# 1

# *Multivariate Survival Data With Censoring.*

**Shulamith Gross and Catherine Huber-Carol**

*Baruch College of the City University of New York, Dept of Statistics and CIS, Box 11-220, 1 Baruch way, 10010 NY.*
*Université Paris V, René Descartes, 45 rue des Saints-Pères, 75 006 Paris, and INSERM U 780.*

**Abstract:** We define a new class of models for multivariate survival data, in continuous time, based on a number of cumulative hazard functions, along the lines of our family of models for correlated survival data in discrete time (Gross and Huber, 2000, 2002). This family is an alternative to frailty and copula models. We establish some properties of our family and compare it to Clayton's and Marshall-Olkin's. Finally we derive non parametric partial likelihood estimates of the hazards involved in its definition and prove, using martingale theory, their asymptotic normality. Simulations will be performed as well as applications to diabetic retinopathy and tumorigenesis in rats.

**Keywords and phrases:** Survival data, clusters, right censoring, continuous time, hazard rates

## 1.1 Introduction

Much attention has been paid to multivariate survival models and inference since the early work of Hougaard, and his recent book (2004) on the subject. Studies on twins lead to the development of papers on bivariate distributions, and, more generally the analysis of family data or clusters data lead to more general models for correlated survival data. One way of dealing with this problem is to use copula or frailty models (see for example Bagdonavicius and Nikulin (2002) for a review of those models). Among the most usual bivariate models, one finds Clayton's, Marshall-Olkin's and Gumbel's models. We shall present here a model for continuous multivariate data based on the same idea as the one we used in the discrete case (Gross and Huber, (2002)), and which is closely related to a multi-state process. We define our class of models in detail for the special case of bivariate data, and generalize this class to any dimension. We then obtain properties of these models and compare them to the usual ones cited above. We then derive NPML estimators for the involved functions and derive their asymptotic properties.

---

## 1.2 Definition of the models

### 1.2.1 Bivariate continuous model

Let $\mathcal{L}$ be the class of continuous univariate cumulative hazard functions on $I\!R^+$:

$$\mathcal{L} = \{\Lambda : I\!R^+ \to I\!R^+, \text{ continuous, non decreasing }, \Lambda(0) = 0, \Lambda(t) \xrightarrow[t\to\infty]{} \infty\}$$

**Definition 1 (bivariate continuous model)**
*Given any five members $\Lambda_{11}^{01}, \Lambda_{11}^{10}, \Lambda_{11}^{00}, \Lambda_{01}^{00}, \Lambda_{10}^{00}$ of $\mathcal{L}$, we define a joint bivariate survival function $S$ on $I\!R^+ \times I\!R^+$ by*

$$\begin{aligned}
\textit{for } x < y\ ,\ dS(x,y) &= exp\{-\Lambda_{11}^{01}(x) - \Lambda_{11}^{10}(x) - \Lambda_{11}^{00}(x)\} d\Lambda_{11}^{01}(x) \\
&\quad exp\{-(\Lambda_{01}^{00}(y) - \Lambda_{01}^{00}(x)\} d\Lambda_{01}^{00}(y) \\
\textit{for } y < x\ ,\ dS(x,y) &= exp\{-\Lambda_{11}^{01}(y) - \Lambda_{11}^{10}(y) - \Lambda_{11}^{00}(y)\} d\Lambda_{11}^{10}(y) \qquad (1.1) \\
&\quad exp\{-(\Lambda_{10}^{00}(x) - \Lambda_{10}^{00}(y)\} d\Lambda_{10}^{00}(x) \\
\textit{for } y = x\ ,\ dS(x,y) &= exp\{-\Lambda_{11}^{01}(x) - \Lambda_{11}^{10}(x) - \Lambda_{11}^{00}(x)\} d\Lambda_{11}^{00}(x)
\end{aligned}$$

We propose the family (1.1) of bivariate probabilities as an alternative to the bivariate probabilities defined by frailties or copulas. It is easy to verify that $S$ thus defined is actually a bivariate survival function, and that a necessary and sufficient condition for the corresponding probability to be absolutely continuous (AC) with respect to $\lambda^2$, the Lebesgue measure on $I\!R^2$, is that $\Lambda_{11}^{00} \equiv 0$. Otherwise, part of the mass is on the diagonal of $I\!R^2$.

