Review of Multivariate Survival Data

Guadalupe Gómez[(1)], M.Luz Calle[(2)], Anna Espinal[(3)] and Carles Serrat[(1)]

(1) Universitat Politècnica de Catalunya
(2) Universitat de Vic
(3) Universitat Autònoma de Barcelona

DR2004/15

Review of Multivariate Survival Data *

Guadalupe Gómez
lupe.gomez@upc.edu
Departament d'Estadística i I.O.
Universitat Politècnica de Catalunya

M.Luz Calle
malu.calle@uvic.es
Departament de Matemàtica i Informàtica
Universitat de Vic

Carles Serrat
carles.serrat@upc.es
Departament de Matemàtica Aplicada I
Universitat Politècnica de Catalunya

Anna Espinal
anna.espinal@uab.es
Departament de Matemàtiques
Universitat Autònoma de Barcelona

September 2004

*Partially supported by grant BFM2002-0027 PB98–0919 from the Ministerio de Ciencia y Tecnologia

Contents

1 Introduction

The goal of this technical report is to review some of the main contributions in the area of multivariate survival data and to propose some possible extensions. In particular, we have concentrated our search and study on those papers that are relevant to the situation where two (or more) consecutive variables are followed until a common day of analysis and subject to informative censoring.

In two different situations we have encountered consecutive times subject to dependent censoring and we have not found standard tools to estimate their corresponding joint distributions, neither to identify prognostic factors for the joint survival. These two instances are briefly described below to motivate the technical report but their specific analysis are not pursued here and will be published elsewhere.

The first example corresponds to the joint estimation of the survival of the first two times without treatment in a clinical trial known as the Tibet project. Tibet is an ongoing trial in which an intermittent therapeutic strategy aiming to improve the quality of life of HIV-infected patients, as well as to reduce the toxic effects of HAART (Highly Active Antiretroviral Therapy), has been assigned to each patient. This strategy defines a sequence of successive stages on which the patients are alternatively without treatment (state OFF) or under treatment (state ON). The process for an individual consists of a series of sojourns in states OFF and ON and various lifetime variables can be defined within this process. Of special clinical interest are the lifetime variables $T_1, T_2, T_3, \ldots, T_m$ defined as the length that a patient stays on each stage. That is, T_1 is the time, from randomization, that a patient stays without treatment. T_2 accounts for the time since the patient restarts treatment until he/she is switched off and T_3 denotes the second time OFF, that is, the time that a patient is without treatment for the second time. One of the goals of such a clinical trial was to characterize the time spent in state ON (T_2) for different sojourns in state OFF (T_1), as well as to study the second time without treatment (T_3) in relation to the first time without treatment (T_1).

The second instance we have approached with a similar methodological problem relates to the estimation of the maturation time and the senescentia time of a climateric fruit and is of great importance in shell life studies. In particular, our coresearchers are interested in bananas which are climateric fruits harvested while still "living", and continue the same metabolic reac-

tions as when it is attached to the plant; it continues growing and maturing (developing attractive flavors and aromas) and improving its eating characteristics. During this period, banana consumer acceptance increases. At one point, catabolic (degenerative/breaking down) biochemical processes replace anabolic (synthetic/building up) processes. This change is called senescence. It causes aging and finally death. In banana, the most obvious changes relating to fruit senescence are peel color change, from yellow to black, and softening pulp. During this period, banana consumer acceptance decreases. We have thus a period of sensorial quality improvement (T_1) followed by a degradative one (T_2). The correct estimation of both periods (consumer acceptance versus consumer rejection) -dependent on the storage time and conditions- is an exceptional tool to correctly manage the post-harvest product, both for banana growers and retailers.

In those studies, where two, or more, consecutive times are observed and the censoring mechanisms acts on their sum, the estimation of the joint survival function of the first, T_1, and second, T_2, duration has to take into account the fact that T_2 is only observed if $T_1 + T_2$ does not exceed the total time of follow–up. That is, the censoring mechanism acting on T_2 will depend on T_1 and if T_1 and T_2 are correlated we will be dealing with a particular case of dependent censoring.

The paper goes as follows. We start reviewing bivariate nonparametric approaches in Section 2. We extend some of the bivariate proposals to the case of two nonconsecutive times in Section 3. We devote Section 4 to introduce the notation and construct the likelihood for the general problem of more than two consecutive survival times. While in this section we state a second order Markov assumption on which we base a Cox model and an alternative partial likelihood, we postpone to Section 5 the formulation of the time dependencies and trends via a Bayesian approach. Finally, in Section 6, three regression models for multivariate survival times are discussed together with the differences among them which will be useful when the main interest is on the effect of covariates on the risk of failure.

2 Review of bivariate nonparametric approaches

2.1 Introduction and notation

The bivariate survival estimator is useful in predicting the joint survival experience, in estimating the degree of dependence, in model building and testing and in strengthening marginal analysis. As Gill (1992) points out is a notoriously tough problem.

Methods for analyzing univariate censored data have been studied by many scientists over the last decades, however, relatively little research has been devoted to the analysis of bivariate observations in the presence of dependent censoring. In many studies two times are observed, examples include twin studies, matched pair studies, studies that record two different failure times and studies in chronic diseases where recurrent and death times are recorded. Other biomedical examples include times to severe visual loss on the left and right eyes, times to cancer detection in the left and right breast. AIDS studies present two important consecutive times: time to HIV infection and subsequent time to AIDS diagnosis.

Most of the published papers concerning the nonparametric estimation of the bivariate distribution functions under random censorship distinguish whether censoring is bivariate –different censoring in each coordinate– or univariate –the same censoring for both coordinates– but only few of them take into account dependent censoring caused by successive times.

We present in subsection 2.2 an overview of several approaches under independent right-censoring. Subsection 2.3 is devoted to the nonparametric estimation of the survival function in the presence of dependent censoring. This section concludes with a brief presentation of dependence measures.

Let T_1 and T_2 be two duration variables. Let F, F_1 and F_2 be the bivariate and marginal distribution functions for (T_1, T_2) and let S, S_1 and S_2 be the bivariate and marginal survival functions for (T_1, T_2):

$$\begin{aligned} F(t_1,t_2) &= \text{Prob}\{T_1 \le t_1, T_2 \le t_2\} \\ S(t_1,t_2) &= \text{Prob}\{T_1 > t_1, T_2 > t_2\} \\ F_1(t_1) &= \text{Prob}\{T_1 \le t_1\} \quad F_2(t_2) = \text{Prob}\{T_2 \le t_2\} \\ S_1(t_1) &= \text{Prob}\{T_1 > t_1\} \quad S_2(t_2) = \text{Prob}\{T_2 > t_2\} \end{aligned}$$

Two different types of censoring can occur. Bivariate censoring implies a different censoring random variable for each of the two coordinates while univariate censoring would assume the same censoring for each coordinate. For the first one consider the population of twins which have suffered a heart attack and the interest relies on the elapsed time to the second heart attach for each one. The censoring here will be, most probably, bivariate since the follow-up time will be independent for each twin. On the other hand, the joint behaviour of the ages of cancer diagnosis in each breast has major health implications. The censoring –due to lost of follow-up or end of study– is here univariate.

