N^{nc}(j,n)+N^c(j,n)}, &(17)\\ \widehat{\Lambda}_{n,|jj'}(x) &= \int_0^x \frac{dN(y,j,j',n)}{Y^{nc}(y,j,j',n)+\widehat{Y}^c(y,j,j',n)},\end{aligned}$$

with

$$\begin{aligned}\widehat{Y}^c(y,j,j',n) &= \sum_{i=1}^n Y_i^c(y,j)\frac{\widehat{\overline{F}}_{n,j'|j}(X^*(i))}{\widehat{\overline{F}}_{n,j}(X^*(i))},\\ \widehat{N}^c(j,j',n) &= \sum_{i=1}^n N_i^c(j)\frac{\widehat{\overline{F}}_{n,j'|j}(X^*(i))}{\widehat{\overline{F}}_{n,j}(X^*(i))}.\end{aligned}$$

The variable $(n^{1/2}(\widehat{p}_{n,jj'}-p^0_{jj'}))_{j'}$ and the process $\{n^{1/2}(\widehat{\Lambda}_{n,|jj'}-\Lambda^0_{|jj'}))_{j'}$ are asymptotically Gaussian, on every interval $[0,\tau]$ such that $\int_0^\tau(\overline{F}^0_{j'|j}\overline{G}^0_j)^{-1}\,d\Lambda^0_{j'|j}<\infty$ (Pons (2003)).

6 A Test of the Hypothesis of Independent Competing Risks.

In the ICR case, the initial probabilities jointly with the survival functions $\overline{F}_{|jj'}$ of the sojourn times conditional on states on both ends, are sufficient to determine completely the law of the process. In the general case, however, the two sets of parameters $p_{jj'}$ and $\overline{F}_{|jj'}$ are independent and may be modeled separately. Our aim is to derive a test of the hypothesis of Independent Competing Risks (ICR):

$$\begin{aligned}H_0 &: \text{The process is ICR}\\ H_1 &: \text{The process is not ICR}\end{aligned}$$

The Kaplan-Meier estimator $\widehat{\overline{F}}_{n,j}$ of $\bar{F}_j$, given in (15), and the estimator $\widehat{F}_{n,j'|j}$ of $F_{j'|j}$, given in (16), are consistent and asymptotically Gaussian both under H_0 and under H_1. It is also true for the straightforward estimator $\widehat{\rho}_{n,j}$ of the initial probabilities. From those estimators, one may derive general estimators of the transition probability $p_{jj'}$ and of the survival function $\overline{F}_{|jj'}$ of the time elapsed between two successive jumps in states j and j'. For these estimators, we shall use the same notations as the estimators of $p_{jj'}$ and $\overline{F}_{|jj'}$ defined in section 5, though they are now given by

$$\widehat{p}_{n,jj'} = \max_t \widehat{F}_{n,j'|j}(t) \tag{18}$$

$$\widehat{\overline{F}}_{n,|jj'}(t) = 1 - \frac{\widehat{F}_{n,j'|j}(t)}{\widehat{p}_{n,jj'}}. \tag{19}$$

In the independent competing risk model, the transition probability $F_{j'|j}$ satisfies (10) and thus may be estimated as

$$\begin{aligned} \widehat{F}^{RC}_{n,j'|j}(t) &= -\int_0^t \prod_{j"\neq j'} \widehat{\overline{F}}_{n,|jj"}(s)\ \ d\widehat{\overline{F}}_{n,|jj'}(s) \\ &= \frac{1}{\prod_{j"} \widehat{p}_{n,jj"}} \int_0^t \prod_{j"\neq j'} \widehat{\overline{F}}_{n,j"|j}(s)\ \widehat{\overline{F}}_{n,j}(s^-)\ d\widehat{\Lambda}_{n,j'|j}(s), \end{aligned} \tag{20}$$

where

$$\widehat{\Lambda}_{n,j'|j}(t) = \int_0^t 1\{Y(s,j,n) > 0\} \frac{dN(s,j,j',n)}{Y(s,j,n)} \tag{21}$$