### 1.2.2 Generalization to $p$ components

When more than two components are involved, say $p$, then our hierarchical class of models is defined in a similar way, involving now a number of cumulative hazards $K(p)$ equal to

$$K(p) = \sum_{k=0}^{p-1} C_p^{p-k} C_{p-k}^1. \qquad (1.2)$$

when the multivariate law is absolutely continuous with respect to $\lambda^p$, the Lebesgue measure on $I\!R^p$, and

$$K(p) = \sum_{k=0}^{p-1} C_p^{p-k} (2^{p-k} - 1). \qquad (1.3)$$

when simultaneous jumps are allowed.

### 1.2.3 Properties of the bivariate family

**Theorem 1** *For all bivariate survival functions defined above and such that $\Lambda_{11}^{00} \equiv 0$, we have the following conditional hazard rates $\forall s < t \in I\!R^+$:*

$$\begin{array}{lcccl} P(X = dt, Y > t | X \geq t, Y \geq t) & = & d\Lambda_{11}^{01}(t) & & \\ P(X > t, Y = dt | X \geq t, Y \geq t) & = & d\Lambda_{11}^{10}(t) & & \\ P(X = dt | X \geq t, Y < t) & = & d\Lambda_{10}^{00}(t) & = & P(X = dt | X \geq t, Y = ds) \\ P(Y = dt | Y \geq t, X < t) & = & d\Lambda_{01}^{00}(t) & = & P(Y = dt | Y \geq t, X = ds) \end{array}$$

*Conversely, if there exist $\Lambda_{11}^{10}, \Lambda_{11}^{01}, \Lambda_{10}^{00}, \Lambda_{01}^{00}$, cumulative hazard functions in $\mathcal{L}$ such that the joint law satisfies the above equations, then the joint survival function of $(X, Y)$ satisfies (1.1).*

**Theorem 2** *If $(X, Y)$ has survival function $S$ given by (1.1), then $X$ and $Y$ are independent and $S$ is absolutely continuous with respect to $\lambda^2$ if and only if*

$$\Lambda_{11}^{00} \equiv 0 \; ; \; \Lambda_{11}^{01} \equiv \Lambda_{10}^{00} \; ; \; \Lambda_{11}^{10} \equiv \Lambda_{01}^{00}.$$

### 1.2.4 General bivariate model

A version of our model (1.1), in discrete time, was introduced in Gross and Huber (2000). The two models are embedded in the following general model. Let $\mathcal{L}^*$ be the set of cumulative hazards with possible jumps on an at most denumerable set of points $\mathcal{D} \in I\!R^+$:

$$\mathcal{L}^* = \{\Lambda : I\!R^+ \to I\!R^+, \Lambda \text{ non decreasing }, \Lambda(0) = 0, \Lambda(t) \xrightarrow[t \to \infty]{} \infty\}$$

**Definition 2 (general bivariate model)**
*Given any five members $\Lambda_{11}^{01}, \Lambda_{11}^{10}, \Lambda_{11}^{00}, \Lambda_{01}^{00}, \Lambda_{10}^{00}$ of $\mathcal{L}^*$ and $\mathcal{D} = \{x_1, x_2, \ldots, x_m, \ldots\}$ the ordered set of discontinuity points of the $\Lambda'$s we define a joint bivariate survival function $S$ on $I\!R^+ \times I\!R^+$ by*

*For $x < y$*

$$\begin{aligned} dS(x, y) &= \prod_{t<x} (1 - \Lambda_{11}^{01}(dt) - \Lambda_{11}^{10}(dt) - \Lambda_{11}^{00}(dt)) \;\; \Lambda_{11}^{01}(dx) \\ & \quad \prod_{x \leq t < y} (1 - \Lambda_{01}^{00}(dt)) \Lambda_{01}^{00}(dy) \end{aligned}$$