For the bivariate censoring define (C_1, C_2) as the right-censor random variables for (T_1, T_2) and let G be the bivariate survival function for (C_1, C_2):

$$G(c_1, c_2) \;=\; \text{Prob}\{C_1 > c_1, C_2 > c_2\}.$$

The observed vector is (Y_1, Y_2, D_1, D_2) where $Y_j = \min\{T_j, C_j\}$ and $D_j = \mathbf{1}\{T_j \leq C_j\}$ $(j = 1, 2)$.

For the univariate censoring define C as the right-censor random variable for (T_1, T_2) and let G_1 be the survival function for C:

$$G_1(c) \;=\; \text{Prob}\{C > c\}. \tag{1}$$

Note that this is a particular case of bivariate censoring where $\text{Prob}\{C = C_1 = C_2\} = 1$. The observed vector is (Y_1, Y_2, D_1, D_2) where $Y_j = \min\{T_j, C\}$ and $D_j = \mathbf{1}\{T_j \leq C\}$ $(j = 1, 2)$.

In either censoring case, assume that we have a random sample $\{(T_{1i}, T_{2i}, C_{1i}, C_{2i}), i = 1, \ldots, n\}$ of (T_1, T_2, C_1, C_2) from which the observed data is $\{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}), i = 1, \ldots, n\}$.

There is a complementary classification which distinguishes whether or not the failures act simultaneously. When a study subject can potentially experience multiple events, the data can fall into one of two categories: parallel and serial (Lin, Sun and Ying, 1999). In the parallel system several possibly dependent failure processes act concurrently, while in the serial system there is a natural ordering of the times of occurrence of events. In the parallel system the censoring can be bivariate or univariate and most of the papers consider the situation under which the vector (C_1, C_2) is independent of (T_1, T_2) (or C independent of (T_1, T_2)).

The serial sampling scheme, which is very useful for describing the evolution of a multistage disease process or a process of recurrent events, is affected

by what is known as dependent censoring and is such that the censoring of the second time depends on the first. We encounter this type of situation both in the shell life of bananas and in the interruption times in the Tibet clinical trial. This case is reviewed in subsection 2.3, where specific notation and considerations about censoring is introduced.

2.2 Nonparametric estimation of the bivariate distribution function under independent censoring

All the papers reviewed in this subsection assume independent bivariate censoring, except the last one (Lin and Ying, 1993) which assumes independent univariate censoring.

- The first paper that tackles the bivariate estimation problem for the distribution function from a nonparametric viewpoint is Campbell (1981). Campbell derives two estimators for the bivariate distribution function under the presence of independent pairs of censoring variables. The first estimator is a reduced-sample estimator based on the factorization of the bivariate distribution function for the observed random vector as a product of the bivariate distribution function for the bivariate vector of interest and the bivariate distribution function for the censoring variables. The second estimator is a discrete bivariate self-consistent estimator. Campbell estimator is in essence a nonparametric MLE (other approaches by Hanley and Parnes (1983)). Although these estimators are shown to be strongly uniform consistent at a rate of convergence equal to that of the empirical distribution function, they are not necessarily monotone increasing in both coordinates and therefore not a "legitimate" bivariate distribution function. Furthermore, the nonparametric MLE does not have a closed form expression and presents a nonuniqueness problem.

- Two years later Campbell and Földes (1982) propose two path-dependent estimators again under the assumption that the censoring times are independent of the failure times. Based on the path decomposition of the bivariate survival function as the product of the conditional times the marginal:

$$S(t_1, t_2) \quad = \quad \text{Prob}\{T_2 > t_2 | T_1 > t_1\} S_1(t_1), \tag{2}$$

Campbell and Földes propose a one-dimensional Kaplan-Meier product limit estimator for each term and their corresponding product as bivariate survival estimator. Their estimator, denoted by $\hat{S}_n$, reduces to the ordinary empirical survival function in the case of no censoring in either coordinate. Note that for the implementation of such estimator standard univariate survival software can be used. For each t_1, $S_1(t_1)$ can be estimated by the ordinary Kaplan-Meier ignoring all the information regarding T_2. To estimate $S_{T_2|T_1}(t_2|t_1) = \text{Prob}\{T_2 > t_2|T_1 > t_1\}$ we calculate the Kaplan-Meier estimator of T_2 based on those individuals for which $T_1 > t_1$. Their estimator is nevertheless uniformly almost surely consistent for the survival $S(t_1, t_2)$ with rate $O(\sqrt{\frac{\log\log n}{n}})$. In the same paper Campbell and Földes propose a second estimator based on the bivariate hazard function. They prove as well that this second estimator is asymptotically equivalent to the first. This estimator suffers from two drawbacks: it is not a legitimate survival function and is dependent on the selected path and ordering of the components.

- Tsai, Leurgans and Crowley (1986) propose a family of closed form estimators that are always survival functions based on a decomposition of the bivariate survival in terms of identifiable survival an subsurvival functions extending Peterson's ideas and some smoothing techniques. Their estimators are fairly complicated and involve the estimation of conditional survival functions using density estimation techniques. Their estimators are kernel and bandwidth dependent. They prove that they are uniformly consistent under bivariate censoring and self-consistent under univariate censoring. However, the rate of convergence of their estimator is slower than Campbell and Földes and Dabrowska and inferior to the rate for the empirical distribution function.

- Burke (1988) modifies Campbell's estimator in such a way that the rate of convergence is preserved but the monotonicity requirements are fulfilled. Burke proposes an estimator based on the representation of the bivariate distribution function as the convolution of the subdistribution which can be naturally estimated by the observed data and the inverse of the bivariate distribution function for the censoring times.

  Define the subdistribution function

$$\tilde{F}(x, y) = \text{Prob}\{Y_1 \leq x, Y_2 \leq y, D_1 = 1, D_2 = 1\}$$

which can be expressed as the convolution of F and G:

$$\tilde{F}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} G(u,v)dF(u,v).$$

The bivariate distribution function F for the random vector (T_1, T_2) can be expressed in terms of $\tilde{F}$ and G and is the basis of Burke's approach:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \{G(u,v)\}^{-1} d\tilde{F}(u,v) \tag{3}$$

The subdistribution function $\tilde{F}(x,y)$ can be naturally estimated from the empirical subdistribution function

$$F_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} D_{1i}D_{2i}\mathbf{1}\{Y_{1i} \le x, Y_{2i} \le y\}.$$

G can be estimated using the Campbell-Földes approach computed from the data $(Y_{1i}, Y_{2i}, 1 - D_{1i}, 1 - D_{2i})$ $(i = 1, \ldots, n)$. Denote the estimator by G_n.

The estimation of F is accomplished plugging $F_n(u,v)$ and $G_n(u,v)$ into the integral (21).

$$\begin{aligned}\hat{F}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} \{G_n(u,v)\}^{-1} dF_n(u,v) \\ &= \frac{1}{n}\sum_{i=1}^{n} D_{1i}D_{2i}\{G_n(Y_{1i}, Y_{2i})\}\mathbf{1}\{Y_{1i} \le x, Y_{2i} \le y\}.\end{aligned}$$

Since F_n can be expressed as an integral of a positive function with respect to a nondecreasing one, it is monotone nondecreasing in both variables. One of the main undesirable features of this approach is that only uses the information provided by the uncensored observations $(T_1 < C_1, T_2 < C_2)$, throwing away the relevant information of censored data points $(T_1 < C_1, T_2 > C_2)$ or $(T_1 > C_1)$.