is the estimator of the cumulative hazard function $\Lambda_{n,j'|j}$ in the general model. A competitor to $\widehat{p}_{n,jj'}$ is deduced as

$$\widehat{p}^{RC}_{n,jj'} = \max_t \widehat{F}^{RC}_{n,j'|j}(t). \tag{22}$$

Proposition 2 *If $p^0_{jj'} > 0$, $\sqrt{n}(\widehat{p}_{n,jj'} - p^0_{jj'})$ is asymptotically distributed as a normal random vector with mean 0, variances and covariances*

$$\begin{aligned} \sigma^2_{jj'} &= \frac{1}{\pi^0_j} \int_0^\infty \frac{1}{\overline{G}^0_j(s)\overline{F}^0_j(s)} \left\{ (\overline{F}^0_{j'|j}(s) - p^0_{j'|j})^2 \frac{dF^0_j(s)}{\overline{F}^0_j(s)} + \{\overline{F}^0_j(s) + 2(\overline{F}^0_{j'|j}(s) - p^0_{j'|j})\}\, d\overline{F}^0_{j'|j}(s) \right\}, \\ \sigma^2_{jj'j"} &= \frac{1}{\pi^0_j} \int_0^\infty \frac{1}{\overline{G}^0_j(s)\overline{F}^0_j(s)} \left\{ (\overline{F}^0_{j'|j}(s) - p^0_{j'|j})(\overline{F}^0_{j"|j}(s) - p^0_{j"|j}) \frac{dF^0_j(s)}{\overline{F}^0_j(s)} \right. \\ &\quad \left. +(\overline{F}^0_{j'|j}(s) - p^0_{j'|j})\}\, d\overline{F}^0_{j"|j}(s) + (\overline{F}^0_{j"|j}(s) - p^0_{j"|j})\}\, d\overline{F}^0_{j'|j}(s) \right\}. \end{aligned}$$

Moreover, $\sqrt{n}(\widehat{p}^{RC}_{n,jj'} - p^0_{jj'})$ is asymptotically distributed as a centered Gaussian variable.

Estimators of the asymptotic variance and covariances of $(\widehat{p}_{n,jj'})_{j'\in J(j)}$ may be obtained by replacing the functions $\overline{F}^0_j$, $F^0_{j'|j}$ and $\Lambda^0_{j'|j}$ by their estimators in the general model, (15), (16) and (21). Due to their intricate formulas, it seems difficult to use an empirical estimator of the asymptotic variance of $\widehat{p}^{RC}_{n,jj'}$ and a bootstrap estimator should be preferred. Asymptotic confidence intervals for $p^0_{jj'}$ at the level α are deduced from the $(1-\alpha/2)$-quantile c_α of their boostrap distributions, $I_{n,jj'}(\alpha)$ in the general case and $I^{RC}_{n,jj'}(\alpha)$ under the null hypothesis of Independent Competing Risks.

A test of the Independent Competing Risks hypothesis may be defined by rejecting H_0 if $I_{n,jj'}(\alpha)$ and $I^{RC}_{n,jj'}(\alpha)$ are not overlapping for some j'. As the estimators of the parameters $p^0_{jj'}$ are not independent, the level α^* of this test with critical region

$$R_{nj}(\alpha) = \cap_{j'=1}^m R_{njj'}(\alpha), \text{ where } R_{njj'}(\alpha) = \{I_{n,jj'}(\alpha) \cap I^{RC}_{n,jj'}(\alpha) \neq \emptyset\},$$

satisfies $\alpha^* \geq 1 - (1-\alpha)^m$.