*and for $x > y$*

$$\begin{aligned} dS(x, y) &= \prod_{t<y} (1 - \Lambda_{11}^{01}(dt) - \Lambda_{11}^{10}(dt) - \Lambda_{11}^{00}(dt)) \;\; \Lambda_{11}^{10}(dy) \\ & \quad \prod_{y \leq t < x} (1 - \Lambda_{10}^{00}(dt)) \Lambda_{10}^{00}(dx) \qquad (1.4) \end{aligned}$$

*Finally for y=x*

$$dS(x, x) = \prod_{t<x} (1 - \Lambda_{11}^{01}(dt) - \Lambda_{11}^{10}(dt) - \Lambda_{11}^{00}(dt)) \;\; \Delta\Lambda_{11}^{00}(x)$$

If $\mathcal{D} = \emptyset$, then (1.4) simplifies to (1.1).

### 1.2.5 Purely discrete model:

**Definition 3 (purely discrete model)**
*Let $\lambda(u) = \Lambda(u^+) - \Lambda(u^-)$, for all five $\Lambda'$s involved in the definition (1.4), assumed to be purely discontinuous, with jumps in $\mathcal{D} = \{x_k, k \in I\!N\}$. Then define*

$$\begin{array}{lcl} \textit{For } x_i < x_j, \quad P(X = x_i, Y = x_j) & = & \prod_{k<i}(1 - \lambda_{11}^{01}(x_k) - \lambda_{11}^{10}(x_k) - \lambda_{11}^{00}(x_k))\lambda_{11}^{01}(x_i) \\ & & \prod_{i<k<j}(1 - \lambda_{01}^{00}(x_k))\lambda_{01}^{00}(x_j) \\ \textit{For } x_i > x_j, \quad P(X = x_i, Y = x_j) & = & \prod_{k<j}(1 - \lambda_{11}^{01}(x_k) - \lambda_{11}^{10}(x_k) - \lambda_{11}^{00}(x_k))\lambda_{11}^{10}(x_j) \\ & & \prod_{j<k<i}(1 - \lambda_{10}^{00}(x_k))\lambda_{10}^{00}(x_i) \\ \textit{For } x_i = x_j, \quad P(X = x_i, Y = x_i) & = & \prod_{k<i}(1 - \lambda_{11}^{01}(x_k) - \lambda_{11}^{10}(x_k) - \lambda_{11}^{00}(x_k))\lambda_{11}^{00}(x_i) \end{array}$$

### 1.2.6 Simple examples of laws of type (1.1)

Let $a, b, c, d$ be four strictly positive constants and $d\Lambda_{11}^{01}(t) = a$, $d\Lambda_{11}^{10}(t) = b$, $d\Lambda_{10}^{00}(t) = c$, $d\Lambda_{01}^{00}(t) = d$. Then, denoting $S$ the bivariate survival, we have:

$$\begin{array}{rcll} \frac{d^2S(x,y)}{dxdy} & = & e^{-(a+b)x}ae^{-d(y-x)}d & \text{if} \quad x < y \\ & = & e^{-(a+b)y}be^{-c(x-y)}c & \text{if} \quad x > y \end{array}$$

Although all four hazard rates are constant, the marginals of these distributions are not exponential. Other simple examples arise from replacing the above exponential hazards by other families, like for example, Weibull or Pareto.

---

## 1.3 Some usual bivariate models

Among the most usual bivariate models, one finds Clayton's, Marschall-Olkin's and Gumbel's models.