- Dabrowska (1988) expresses the bivariate survival function in terms of the joint distribution of the observable variables. She develops an empirical estimator which is such that the marginals are given by the univariate Kaplan-Meier, in the absence of censoring reduces to the empirical survival function and is almost sure consistent. The undesirable property is that it throws important part of the data away.

Dabrowska defines a bivariate cumulative hazard function (a vector function $\Lambda(t_1, t_2) = (\Lambda_{10}(t_1, t_2), \Lambda_{01}(t_1, t_2), \Lambda_{11}(t_1, t_2)))$ and shows that it determines the bivariate survival function. Furthermore she shows that identifiability of S will follow if the bivariate hazard function can be expressed in terms of joint distribution of the observables. Under the assumption of independent bivariate censoring, she proves that S is identifiable on the support of the bivariate survival function of the observables.

The corresponding instantaneous bivariate hazard has as well three components: $\lambda(t_1, t_2) = (\lambda_{10}(t_1, t_2), \lambda_{01}(t_1, t_2), \lambda_{11}(t_1, t_2))$ which represent, respectively, the instantaneous rate of a single failure at time t_1 given that $(T_1 \geq t_1, T_2 > t_2)$, the instantaneous rate of a single failure at time t_2 given that $(T_1 > t_1, T_2 \geq t_2)$ and the instantaneous rate of a double failure at point (t_1, t_2) given that individuals were alive at times $T_1 = t_1-$ and $T_2 = t_2-$, that is given that $(T_1 \geq t_1, T_2 \geq t_2)$.

Dabrowska's estimators admits the following expression

$$\hat{S}(t_1, t_2) \;=\; \hat{S}_1(t_1)\hat{S}_2(t_2) \prod_{0<u\leq t_1, 0<v\leq t_2} [1 - \hat{L}(\Delta u, \Delta v)]$$

where $\hat{S}_1(t_1)$ and $\hat{S}_2(t_2)$ are the corresponding univariate Kaplan-Meier estimators and $\hat{L}$ is function of the empirical counterparts of $(\Lambda_{10}(t_1, t_2), \Lambda_{01}(t_1, t_2), \Lambda_{11}(t_1, t_2))$.

The Dabrowska estimator overcomes some of the undesirable features of other proposals: nonuniqueness, inconsistency and lack of weak convergence and for these reasons is to be preferred. Unfortunately, its computation is quite complicated and the covariance function of these estimators cannot be estimated analytically.

- Prentice and Cai (1992), while approaching the problem of the characterization of the dependence between T_1 and T_2 using an estimator for the covariance function for T_1 and T_2, propose a new estimator for the bivariate survival function. They provide a representation for S in terms of the marginal survivor functions and a differential function which can be easily estimated:

$$S(t_1, t_2) \;=\; S_1(t_1)S_2(t_2)L(t_1, t_2)$$

where $L(t_1, t_2)$ is function of estimable differential functions. Furthermore, they provide another representation allowing to generate joint survivor functions having specified marginal distributions and covariance function.

$$S(t_1, t_2) \quad = \quad S_1(t_1)S_2(t_2)[1 + \int_0^{t_1} \int_0^{t_2} \{S_1(u)S_2(v)\}^{-1} C(du, dv)]$$

- Lin and Ying (1993) provide a simple nonparametric estimator of the bivariate survival function under univariate censoring. It works under the assumption that $C = C_1 = C_2$ right-censors independently the random variables (T_1, T_2). The observed vector is (Y_1, Y_2, D_1, D_2) where $Y_j = \min\{T_j, C\}$ and $D_j = \mathbf{1}\{T_j \leq C\}$, $(j = 1, 2)$. Their estimator is based on the following representation:

$$S(t_1, t_2) \quad = \quad \text{Prob}\{Y_1 \geq t_1, Y_2 \geq t_2\}/G(\max(t_1, t_2)),$$

where G is the survival function for C and it follows straightforwardly from the independence between (T_1, T_2) and C. It is also natural to estimate $\text{Prob}\{Y_1 \geq t_1, Y_2 \geq t_2\}$ by the empirical survival function: $n^{-1} \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} \geq t_1, Y_{2i} \geq t_2\}$ and G can be estimated by the Kaplan-Meier estimator from the data (C_i^*, δ_i^c) where

$$C_i^* = \min\{C_i, \max(T_{1i}, T_{2i})\} = \max(Y_{1i}, Y_{2i})$$

and $\delta_i^c = \mathbf{1}\{C_i \leq \max(Y_{1i}, Y_{2i})\} = 1 - D_{1i}D_{2i}$.

This estimator is far simpler than the others, reduces to the empirical survival function in the absence of censoring, converges weakly to a zero-mean Gaussian process (properly normalized) and if the support of C contains the support of (T_1, T_2), we will have uniform consistency of the estimator over the entire support of S. Simulation studies suggest that Lin and Ying estimator behaves similarly to Prentice-Cai and Dabrowska estimator.

2.3 Nonparametric estimation of the bivariate distribution function under dependent right-censoring

We assume that an individual may experience two consecutive events at times T_1 and $T_1 + T_2$ which are measured from the start of the follow-up. Assume

as well that the follow-up time is subject to independent right censoring by C, which implies that T_2 is subject to right censoring by $C - T_1$, which is naturally correlated with T_1 unless T_1 is independent of T_2. The marginal distribution of T_2 cannot therefore be estimated by the Kaplan–Meier method, and neither can the joint distribution of (T_1, T_2) be estimated by any existing estimator for parallel events.

Furthermore, the bivariate distribution function $F(t_1, t_2)$ is not estimable if $t_1 + t_2 > \tau_C$ where $\tau_C = \sup\{t : G(t) > 0\}$. In other words, for any $T_1 = t_1 < \tau_C$, T_2 is only observable while $T_2 \leq \tau_C - t_1$, and there is no information in the data to estimate $T_2 > \tau_C - t_1$. This restriction is analogous to the non-estimability of the Kaplan-Meier estimators beyond those values larger than the total follow-up time.

The observed vector is (Y_1, Y_2, D_1, D_2) where $Y_j = \min\{T_j, C_j\}$, $D_j = \mathbf{1}\{T_j \leq C_j\}$, $C_1 = C$ and $C_2 = (C_1 - T_1)\mathbf{1}\{T_1 \leq C_1\}$. A different way of representing the observables is via (Y_1, Y_2, δ) where

$$\delta = \begin{cases} 1 & \text{if} \quad T_1 > C \\ 2 & \text{if} \quad T_1 \leq C < T_1 + T_2 \\ 3 & \text{if} \quad T_1 + T_2 \leq C. \end{cases}$$

Note that when

1. $\delta = 1 \iff D_1 = 0 = D_2$: the two durations are right–censored and thus $Y_1 = C$, $Y_2 = 0$ and no information about T_2 is available

2. $\delta = 2 \iff D_1 = 1, D_2 = 0$, T_1 is observed while T_2 is right–censored by $C - T_1$, which implies that T_2 is right–censored by a dependent variable if T_1 and T_2 are correlated.

3. $\delta = 3 \iff D_1 = 1, D_2 = 1$, T_1 and T_2 are observed.

Assume that we have a random sample $\{(T_{1i}, T_{2i}, C_i), i = 1, \ldots, n\}$ of (T_1, T_2, C) from which the observed data is $\{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}), i = 1, \ldots, n\}$.