7 Simulation results and example

The following simulations illustrate the preceding paragraph in a simple case. We consider a three states model, with transition probabilities p_{12}, p_{13} and p_{23}. From state 1 to state 2 the distribution function of the sojourn time in state 1 follows a Weibull distribution, with shape parameter equal to α_{12} and scale parameter λ_{12}, so that $\overline{F}_{|12}(t) = \exp\{-\lambda_{12}t)^{\alpha_{12}}\}$, and the analogues for the transition from state 1 to state 3 (Table 1). We assume that the sojourn time in state 1 is censored by an exponential with hazard rate λ_{cens1}, as well as the sojourn time in state 2, with parameter λ_{cens2}. Moreover, the initial probabilities are equal in all four simulations to $\rho_1 \equiv P(J_0 = 1) = 0.8$ and

$\rho_2 \equiv P(J_0 = 1) = 0.2$. The amount of censoring while in state 1, resulting from the choice of the parameters, ranges from 7% to 30%. The first two simulations deal with non ICR models and the two following ones with ICR models. In the non ICR case, p_{12}, p_{13} and p_{23} are chosen to be equal respectively to 0.75, 0.25 and 1, while they are determined by the sojourn time distributions in the ICR case. The transition probabilities and the survival functions are estimated from simulated samples of size $n = 500$, with 500 replicates for each model, assuming an ICR model (assumption H_0) or a general model. The estimates of the transition probabilities are based on (17) and (18) for the general case and on (22) for the ICR case. The estimates of the survival functions are based on (16) for the general case and on the Kaplan-Meier estimator of section 3 for the ICR case.

insert here [table 1, 2 and 3]

Tables 2 and 3 give the median and the 95%-confidence intervals for the transition probabilities over the 500 replications. They show a serious gap between the simulated confidence intervals of the transition probabilities obtained using the ICR and the general estimators under the alternative. The estimators (17) and (18) always give the same numerical value. For the simulations under H_0, both confidence intervals are widely overlapping and the general estimator of the transition probabilities is always more precise than the ICR estimator.

The confidence intervals of the survival functions in the non ICR models happened to be very close for the estimates of $\overline{F}_{12}$ in simulations 1 and 2 (Figures 2 and 4). A test for H_0 based on their comparison would lead to a wrong conclusion. A test for H_0 based on a comparison of the estimated survival functions would again be misleading in the ICR model 4 (Figures 8) for $\overline{F}_{13}$. In the ICR models, the general estimator of the survival functions $\overline{F}_{jj'}$ have always less bias and smaller confidence intervals than the Kaplan-Meier estimator. This poor performance of the Kaplan-Meier estimator seems to be an effect of the large number of censoring times in the competing risks models, the actual censoring times and the sojourn times before a transitions to the other state $j" \neq j'$. The estimator (20) has also be computed for the ICR case, it presents a strong bias in all the simulated models, as an effect of the censorship in each term of the product.

insert here [Figures 1 to 6]

The method has been applied to the heart transplant data (Kalbfleisch and Prentice, 1980) where patients evolved in a three states model: waiting for a heart transplant (state 1), heart transplanted (state 2) or dead (state 3). The total number of patients was $n = 103$ and all were in state 1 at t_0. The patients had 69 transitions from state 1 to state 2, 30 transitions from state 1 to state 3 and 45 transitions from state 2 to state 3, moreover 4 patients were censored in state 1 and 24 were censored in state 2. The estimates of the transition probabilities (18) for the general case and on (22) for the ICR case and their bootstrap confidence intervals were computed : $\widehat{p}_{12} = 0.6699029$, with confidence interval $(0.5825243, 0.7572816)$ and $\widehat{p}_{12}^{RC} = 0.4468599$, with confidence interval $(0.3109207, 0.5773497)$, $\widehat{p}_{13} = 0.2912621$, with confidence interval $(0.2038835, 0.3786408)$ and $\widehat{p}_{12}^{RC} = 0.5386473$, with confidence interval $(0.4000553, 0.6703223)$, $\widehat{p}_{13} = 0.8134351$, with confidence interval $(0.6716397, 0.9397728)$. The hypothesis of ICR for the transitions from 1 was therefore rejected for both states 2 and 3.
insert here [figure 9]

Table 1. Parameters of the four simulations.						
	α_{12}	λ_{12}	α_{13}	λ_{13}	λ_{cens1}	λ_{cens2}
Simulation 1 (non ICR) and 3 (ICR)	1	0.5	1	4	0.2	0.2
Simulation 2 (non ICR) and 4 (ICR)	1.5	0.5	1.5	4	0.2	0.2