### 1.3.1 Clayton bivariate distribution

Clayton survival function (1978), parametrized by Oakes (1989) is given by:

$$S(x, y) = P(X_1 > x, X_2 > y) = [S_1(x)^{-(\theta-1)} + S_2(y)^{-(\theta-1)} - 1]^{\frac{-1}{\theta-1}} \qquad (1.5)$$

where $\theta \in ]1 + \infty[$ and $S_1(x)$ and $S_2(y)$ are the marginal survival functions for $X_1$ and $X_2$. The limiting distribution, when $\theta \to 0$ has independent components. We change parameter, letting

$$\begin{array}{rcl} a & = & \frac{1}{\theta-1} \\ \theta & > & 1 \; ; \; a \; > \; 0. \end{array}$$

Genest et al propose a pseudo likelihood (PL) estimate for $a$. Their PL is based on a copula defining the joint distribution function $F(x,y) = c_\alpha(F_1(x), F_2(y))$. It is the product, for all observations $(x_i, y_i)$, of the second partial derivative of $c_\alpha(u,v)$ with respect to $u$ and $v$. $u$ and $v$ are further respectively replaced by $\widehat{F_1(x)}$ and $\widehat{F_2(y)}$. With our copula acting on the survival rather than on the d.f., the corresponding PL is derived below. If

$$S_a(u,v) \quad = \quad [u^{-1/a} + v^{-1/a} - 1]^{-a},$$

the pseudo-likelihood is equal to

$$\prod_{i=1}^{n} [\frac{\partial^2 S_a(u,v)}{\partial u \partial v}]_{u=\widehat{S_1}(x_i), v=\widehat{S_2}(y_i)}$$

As

$$\frac{\partial^2 S_a(u,v)}{\partial u \partial v} = (1+\frac{1}{a})\frac{1}{uv} e^{-\frac{1}{a}\log(uv)} [u^{-\frac{1}{a}} + v^{-\frac{1}{a}} - 1]^{-a-2}$$

one can compute easily the PL substituting the Kaplan-Meier estimates $\widehat{S_1}$ and $\widehat{S_2}$ for $S_1$ and $S_2$.

### 1.3.2 Marshall-Olkin bivariate distribution:

Let $\lambda_1$, $\lambda_2$ and $\lambda_{12}$ be three positive constants and $S$ the bivariate survival function

$$S(x_1, x_2) = P(X_1 \geq x_1, X_2 \geq x_2) = e^{-\lambda_1 x_1 - \lambda_2 x_2 - \lambda_{12}(x_1 \vee x_2)} \tag{1.6}$$

It is clear that the bivariate Marshall-Olkin is not absolutely continuous with respect to $\lambda^2$ as

$$P(X_1 > X_2) + P(X_2 > X_1) = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_{12}} \tag{1.7}$$

Moreover, denoting $min = 1\{x_1 < x_2\} + 2 * 1\{x_2 < x_1\}$ and $max = 1\{x_1 > x_2\} + 2 * 1\{x_2 > x_1\}$, the density at point $(x_1, x_2)$ ; $x_1 \neq x_2$ may be written as:

$$f(x_1, x_2) = \lambda_{min}(\lambda_{max} + \lambda_{12}) e^{-(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_{12} x_{max})} \tag{1.8}$$

The deficit to one is due to the fact that there is a probability mass equal to

$$\frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}}$$

on the diagonal. The linear density on the diagonal is equal to

$$f_0(x) = \lambda_{12} e^{-(\lambda_1 x + \lambda_2 x + \lambda_{12} x)} \tag{1.9}$$

as can be derived from looking at the following limit:

$$\lim_{dt\to 0} \frac{1}{dt}(S(t,t) - S(t,t+dt) - S(t+dt,t) + S(t+dt,t+dt))$$

The corresponding hazards in our scheme would be

$$\begin{array}{rclcl} \lambda_{11}^{01}(t) &=& \frac{P(X_1=t,X_2>t)}{P(X_1\geq t,X_2\geq t)} &=& \lambda_1 \quad ; \quad \lambda_{11}^{10}(t) = \frac{P(X_1>t,X_2=t)}{P(X_1\geq t,X_2\geq t)} = \lambda_2 \\ \lambda_{10}^{00}(t) &=& \frac{P(X_1=t,X_2\leq t)}{P(X_1\geq t,X_2\leq t)} &=& \lambda_1+\lambda_{12} \quad ; \quad \lambda_{01}^{00}(t) = \frac{P(X_2=t,X_1\leq t)}{P(X_1\leq t,X_2\geq t)} = \lambda_2+\lambda_{12} \\ \lambda_{11}^{00}(t) &=& \lambda_{12}. \end{array}$$