In what follows we review three approaches: Visser's approach which is valid under the assumption of discrete times and which is extended in subsection 3.1 to the case of two non consecutive times, Wang and Wells's weighting approach, similar to the original way of defining the well–known Kaplan–Meier proposed by Satten and Datta (2001) as an inverse probability of censoring weighted average, and Ling, Sun and Ying which is a natural extension of the univariate Kaplan–Meier method.

- Visser (1996) proposes a nonparametric estimator for the bivariate survival function when the two duration variables are successive, are always observed in a particular order, and the censoring mechanisms acts on their sum.

  Visser starts assuming that T_1, T_2 and C are discrete random variables taking values in $\{0, 1, 2, \ldots, K\}$, and therefore Y_1, Y_2 are discrete as well. We introduce the following notation for the bivariate survival function of (T_1, T_2), the survival function of T_1, the conditional survival function of T_2 given $T_1 \geq k$, as well as for the hazard function of T_1, the conditional hazard function of T_2 given $T_1 \geq k$ and the survival function of C:

$$\begin{aligned}
S_{T_1,T_2}(k,l) &= \text{Prob}\{T_1 \geq k, T_2 \geq l\} \\
S_{T_1}(k) &= \text{Prob}\{T_1 \geq k\} \\
S_{T_2|T_1}(l|k) &= \text{Prob}\{T_2 \geq l | T_1 \geq k\} \\
S_{T_2|T_1=k}(l) &= \text{Prob}\{T_2 \geq l | T_1 = k\} \\
\lambda_{T_1}(k) &= \text{Prob}\{T_1 = k | T_1 \geq k\} \\
\lambda_{T_2|T_1}(l|k) &= \text{Prob}\{T_2 = l | T_1 \geq k, T_2 \geq l\} \\
\lambda_{T_2|T_1=k}(l) &= \text{Prob}\{T_2 = l | T_1 = k, T_2 \geq l\} \\
S_C(k) &= \text{Prob}\{C \geq k\}
\end{aligned}$$

  Note that, due to the fact that the random variables T_1, T_2 and C are supposed to be discrete and finite, Visser defines the corresponding survival distributions at each time t as the probability of being greater or equal than t. It does not mean lose of generality for the subsequent results.

  Two different conditional survival functions are defined, namely $S_{T_2|T_1}(l|k)$ and $S_{T_2|T_1=k}(l)$, and the corresponding hazard functions, namely $\lambda_{T_2|T_1}(l|k)$ and $\lambda_{T_2|T_1=k}(l)$.

  To achieve the main goal, that is the estimation of $S_{T_1,T_2}(k,l)$, he uses the fact that the joint survival can be factorized in terms of the marginal and the conditional survival functions. The following equalities are relevant:

$$S_{T_1,T_2}(k,l) = S_{T_1}(k) S_{T_2|T_1}(l|k) \tag{4}$$

Furthermore, he takes advantage of the product limit factorization of the survival in terms of the hazard functions. For $k, l = 1, 2, \ldots, K$:

$$\begin{aligned} S_{T_1}(k) &= (1 - \lambda_{T_1}(0)) \ldots (1 - \lambda_{T_1}(k-1)) \quad (5) \\ S_{T_2|T_1=k}(l) &= (1 - \lambda_{T_2|T_1=k}(0)) \ldots (1 - \lambda_{T_2|T_1=k}(l-1)) \quad (6) \end{aligned}$$

It is also worth noticing that $S_{T_2|T_1}(l|k)$ can be written as follows:

$$\begin{aligned} S_{T_2|T_1}(l|k) &= \text{Prob}\{T_2 \geq l | T_1 \geq k\} = \frac{\text{Prob}\{T_2 \geq l, T_1 \geq k\}}{\text{Prob}\{T_1 \geq k\}} \\ &= (S_{T_1}(k))^{-1} \sum_{j=k}^{K} \text{Prob}\{T_2 \geq l | T_1 = j\} \text{Prob}\{T_1 = j\} \\ &= (S_{T_1}(k))^{-1} \sum_{j=k}^{K} S_{T_2|T_1=j}(l) \text{Prob}\{T_1 = j\} \\ &= (S_{T_1}(k))^{-1} \sum_{j=k}^{K} S_{T_2|T_1=j}(l)(S_{T_1}(j) - S_{T_1}(j+1)) \quad (7) \end{aligned}$$

Equalities (4) and (7) imply that in order to estimate $S_{T_1,T_2}(k, l)$ we only need to estimate $S_{T_1}(k)$ and $S_{T_2|T_1=j}(l)$. The estimation of $S_{T_1}(k)$ is straightforward though the Kaplan-Meier estimator.

In general estimators for $\lambda_{T_1}(k)$ and for $\lambda_{T_2|T_1=k}(l)$ can be derived based on the observations $(Y_{1i}, Y_{2i}, \delta_i)$ $(i = 1, 2, \ldots, n)$.

Denote by n_{1k}, n_{2kl}, n_{3kl} the following counting processes:

$$\begin{aligned} n_{1k} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, \delta_i = 1\} \\ n_{2kl} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} = l, \delta_i = 2\} \\ n_{3kl} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} = l, \delta_i = 3\}, \end{aligned}$$

that is, n_{1k} counts the number of censored individuals at k months (for these individuals $T_1 > k$ and T_2 is not defined), n_{2kl} counts the number of patients whose first duration is equal to k months and who are censored after $k+l$ months (for these individuals $T_1 = k$ and $T_2 > l$)

and n_{3kl} counts the number of patients with a first duration equal to k months and a second duration equal to l months (for these individuals $T_1 = k$ and $T_2 = l$). Denote as well $n_{k\cdot} = \sum_{l=1}^{K}(n_{2kl} + n_{3kl})$ which counts the total number of individuals whose $T_1 = k$ irrespective of their status on T_2.

The likelihood for the n observations is as follows:

$$\begin{aligned} L &= \prod_{i=1}^{n} \Big\{ \text{Prob}\{T_1 > y_{1i}, C = y_{1i}\}^{\mathbf{1}\{\delta_i=1\}} \\ &\quad \text{Prob}\{T_1 = y_{1i}, T_2 > y_{2i}, C - y_{1i} = y_{2i}\}^{\mathbf{1}\{\delta_i=2\}} \\ &\quad \text{Prob}\{T_1 = y_{1i}, T_2 = y_{2i}, C > y_{1i} + y_{2i}\}^{\mathbf{1}\{\delta_i=3\}} \Big\} \end{aligned}$$

and the corresponding log likelihood looks like as

$$\begin{aligned} \mathcal{L} &= \log L = \sum_{i=1}^{n} \{\mathbf{1}\{\delta_i = 1\} \log \text{Prob}\{T_1 > y_{1i}, C = y_{1i}\} \\ &+ \mathbf{1}\{\delta_i = 2\} \log \text{Prob}\{T_1 = y_{1i}, T_2 > y_{2i}, C - y_{1i} = y_{2i}\} \\ &+ \mathbf{1}\{\delta_i = 3\} \log \text{Prob}\{T_1 = y_{1i}, T_2 = y_{2i}, C > y_{1i} + y_{2i}\}\}. \end{aligned}$$