Table 2. Estimation of the transition probabilities.
(non ICR cases)

	p_{12}						
Simulation	True value	ICR quantiles			general quantiles		
		0.025	0.5	0.975	0.025	0.5	0.975
1	0.75	0.080	0.111	0.143	0.694	0.747	0.787
2	0.75	0.025	0.042	0.062	0.702	0.748	0.791

	p_{13}						
Simulation	True value	ICR quantiles			general quantiles		
		0.025	0.5	0.975	0.025	0.5	0.975
1	0.25	0.849	0.882	0.9123	0.207	0.249	0.294
2	0.25	0.932	0.953	0.973	0.209	0.250	0.295

Table 3. Estimation of the transition probabilities.
(ICR cases)

	p_{12}						
Simulation	True value	ICR quantiles			general quantiles		
		0.025	0.5	0.975	0.025	0.5	0.975
3	0.111	0.053	0.110	0.193	0.075	0.109	0.064
4	0.042	0.006	0.0400	0.096	0.019	0.041	0.145

	p_{13}						
Simulation	True value	ICR quantiles			general quantiles		
		0.025	0.5	0.975	0.025	0.5	0.975
3	0.889	0.802	0.887	0.944	0.856	0.889	0.922
4	0.958	0.902	0.958	0.992	0.936	0.958	0.978

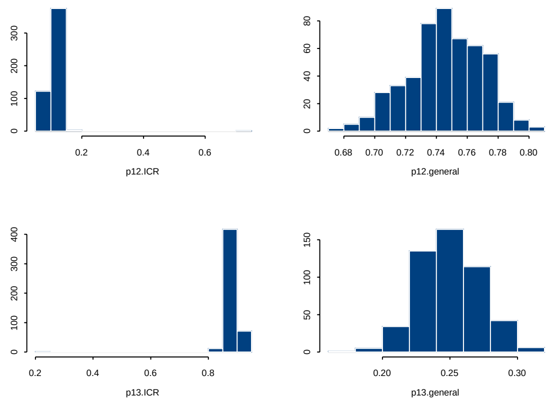

Figure 1: Histograms of transition probabilities estimations using general and ICR models. Simulation 1.

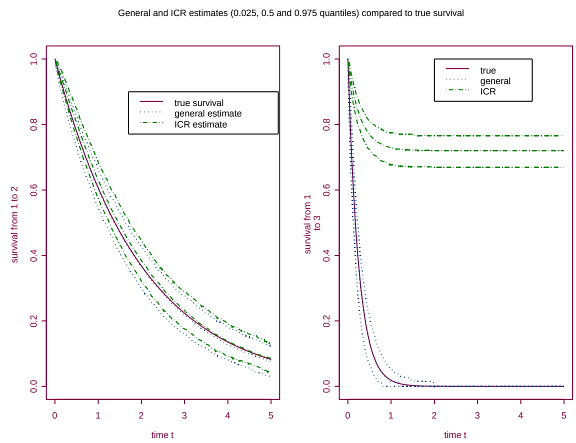

Figure 2: Comparison of survival curves estimations using general and ICR estimators. Simulation 1.

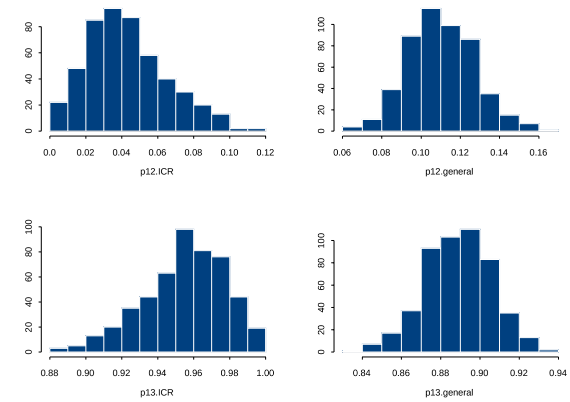

Figure 3: Comparison of survival curves estimations using general and ICR estimators. Simulation 2.