It can be seen from the following example of our class of distributions thus defined:

### 1.3.3 Our quasi-Marshall-Olkin bivariate distribution:

Let a bivariate distribution be defined as in (1.1), the hazards being equal to:

$$\begin{array}{ccccc} (11) & \stackrel{\lambda_{11}^{01}\equiv\lambda_1\equiv\alpha}{\longrightarrow} & (01) & \stackrel{\lambda_{01}^{00}\equiv\lambda_2+\lambda_{12}\equiv\beta+\gamma}{\longrightarrow} & (00) \\ (11) & \stackrel{\lambda_{11}^{10}\equiv\lambda_2\equiv\beta}{\longrightarrow} & (10) & \stackrel{\lambda_{01}^{00}\equiv\lambda_1+\lambda_{12}\equiv\alpha+\gamma}{\longrightarrow} & (00) \end{array}$$

and $\lambda_{11}^{00}$ being identically null. Let us denote

$$\begin{array}{rcl} Y &=& min(X_1,X_2) \\ Z &=& max(X_1,X_2). \end{array}$$

Following the log-likelihood $L$ derived predecingly, we obtain:

$$\begin{array}{rcl} L(x_1,x_2) &=& 1\{x_1 < x_2\} * [-(\alpha+\beta)y + \log(\alpha) - (\beta+\gamma)(z-y) + \log(\beta+\gamma)] \\ && +1\{x_2 < x_1\} * [-(\alpha+\beta)y + \log(\beta) - (\alpha+\gamma)(z-y) + \log(\alpha+\gamma)] \\ &=& -(\alpha+\beta)y - \gamma(z-y) \\ && +1\{x_1 < x_2\} * [\log(\alpha(\beta+\gamma)) - \beta(z-y)] \\ && +1\{x_2 < x_1\} * [\log(\beta(\alpha+\gamma)) - \alpha(z-y)] \end{array}$$

In order to compare our distribution to Marshall-Olkin's, let $g(x_1,x_2)$ be the density:

$$g(x_1,x_2) = \lambda_{min}(\lambda_{max}+\lambda_{12})e^{-(\lambda_1 x_1+\lambda_2 x_2+\lambda_{12}(x_{max}-x_{min}))} \tag{1.10}$$

One can see that only $x_{max}$ is replaced by $x_{max} - x_{min}$. As a result,

$$\int_0^\infty g(x_1,x_2)dx_1 dx_2 = 1.$$

It is an A.C. distribution. If we add the $\lambda_{11}^{00}$ as in the preceding paragraph, we get the Marshall-Olkin distribution.

### 1.3.4 Gumbel bivariate distribution

Gumbel bivariate exponential distribution is part of the general Morgenstern proposal for bivariate distributions:

$$F(x,y) = F_1(x)F_2(y)[1+\alpha(1-F_1(x))(1-F_2(y))], \quad (-1 \leq \alpha \leq 1). \tag{1.11}$$

where $F_1$ and $F_2$ are the respective marginal distribution functions for $X$ and $Y$. Gumbel bivariate exponential is thus equal to:

$$F(x,y) = (1-e^{-x})(1-e^{-y})[1+\alpha e^{-x}e^{-y}]/;/;(x \geq 0, y \geq 0, -1 \leq \alpha \leq 1). \tag{1.12}$$

In order to simulate this law, one may notice that the conditional distribution of $Y$ with respect to $X$ is given by:

$$P(Y \leq y | X = x) = (1-\alpha(2e^{-x}-1))(1-e^{-y}) + \alpha(2e^{-x}-1)(1-e^{-2y}). \tag{1.13}$$