Since our variables are discrete, the possible values for y_{1i}, y_{2i} and C are only $\{0, 1, 2, \ldots, K\}$, thus $\mathcal{L}$ can be rewritten as

$$\begin{aligned} \mathcal{L} &= \sum_{k=1}^{K} n_{1k} \log \text{Prob}\{T_1 > k, C = k\} \\ &+ \sum_{k=1}^{K} \sum_{l=1}^{K} n_{2kl} \log \text{Prob}\{T_1 = k, T_2 > l, C - k = l\} \\ &+ \sum_{k=1}^{K} \sum_{l=1}^{K} n_{3kl} \log \text{Prob}\{T_1 = k, T_2 = l, C > k + l\}. \end{aligned}$$

Factorizing each probability into the conditional probabilities we can rewrite:

$$\mathcal{L} = \sum_{k=1}^{K} n_{1k} \log \text{Prob}\{T_1 > k | C = k\} + \sum_{k=1}^{K} n_{1k} \log \text{Prob}\{C = k\} +$$

$$\sum_{k=1}^{K}\sum_{l=1}^{K} n_{2kl}\log \text{Prob}\{T_2 > l | T_1 = k, C - k = l\} + \sum_{k=1}^{K}\sum_{l=1}^{K} n_{2kl}\log \text{Prob}\{T_1 = k | C - k = l\}$$
$$\sum_{k=1}^{K}\sum_{l=1}^{K} n_{2kl}\log \text{Prob}\{C - k = l\} + \sum_{k=1}^{K}\sum_{l=1}^{K} n_{3kl}\log \text{Prob}\{T_2 = l | T_1 = k, C > k + l\} +$$
$$\sum_{k=1}^{K}\sum_{l=1}^{K} n_{3kl}\log \text{Prob}\{T_1 = k | C > k + l\} + \sum_{k=1}^{K}\sum_{l=1}^{K} n_{3kl}\log \text{Prob}\{C > k + l\}$$

and since C is independent of (T_1, T_2), it follows

$$\begin{aligned}
\mathcal{L} &= \sum_{k=1}^{K} n_{1k}(\log \text{Prob}\{T_1 > k\} + \log \text{Prob}\{C = k\}) + \\
&\sum_{k=1}^{K}\sum_{l=1}^{K} n_{2kl}(\log \text{Prob}\{T_2 > l | T_1 = k\} + \log \text{Prob}\{T_1 = k\} + \log \text{Prob}\{C - k = l\}) + \\
&\sum_{k=1}^{K}\sum_{l=1}^{K} n_{3kl}(\log \text{Prob}\{T_2 = l | T_1 = k\} + \log \text{Prob}\{T_1 = k\} + \log \text{Prob}\{C > k + l\}) \\
&= \sum_{k=1}^{K}\{n_{k\cdot}\log \text{Prob}\{T_1 = k\} + n_{1k}\log \text{Prob}\{T_1 > k\}\} + \qquad (8) \\
&\sum_{k=1}^{K}\sum_{l=1}^{K}\{n_{3kl}\log \text{Pr}\{T_2 = l | T_1 = k\} + n_{2kl}\log \text{Pr}\{T_2 > l | T_1 = k\}\} + \qquad (9) \\
&\sum_{k=1}^{K}\{n_{1k}\log \text{Pr}\{C = k\} + \sum_{l=1}^{K}\{n_{2kl}\log \text{Pr}\{C = k + l\} + n_{3kl}\log \text{Pr}\{C > k + l\}\}\} \qquad (10) \\
= \;& \mathcal{L}_{T_1} + \mathcal{L}_{T_2|T_1} + \mathcal{L}_C.
\end{aligned}$$

All the expressions for the probabilities can be replaced by functions containing uniquely $\lambda_{T_1}(k)$ and to $\lambda_{T_2|T_1=k}(l)$. For instance,

$$\begin{aligned}
\mathcal{L}_{T_1} &= \sum_{k=1}^{K}\{n_{k\cdot}\log \text{Prob}\{T_1 = k\} + n_{1k}\log \text{Prob}\{T_1 > k\}\} \\
&= \sum_{k=1}^{K}\{n_{k\cdot}\log \lambda_{T_1}(k) S_{T_1}(k) + n_{1k}\log S_{T_1}(k+1)\} \\
&= \sum_{k=1}^{K}\{n_{k\cdot}(\log \lambda_{T_1}(k) + \log \prod_{j=0}^{k-1}(1 - \lambda_{T_1}(j))) + n_{1k}\log \prod_{j=0}^{k}(1 - \lambda_{T_1}(j))\}
\end{aligned}$$

$$= \sum_{k=1}^{K} n_{k\cdot} \log \lambda_{T_1}(k) + \sum_{k=1}^{K} (n_{1k} + n_{k\cdot}) \sum_{j=0}^{K-1} \log(1 - \lambda_{T_1}(j)) + \sum_{k=1}^{K} n_{1k} \log(1 - \lambda_{T_1}(k)).$$

The nonparametric estimators for the hazard functions are obtained after maximizing the log likelihood $\mathcal{L} = \mathcal{L}_{T_1} + \mathcal{L}_{T_2|T_1} + \mathcal{L}_C$. Note that we are in fact maximizing $\log L$ with respect to $\lambda_{T_1}(k)$ and to $\lambda_{T_2|T_1=k}(l)$, and because the terms act additively we can maximize first with respect to $\lambda_{T_1}(k)$ and then with respect to $\lambda_{T_2|T_1=k}(l)$.

Visser proves that the nonparametric MLE for $\lambda_{T_1}(k)$ is given by

$$\hat{\lambda}_{T_1}(k) = \frac{\sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, \delta_i \geq 2\}}{\sum_{i=1}^{n} \mathbf{1}\{Y_{1i} \geq k\}} = \frac{\sum_{l=1}^{K} (n_{2kl} + n_{3kl})}{n_{1k} + \sum_{l=1}^{K} (n_{2kl} + n_{3kl})} \qquad (11)$$

which yields the discrete time Kaplan-Meier estimator after replacing it in (5). On the other hand, the nonparametric MLE for $\lambda_{T_2|T_1=k}(l)$ is given by

$$\hat{\lambda}_{T_2|T_1=k}(l) = \sum_{i=1}^{n} \frac{\mathbf{1}\{Y_{1i} = k, Y_{2i} = l, \delta_i = 3\}}{\sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} \geq l\}}. \qquad (12)$$

Replacing $\hat{\lambda}_{T_2|T_1=k}(l)$ in (6) provides the MLE for $S_{T_2|T_1=k}(l)$, which in turn can be replaced in (7) to obtain an estimator for $S_{T_2|T_1}(l|k)$. Finally everything could be replaced in (4) to get the bivariate nonparametric estimator for $S_{T_1,T_2}(k,l)$.

Visser proves that both estimators, $\hat{\lambda}_{T_1}(k)$ and $\hat{\lambda}_{T_2|T_1=k}(l)$, are consistent and asymptotically normal after normalizing by $\sqrt{n}$. He also claims that both estimators are asymptotically independent. These facts, together with the δ method, allow him to prove that $\sqrt{n}(\hat{S}_{T_1,T_2}(k,l) - S_{T_1,T_2}(k,l))$ is asymptotically normal, mean zero and with an asymptotic variance that can be estimated replacing the unknown functions by their estimators.

The survival function S_C of the censoring variable appears in the expression for the variances. It may be estimated by the product-limit method.

- Wang and Wells (1998) propose a path-dependent (nonparametric) estimate for the joint survival function of two duration variables.