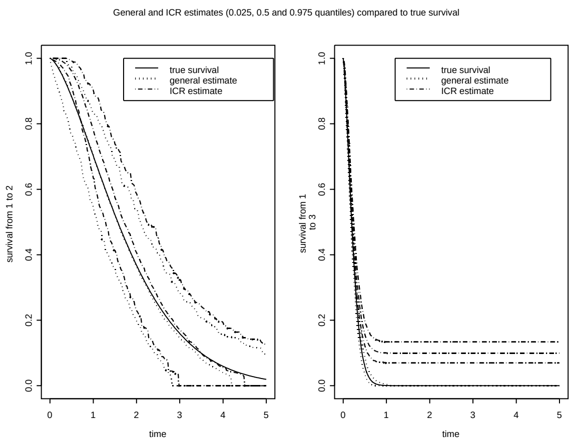

Figure 4: Comparison of survival curves estimations using general and ICR estimators. Simulation 2.

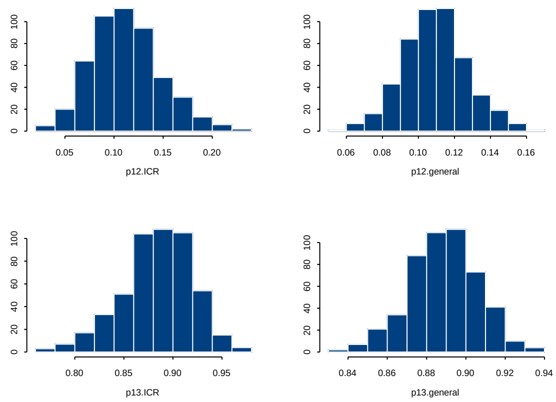

Figure 5: Histograms of transition probabilities estimations using general and ICR models. Simulation 3.

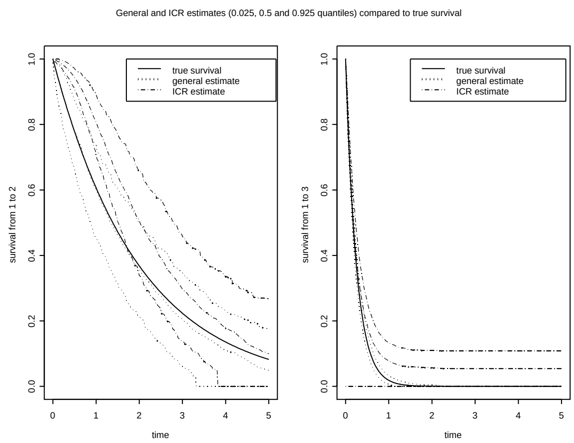

Figure 6: Comparison of survival curves estimations using general and ICR estimators. Simulation 3.

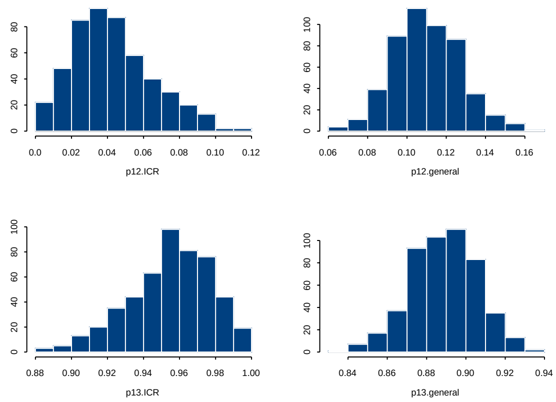

Figure 7: Histograms of transition probabilities estimations using general and ICR models. Simulation 4.

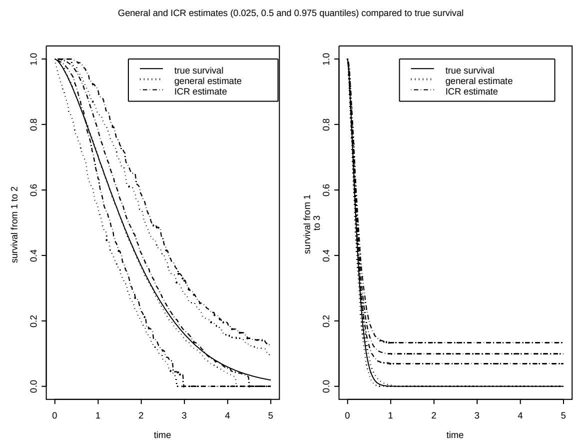

Figure 8: Comparison of survival curves estimations using general and ICR estimators. Simulation 4.