---

## 1.4 NPML estimation

### 1.4.1 Likelihood for the bivariate case

Let $X = (X_{i1}, X_{i2})$ be the bivariate survival time of cluster $i$, $i \in \{1, 2, \ldots, n\}$. The clusters are assumed to be independent. $X_{i1}$ and $X_{i2}$ may possibly be right censored by a bivariate censoring time $C = (C_{i1}, C_{i2})$, independent of $X$, so that the observed bivariate time is $T = ((X_{i1} \wedge C_{i1}, X_{i2} \wedge C_{i2}) \equiv (T_{i1}, T_{i2})$. The indicator of non censoring is denoted $\delta = (\delta_{i1}, \delta_{i2}) \equiv (1\{T_{i1} = X_{i1}\}, 1\{T_{i2} = X_{i2})\}$. Let then $R(t) = (R_{i1}(t), R_{i2}(t))$ and $N(t) = (N_{i1}(t), N_{i2}(t))$ be respectively the associated at risk and counting processes defined for $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2\}$ as

$$\begin{aligned} R_{ij}(t) &= 1\{t < T_{ij}\} \\ N_{ij}(t) &= \delta_{ij} 1\{t \geq T_{ij}\} \end{aligned}$$

The likelihood will be expressed in terms of the following hazards defined for $X = (X_1, X_2)$:

$$\begin{aligned} \lambda_{11}^{01}(t)dt &= P(t \leq X_1 \leq t+dt | X_1 \geq t, X_2 > t) \\ \lambda_{11}^{10}(t)dt &= P(t \leq X_2 \leq t+dt | X_1 > t, X_2 \geq t) \\ \lambda_{10}^{00}(t)dt &= P(t \leq X_1 \leq t+dt | X_1 \geq t, X_2 < t) \\ \lambda_{01}^{00}(t)dt &= P(t \leq X_2 \leq t+dt | X_1 < t, X_2 \geq t) \end{aligned}$$

The likelihood for the $n$ clusters is the product $V = \prod_{i=1}^n V_i$ where each $V_i$ may be written as

$$\begin{aligned} V_i &= \prod_t (1-\lambda_{11}^{10}(t)dt - \lambda_{11}^{01}(t)dt)^{R_1(t)R_2(t)} (\lambda_{11}^{10}(t))^{R_1(t^-)R_2(t^-)dN_1(t)} \\ &\quad (\lambda_{11}^{01}(t))^{R_1(t^-)R_2(t^-)dN_2(t)} \prod_t (1-\lambda_{10}^{10}(t)dt)^{R_1(t)(1-R_2(t))\delta_2} \\ &\quad \prod_t (1-\lambda_{01}^{01}(t)dt)^{R_2(t)(1-R_1(t))\delta_1} (\lambda_{10}^{10}(t)dt)^{R_1(t)(1-R_2(t))\delta_2 dN_1(t)} \\ &\quad (\lambda_{01}^{01}(t)dt)^{R_2(t)(1-R_1(t))\delta_1 dN_2(t)} \end{aligned}$$

### 1.4.2 NPML estimation

Maximization of the log-likelihood (NPML) implies jumps of the $\Lambda's$ at (ordered) times $T_k$ , $k = 1, 2 \cdots, K$ when an event occurred ($\delta_{ij} = 1$ for some $(i,j)$). Let us introduce the quantities:

$$\begin{array}{lll} \tau_1(i) = 1\{T_{i1} < T_{i2}\} \ ; \ \tau_2(i) = 1\{T_{i2} < T_{i1}\} & ; & \tau(i) = 1\{T_{i1} = T_{i2}\} \\ a_k = \Lambda_{11}^{01}(T_k^+) - \Lambda_{11}^{01}(T_k^-) & ; & b_k = \Lambda_{11}^{10}(T_k^+) - \Lambda_{11}^{10}(T_k^-) \ ; \\ c_k = \Lambda_{10}^{00}(T_k^+) - \Lambda_{10}^{00}(T_k^-) & ; & d_k = \Lambda_{01}^{00}(T_k^+) - \Lambda_{01}^{00}(T_k^-) \ . \end{array}$$