Denote the observed sample by $\mathcal{S} = \{(Y_{1i}, Y_{2i}, D_{1i}, D_{2i}), i = 1, \ldots, n\}$. Consider a subset of $\mathcal{S}$ consisting of observations for which T_1 is observed, that is $\mathcal{S}^* = \{(Y_{1i}, Y_{2i}, D_{1i} = 1, D_{2i}), i = 1, \ldots, n\}$.

Following the notation introduced in subsection 2.1, they consider the following path-dependent decomposition

$$\begin{aligned} S(t_1, t_2) &= \text{Prob}\{T_2 > t_2 | T_1 > t_1\}\text{Prob}\{T_1 > t_1\} \\ &= \prod_{v \leq t_2} \{1 - \Lambda_{T_2|T_1>t_1}(dv)\} S_1(t_1) \end{aligned} \tag{13}$$

where $\Lambda_{T_2|T_1>t_1}(dv)$ is the cumulative conditional hazard of T_2 given $T_1 > t_1$. Wang and Wells propose to estimate $S(t_1, t_2)$ via estimable components for $\Lambda_{T_2|T_1>t_1}(dv)$ and for $S_1(t_1)$ and to plug them into (13).

The estimation of the marginal $S_1(t_1)$ is accomplished via the Kaplan–Meier estimator based on the observables (Y_{1i}, D_{1i}) $(i = 1, \ldots, n)$. Note that if T_1 is right–censored then $\delta_i = 1$ and $D_{1i} = 0$ while if T_1 is observed then $\delta_i = 2$ or 3 $(D_{1i} = 1)$.

The estimator for $\Lambda_{T_2|T_1>t_1}(dv)$ extends Campbell and Földes estimator so that dependent censoring is taking into account. First note that if we let $R_{T_2}(v|t_1)$ be the risk set of T_2 at time v given $T_1 > t_1$, if $v > 0$ then $R_{T_2}(v|t_1) \subset \mathcal{S}^*$. An observation i with the first duration $T_1 = t_{1i}$ affects the probability of the corresponding T_{2i} being included in $R_{T_2}(v|t_1)$ as we see in the following expression

$$\begin{aligned} \text{Prob}\{i \in R_{T_2}(v|t_1)\} &= \text{Prob}\{Y_{1i} \in t_{1i}, t_{1i} > t_1, D_{1i} = 1, Y_{2i} \geq v\} \\ &= \text{Prob}\{T_1 \in t_{1i}, t_{1i} > t_1, T_2 \geq v\}\text{Prob}\{C_1 > t_{1i} + v\} \\ &= \text{Prob}\{T_1 \in t_{1i}, t_{1i} > t_1, T_2 \geq v\} G_1(t_{1i} + v). \end{aligned}$$

Hence we will adjust this heterogeneity weighting each observation in $R_{T_2}(v|t_1)$ by an estimate of $1/G_1(t_{1i} + v)$.

Wang and Wells' estimator for $\Lambda_{T_2|T_1>t_1}(dv)$ can be expressed as follows:

$$\begin{aligned} \hat{\Lambda}_{T_2|T_1>t_1}(\Delta v) &= \frac{\sum_{i \in R_{T_2}(v|t_1)} \mathbf{1}\{Y_{2i} = v, D_{2i} = 1\}/\hat{G}_1(t_{1i} + v)}{\sum_{i \in R_{T_2}(v|t_1)} \mathbf{1}\{Y_{2i} \geq v\}/\hat{G}_1(t_{1i} + v)} \\ &= \frac{\sum_{i=1}^n \mathbf{1}\{\hat{G}_1(Y_{1i} + v) > 0\}\mathbf{1}\{Y_{1i} > t_1, D_{1i} = 1, Y_{2i} = v, D_{2i} = 1\}/\hat{G}_1(Y_{1i} + v)}{\sum_{i=1}^n \mathbf{1}\{\hat{G}_1(Y_{1i} + v) > 0\}\mathbf{1}\{Y_{1i} > t_1, D_{1i} = 1, Y_{2i} \geq v\}/\hat{G}_1(Y_{1i} + v)} \end{aligned} \tag{14}$$

where $\hat{G}_1(\cdot)$ is the Kaplan-Meier estimator of $G_1(\cdot)$ computed from the data $(Y_{1i} + Y_{2i}, 1 - D_{1i}D_{2i})$ $(i = 1, \ldots, n)$. Wang and Wells' estimator for $S(t_1, t_2)$ is given by

$$\hat{S}(t_1, t_2) = \prod_{v \leq t_2} \{1 - \hat{\Lambda}_{T_2|T_1 > t_1}(dv)\} \hat{S}_1(t_1)$$

Their estimator uses the information on the first duration to weight each observation to unbias the effect of dependent censoring. The marginal survivor function is estimated by $\hat{S}_2(t_2) = \hat{S}(0, t_2)$. This estimator has a potential problem with the existence of $\hat{S}(t_1, t_2)$ when $\hat{G}_1(\cdot) = 0$. If the largest value of $Y_{1i} + Y_{2i}$, say $c_{(n)}$, is censored ($D_{1i}D_{2i} = 0$ or $\delta_i = 3$), then the largest observation of the censoring variables is observed ($1 - D_{1i}D_{2i} = 1$) and hence $\hat{G}_1(c_{(n)}) = 0$.

Wang and Wells show that $\hat{S}(t_1, t_2)$ converges in probability to $S(t_1, t_2)$ and claim that the limit distribution of $\sqrt{n}(\hat{S}(t_1, t_2) - S(t_1, t_2))$ converges weakly to a zero-mean Gaussian process, but the variance of the limiting process is quite complex and is not given.

- Lin, Sun and Ying (1999) propose an estimator for the bivariate distribution $F(t_1, t_2)$ based on the relationship $F(t_1, t_2) = L(t_1, 0) - L(t_1, t_2)$, where $L(t_1, t_2) = \text{Prob}\{T_1 \leq t_1, T_2 > t_2\}$. The estimator for L is defined as

$$\hat{L}(t_1, t_2) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}\{Y_{1i} \leq t_1, Y_{2i} > t_2\}}{\hat{G}_1(Y_{1i} + t_2)}$$

where $\hat{G}_1(\cdot)$ is the Kaplan-Meier estimator of $G_1(\cdot)$ computed from the data $(Y_{1i}, 1 - D_{1i})$ or $(Y_{2i}, 1 - D_{2i})$ $(i = 1, \ldots, n)$. $\hat{L}(\cdot, \cdot)$ would be an unbiased estimator of $L(\cdot, \cdot)$ if G_1 were known. Hence the estimator for the bivariate distribution is defined *ad hoc* as

$$\hat{F}(t_1, t_2) = \hat{L}(t_1, 0) - \hat{L}(t_1, t_2)$$

and is as well unbiased if G_1 were known.

Note that for the estimation of $F(t_1, t_2)$ when $t_2 < \infty$ only the observations for which T_1 is observed are considered, that is only the

observations inside the subset $\mathcal{S}^*$, defined as in Wang and Wells, will be used for the computations. However, if we are interested in the marginal $F_1(t_1) = F(t_1, \infty)$, the estimator

$$\hat{F}(t_1, \infty) = \hat{L}(t_1, 0) = \frac{1}{n} \sum_{i=1}^{n} \frac{D_{1i} \mathbf{1}\{Y_{1i} \leq t_1\}}{\hat{G}_1(Y_{1i})}$$

is identical to the Kaplan-Meier estimator of F_1 if $\hat{G}_1(\cdot)$ is the Kaplan-Meier estimator of $G_1(\cdot)$ computed from $(Y_{1i}, 1 - D_{1i})$ $(i = 1, \ldots, n)$.