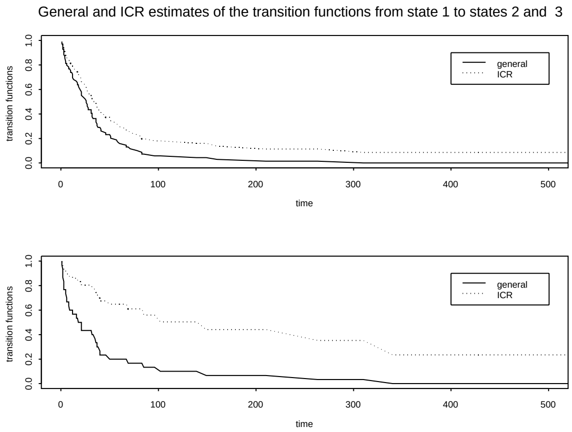

Figure 9: Estimates of transition functions for heart transplant data.

References

[1] P.K. Andersen and N. Keiding, "Event history analysis in Continuous Time," *Technical report, Department of Biostatistics, University of Copenhagen*,2000.

[2] V. Bagdonavicius and M.N. Nikulin, "Accelerated Life Models, Modeling and Statistical Analysis," Chapman & Hall/CRC ed., 2002.

[3] R. Gill, " Nonparametric estimation based on censored observations of a markov renewal process," *Z. Wahrsch. verw. Gebiete* vol. 53 pp. 97-116,1980.

[4] R. Gill, "Large sample behaviour of the product-limit estimator on the whole line," *Ann. Statist.* vol. 11 pp. 49-58, 1983.

[5] N. Heutte and C. Huber-Carol, "Semi-Markov Models for Quality of Life Data with Censoring," in *Statistical Methods for Quality of Life Studies*, Kluwer Academic Publishers, pp. 207-218, 2002.

[6] J.D. Kalbfleisch, and R.L. Prentice "The Statistical Analysis of Failure Time Data," Wiley, New York, 1980.

[7] N. Limnios and G. Oprisan, "Semi-Markov Processes and Reliability," Birkhaser ed.,222 p., 2001.

[8] O. Pons, "Estimation of semi-Markov models with right-censored data," In *Handbook of Statistics 23, "Survival Analysis"*, Elsevier *(to appear)*, 2003.

[9] S.W. Lagakos, C. Sommer and M. Zelen, "Semi-markov models for partially censored data.," *Biometrika* vol. 65 pp. 311-317, 1978.

[10] R. Pyke, "Markov renewal processes: definitions and preliminary properties," *Ann. Math. Statist.* vol. 32 pp. 1231-1342, 1961.

[11] A. Tsiatis, "A nonidentifiability aspect of the problem of competing risks," *Proc. Nat. Acad. Sci. USA* vol. 37 pp. 20-22, 1975.

[12] J. Voelkel and , J. Crowley, "Nonparametric inference for a class of semi-markov processes with censored observations," *Ann. Statist.* vol. 12 pp. 142-160, 1984.

Key-words:

semi-Markov process, competing risks, right-censoring, goodness-of-fit test.

Corresponding author:

Professeur C. Huber
Université René Descartes UFR Biomédicale
45 rue des Saints-Pres
75270 Paris Cedex 06
and U472 INSERM
16 av. P-V Couturier
94800 Villejuif
tl. : 01-42-86-21-01
Fax : 01-42-86-04-02
E-mail: huber@biomedicale.univ-paris5.fr

Affiliation of the authors:

[a] Université René Descartes, Paris 5, and U472 INSERM, 94800 Villejuif, France.

[b] Université René Descartes, Paris 5, and INRA, Jouy-en-Josas, France.

[c] Université de Caen, IUT, 14 000 Lisieux, and Université Paris 5.