and the counts:

$$\begin{array}{ll} s_1(i) = \sum_{i'} 1\{T_{i1} \le T_{i'1} \wedge T_{i'2}\} & ; s_2(i) = \sum_{i'} 1\{T_{i2} \le T_{i'1} \wedge T_{i'2}\} \ ; \\ s_3(i) = \sum_{i'} \tau_2(i') 1\{T_{i'2} \le T_{i1} \le T_{i'1}\}\} & ; s_4(i) = \sum_{i'} \tau_1[i'] 1\{T_{i'1} \le T_{i2} \le T_{i'2}\}\}. \end{array}$$

Then the log-likelihod is equal to

$$\begin{aligned} L = & -\sum_i a_i \delta_{i1} \tau_1(i) s_1(i) - \sum_i b_i \delta_{i2} \tau_2(i) s_2(i) + \sum_i \delta_{i1} \tau_1(i) \log(a_i) \\ & \sum_i \delta_{i2} \tau_2(i) \log(b_i) - \sum_i c_i \delta_{i1} \tau_2(i) b_i s_3(i) - \sum_i d_i \delta_{i2} \tau_1(i) b_i s_4(i) \\ & \sum_i \delta_{i1} \tau_2(i) \log(c_i) + \sum_i \delta_{i2} \tau_1(i) \log(d_i) \end{aligned}$$

By derivation of $L$ with respect to the jumps $a_i, b_i, c_i, d_i$, we obtain the following NPML estimates:

$$\widehat{a_i} = \frac{\delta_{i1}\tau_1(i)}{s_1(i)} \ ; \ \widehat{b_i} = \frac{\delta_{i2}\tau_2(i)}{s_2(i)} \ ; \ \widehat{c_i} = \frac{\delta_{i1}\tau_2(i)}{s_3(i)} \ ; \ \widehat{d_i} = \frac{\delta_{i2}\tau_1(i)}{s_4(i)} \ .$$

In order to derive the asymptotic properties of the NPML estimates, one rewrites them in terms of the following associated counting processes.

### 1.4.3 Associated point processes and martingales

Let

$$\mathcal{F}(t) = \sigma(N_{i1}(s), N_{i2}(s), R_{i1}(s), R_{i2}(s), s < t) \tag{1.14}$$

be a filtration and define four point processes $N$ with associated presence at risk processes $Y$, for each case: jump of individual 1 (respectively 2) in the presence (respectively absence) of the other element of the pair:

$$
\begin{array}{lcllcl}
N_{i,11:01}(t) & = & 1\{X_{i1} \leq t, X_{i1} < X_{i2} \wedge C_{i1} \wedge C_{i2}\} & = & \int_0^t R_{i1}(s)R_{i2}(s)dN_{i1}(s) \\
Y_{i,11}(t) & = & 1\{X_{i1} \wedge X_{i2} \wedge C_{i1} \wedge C_{i2} \geq t\} & = & R_{i1}(t)R_{i2}(t) \\
M_{i,11:01}(t) & = & N_{i,11:01}(t) - \int_0^t Y_{i,11}(u)d\Lambda_{11}^{01}(u) & & \\
\\
N_{i,11:10}(t) & = & 1\{X_{i2} \leq t, X_{i2} < X_{i1} \wedge C_{i1} \wedge C_{i2}\} & = & \int_0^t R_{i1}(s)R_{i2}(s)dN_{i2}(s) \\
Y_{i,11}(t) & = & 1\{X_{i1} \wedge X_{i2} \wedge C_{i1} \wedge C_{i2} \geq t\} & = & R_{i1}(t)R_{i2}(t) \\
M_{i,11:10}(t) & = & N_{i,11:10}(t) - \int_0^t Y_{i,11}(u)d\Lambda_{11}^{10}(u) & & \\
\\
N_{i,10:00}(t) & = & 1\{X_{i2} < X_{i1} \wedge C_{i1}, X_{i2} \leq C_{i2} \wedge t\} & = & \int_0^t R_{i1}(s)(1 - R_{i2}(s))dN_{i1}(s) \\
Y_{i,10}(t) & = & 1\{X_{i1} \wedge C_{i1} \geq t, X_{i2} < t, X_{i2} \leq C_{i2}\} & = & R_{i1}(t)(1 - R_{i2}(t)) \\
M_{i,10:00}(t) & = & N_{i,10:00}(t) - \int_0^t Y_{i,10}(u)d\Lambda_{10}^{00}(u) & & \\
\\
N_{i,01:00}(t) & = & 1\{X_{i1} < X_{i2} \wedge C_{i2}, X_{i1} \leq C_{i1} \wedge t\} & = & \int_0^t (1 - R_{i1}(s))R_{i2}(s)dN_{i2}(s) \\
Y_{i,01}(t) & = & 1\{X_{i2} \wedge C_{i2} \geq t, X_{i1} < t, X_{i1} \leq C_{i1}\} & = & (1 - R_{i1}(t))R_{i2}(t) \\
M_{i,01:00}(t) & = & N_{i,01:00}(t) - \int_0^t Y_{i,01}(u)d\Lambda_{01}^{00}(u) & &
\end{array}
$$