Lin, Sun and Ying's estimator reduces to the usual empirical distribution function in the absence of censoring, is a natural extension of the univariate Kaplan-Meier approach, is strongly consistent and the process $\sqrt{n}(\hat{F}(\cdot, \cdot) - F(\cdot, \cdot))$ converges weakly to a bivariate zero-mean Gaussian process. The authors give an explicit expression, as well as a consistent estimator, for the covariance function. This estimator, however, is not always a proper distribution since it can yield negative mass points.

2.4 Average relative risk dependence measure

Once we have an estimator for the joint survival for (T_1, T_2), it is of interest to have a measure which summarizes the degree of dependency or concordance level between T_1 and T_2. To this end several proposals can be made.

Clayton (1978) and Oakes (1989) propose the cross ratio as a measure of local dependence at (s_1, s_2), that is,

$$\begin{aligned} c^*(s_1, s_2) &= \frac{F(ds_1, ds_2) F(s_1^-, s_2^-)}{F(ds_1, s_2^-) F(s_1^-, ds_2)} \\ &= \lambda_2(s_2 | T_1 = s_1) / \lambda_2(s_2 | T_1 \geq s_1), \end{aligned}$$

where $F(s_1, s_2) = \text{Prob}\{T_1 > s_1, T_2 > s_2\}$ is the joint survival function. The cross ratio is a type of hazard ratio or relative risk. An average of the cross ratio over a period of time can be computed as a summary.

Another local measure could be based on Kendall's coefficient of concordance.

Fan, Hsu and Prentice (2000) propose a summary dependence measure $C(t_1, t_2)$ based on the reciprocal of the cross ratio $c^*(s_1, s_2)$. The reason of using the reciprocal cross ratio is that it can be consistently estimated.

The average relative risk dependence measure over $[0,t_1]x[0,t_2]$ is defined as follows:

$$\begin{aligned} C(t_1,t_2) &= \frac{\int_0^{t_1}\int_0^{t_2}\frac{1}{c^*(s_1,s_2)}F(ds_1,ds_2)}{\int_0^{t_1}\int_0^{t_2}F(du_1,du_2)} \\ &= \frac{\int_0^{t_1}\int_0^{t_2}\frac{F(ds_1,s_2^-)F(s_1^-,ds_2)}{F(ds_1,ds_2)F(s_1^-,s_2^-)}F(ds_1,ds_2)}{1-F(t_1,0)-F(0,t_2)+F(t_1,t_2)} \\ &= \frac{\int_0^{t_1}\int_0^{t_2}F(s_1^-,s_2^-)\Lambda_{10}(ds_1,s_2^-)\Lambda_{01}(s_1^-,ds_2)}{1-F(t_1,0)-F(0,t_2)+F(t_1,t_2)} \end{aligned}$$

where $\Lambda_{10}(ds_1,s_2^-) = -F(ds_1,s_2^-)/F(s_1^-,s_2^-)$ and $\Lambda_{01}(s_1^-,ds_2) = -F(s_1^-,ds_2)/F(s_1^-,s_2^-)$ are the failure hazards at (t_1,t_2).

Under independent random censorship the measure $C(t_1,t_2)$ can be estimated nonparametrically by

$$\hat{C}(t_1,t_2) = \frac{\int_0^{t_1}\int_0^{t_2}\hat{F}(s_1^-,s_2^-)\hat{\Lambda}_{10}(ds_1,s_2^-)\hat{\Lambda}_{01}(s_1^-,ds_2)}{\{1-\hat{F}(t_1,0)-\hat{F}(0,t_2)+\hat{F}(t_1,t_2)\}}$$

where

- $\hat{F}$ is a strongly consistent nonparametric estimator of F. Both Dabrowska (1988) and Prentice and Cai (1992) propose such a $\hat{F}$.
- $\hat{\Lambda}_{10}$ and $\hat{\Lambda}_{01}$ are the Nelson-Aalen counterparts of Λ_{10} and Λ_{01}.
- Under weak conditions, $\hat{C}(t_1,t_2)$ is strongly consistent for $C(t_1,t_2)$ and $K^{1/2}(\hat{C}(t_1,t_2)-C(t_1,t_2))$ converges in distribution to a mean zero Gaussian process, where K is the sample size.
- The covariance of $\hat{C}$ is very complicated due to the complexity of $\hat{F}$.

$C(t_1,t_2)$ can be interpreted as a weighted average of the hazard ratio $\frac{1}{c^*(s_1,s_2)}$ with weight proportional to the failure time density at (t_1,t_2). Also $C(t_1,t_2)=1$ if T_1 and T_2 are independent, while if they are locally positively dependent $0 \le C(t_1,t_2) < 1$ and if they are locally negatively dependent $C(t_1,t_2) > 1$.

Another meaningful summary dependence measure would be $C^{-1}(t_1,t_2)$ since its range of values for positive and negative dependence and independence agrees with the usual cross ratio local dependency measure.

3 Extensions of the bivariate survival estimator

3.1 Estimation of the bivariate survival function for two not consecutive gap times. A discrete approach

Consider the situation in which individuals can experience at most three events. Here we focus our interest on the successive times, T_1, T_2 and T_3, between these three events. It is convenient to define three different stages as follows: An individual is in Stage 1 while he does not experienced the first event, is in Stage 2 after he/she has experienced the first event and before the second one, and is in Stage 3 if he/she has experienced the fisrts and second events and before the third one.

Due to the possible different nature of T_1 and T_3 compared to T_2 we start focusing our attention on the joint behaviour of (T_1, T_3), taking into account T_2, and therefore looking at the distribution of (T_1, T_3) as the marginal bivariate distribution of (T_1, T_2, T_3).

We generalize Visser's nonparametric estimator for the bivariate survival function which takes into account the dependent censoring implied by the successive durations. As in Visser we assume that T_1, T_2, T_3 and C are discrete random variables taking values in $\{0, 1, 2, \ldots, K\}$, and therefore Y_1, Y_2 and Y_3 are discrete as well.