The whole asymptotic normal theory holds as the estimates of the cumulative $\Lambda' s$, properly normalized converge to independent gaussian martingales with estimable auto-covariances.

---

## 1.5 Concluding remarks

The proposed model could be considered as a multistate model, where the successive states are the actual composition of the subset of the cluster that is *still at risk* after some members have experienced the expected event. In a future work, we shall introduce covariates such as cluster and individual covariates as well as the time elapsed between two successive states of the cluster Let us finally remark that the parallel with semi-Markov models for multistate models is not straightforward. This is due to the fact that, for example in the bivariate case, when the pair is in state $(0,1)$ the cumulative hazard $\Lambda_{01}^{00}$ starts from 0 and not from the time $s$ at which the first member of the pair experienced the event. Making the parallel perfect would lead to a new family of models having all properties of semi-markov multistate models, to which could be applied all results already obtained for example by Huber, Pons et Heutte (2006).

---

## References

1. Bagdonavicius, V. and Nikulin, M. (2002). *Accelerated Life Models*, Chapman and Hall/CRC, Boca Raton.

2. Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**, 543–52.

3. Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion), *Journal of the Royal Statistical Society A***148**, 82–117.

4. Genest, C., Ghoudi, K. and Rivest L-P (1995). A semi-parametric estimation procedure of dependence parameter in multivariate families of distributions, *Biometrika***82(3)**, 543–552.

5. Gross, S. and Huber-Carol, C. (2000). Hierarchical Dependency Models for Multivariate Survival data with Censoring, *Lifetime Data Analysis***6**, 299–320.

6. Gross, S. and Huber-Carol, C. (2002). A new family of Multivariate Distributions for Survival Data, *Goodness of Fit tests and Model Validity*, ed. Birkhauser, 255–266.

7. Hougaard, P. (2004). *Analysis of Multivariate Survival Data*, Springer Verlag ed.

8. Huber-Carol, C., Pons, O. and Heutte N. (2006). Inference for a general semi-Markov model and a sub-model for independent competing risks, In *Probability, Statistics and Modelling in Public Health* (Ed., M. Nikulin, D. Commenges, C. Huber), pp. 231–244, Springer.

9. Lee, Wei, and Ying (1993) Linear regression analysis for highly stratified failure time data, *Biometrics*,**58**, 643-649.

10. Marschall, A.W. and Olkin, I. (1967). A multivariate exponential distribution, *J.A.S.A.*,**62**, 30–44.

11. Oakes, D. (1989). Bivariate survival models induced by frailties, *Journal of the American Statistical Association*,**84**, 487–493.

12. Ross, E.A. and Moore, D. (1999) Modeling clustered, discrete, or grouped time survival data with covariates, *Biometrics*,**55(3)**,813–819.