We introduce the notation $S_{123}(a, b, c)$ for the joint survival function of (T_1, T_2, T_3) and $S_{jk}(a, b)$ for the bivariate survival function of (T_j, T_k) $(1 \leq j, k \leq 3)$, that is,

$$\begin{aligned} S_{123}(a, b, c) &= \text{Prob}\{T_1 \geq a, T_2 \geq b, T_3 \geq c\}, \\ S_{jk}(a, b) &= \text{Prob}\{T_j \geq a, T_k \geq b\}. \end{aligned}$$

The conditional survival functions for any subset of random variables conditioned to any other subset can be defined straightforwardly. Due to the nature of our random variables, we will only be interested in conditioning to the past, that is T_3 given T_2 and T_1, T_3 given T_1, etc. For instance, the conditional survival function of T_3 given $T_1 \geq a$,

$$S_{T_3|T_1}(c|a) = \text{Prob}\{T_3 \geq c|T_1 \geq a\}.$$

We also define the survival functions of T_3 conditioned to $(T_2 = b, T_1 = a)$ and of T_2 conditioned to $T_1 = a$ as follows

$$\begin{aligned} S_{T_3|T_2=b,T_1=a}(c) &= \text{Prob}\{T_3 \geq c | T_2 = b, T_1 = a\} \\ S_{T_2|T_1=a}(b) &= \text{Prob}\{T_2 \geq b | T_1 = a\}. \end{aligned}$$

Analogously we define the following hazards and conditional hazards and the survival function of C as follows:

$$\begin{aligned} \lambda_{T_1}(a) &= \text{Prob}\{T_1 = a | T_1 \geq a\} \\ \lambda_{T_2|T_1}(b|a) &= \text{Prob}\{T_2 = b | T_1 \geq a, T_2 \geq b\} \\ \lambda_{T_3|T_2,T_1}(c|b,a) &= \text{Prob}\{T_3 = c | T_1 \geq a, T_2 \geq b, T_3 \geq c\} \\ \lambda_{T_2|T_1=a}(b) &= \text{Prob}\{T_2 = b | T_1 = a, T_2 \geq b\} \\ \lambda_{T_3|T_2=b,T_1=a}(c) &= \text{Prob}\{T_3 = c | T_1 = a, T_2 = b, T_3 \geq c\} \\ S_C(k) &= \text{Prob}\{C \geq k\} \end{aligned}$$

Note that the following equalities are straightforward:

$$S_{T_1,T_3}(a,c) = S_{T_1}(a) S_{T_3|T_1}(c|a) \tag{15}$$

$$S_{T_3|T_1}(c|a) = \text{Prob}\{T_3 \geq c | T_1 \geq a\} = \sum_{k=0}^{K} \text{Prob}\{T_3 \geq c, T_2 = k | T_1 \geq a\} \tag{16}$$

$$= (S_{T_1}(a))^{-1} \cdot \sum_{k=0}^{K} \sum_{j=a}^{K} \text{Prob}\{T_3 \geq c | T_2 = k, T_1 = j\} \text{Prob}\{T_2 = k | T_1 = j\} \text{Prob}\{T_1 = j\}$$

$$= (S_{T_1}(a))^{-1} \cdot \sum_{k=0}^{K} \sum_{j=a}^{K} S_{T_3|T_2=k,T_1=j}(c)(S_{T_2|T_1=j}(k) - S_{T_2|T_1=j}(k+1))(S_{T_1}(j) - S_{T_1}(j+1))$$

$$S_{T_1}(a) = (1 - \lambda_{T_1}(0)) \ldots (1 - \lambda_{T_1}(a-1)) \tag{17}$$

Equalities (15) and (16) show that to provide an estimator for $S_{T_1,T_3}(a,c)$ we only need to take care of the estimation of $S_{T_1}(j)$, $S_{T_2|T_1=j}(k)$ and $S_{T_3|T_2=k,T_1=j}(c)$ for $a \leq j \leq K+1$ and $0 \leq k \leq K+1$.

By definition we will use

$$S_{T_1}(0) = 1 = S_{T_2|T_1=j}(0) = S_{T_3|T_2=k,T_1=j}(0)$$

and

$$S_{T_1}(K+1) = 0 = S_{T_2|T_1=j}(K+1) = S_{T_3|T_2=k,T_1=j}(K+1)$$

On the other hand the product decomposition (17) of the survival function in terms of the hazards implies that to derive an estimator for $S_{T_1,T_3}(a,c)$, $0 \leq a,c \leq K$, it is only necessary to estimate $\lambda_{T_1}(j)$, $\lambda_{T_2|T_1=j}(k)$ and $\lambda_{T_3|T_2=k,T_1=j}(c)$ for $a \leq j \leq K$ and $0 \leq k \leq K$.

Denote by $n_{1k}, n_{2kl}, n_{3klm}, n_{4klm}$, $0 \leq k,l,m \leq K$ the number of observations for which $(Y_1 = k, \delta = 1)$, $(Y_1 = k, Y_2 = l, \delta = 2)$, $(Y_1 = k, Y_2 = l, Y_3 = m, \delta = 3)$, $(Y_1 = k, Y_2 = l, Y_3 = m, \delta = 4)$ respectively. That is, n_{1k} counts how many individuals in Stage 1 are censored at k months, n_{2kl} counts how many individuals in Stage 2 have stayed k units of time in the first period and l units of time in the second period being censored at the end of the study. n_{3klm} counts how many individuals in Stage 3 have stayed k units of time in the first period, l units of time in the second period and m units of time in the third one being censored at the end of the study, and finally n_{4klm} counts how many individuals in Stage 4 have stayed k units of time in the first period, l units of time in the second period and exactly m units of time in the third one, before the end of the study.

Summarizing,

$$\begin{aligned}
n_{1k} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, \delta_i = 1\} \\
n_{2kl} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} = l, \delta_i = 2\} \\
n_{3klm} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} = l, Y_{3i} = m, \delta_i = 3\} \\
n_{4klm} &= \sum_{i=1}^{n} \mathbf{1}\{Y_{1i} = k, Y_{2i} = l, Y_{3i} = m, \delta_i = 4\}
\end{aligned}$$

The likelihood of the n observations is as follows:

$$\begin{aligned}
L &= L_1 L_2 L_3 L_4 \\
&= \prod_{i=1}^{n} \Big\{ \mathrm{Prob}\{T_1 > y_{1i}, C = y_{1i}\}^{\mathbf{1}\{\delta_i=1\}} \\
&\quad \mathrm{Prob}\{T_1 = y_{1i}, T_2 > y_{2i}, C - y_{1i} = y_{2i}\}^{\mathbf{1}\{\delta_i=2\}} \\
&\quad \mathrm{Prob}\{T_1 = y_{1i}, T_2 = y_{2i}, T_3 > y_{3i}, C - (y_{1i} + y_{2i}) = y_{3i}\}^{\mathbf{1}\{\delta_i=3\}} \\
&\quad \mathrm{Prob}\{T_1 = y_{1i}, T_2 = y_{2i}, T_3 = y_{3i}, C > y_{1i} + y_{2i} + y_{3i}\}^{\mathbf{1}\{\delta_i=4\}} \Big\}
\end{aligned}$$

Denote by $\mathcal{L}$ the log likelihood, by $\mathcal{L}_j = \log L_j$ $(j = 1, \ldots, 4)$ and by $\mathrm{logP} = \log \mathrm{Prob}$.

All the expressions for the probabilities can be replaced by functions containing uniquely $\lambda_{T_1}(a)$, $\lambda_{T_2|T_1=k}(l)$, $\lambda_{T_3|T_2=l,T_1=k}(m)$ and $S_C(a)$. Indeed,

$$\begin{aligned}
\mathcal{L}_1 &= \sum_{i=1}^{n} \mathbf{1}\{\delta_i = 1\} \log \mathrm{P}\{T_1 > y_{1i}, C = y_{1i}\} \\
&= \sum_{k=0}^{K} n_{1k} (\log \mathrm{P}\{T_1 > k\} + \log \mathrm{P}\{C = k\}) \\
&= \sum_{k=0}^{K} n_{1k} (\log \prod_{a=0}^{k} (1 - \lambda_{T_1}(a)) + \log \mathrm{P}\{C = k\}) \\
&= \sum_{k=0}^{K} n_{1k} \left(\sum_{a=0}^{k} \log(1 - \lambda_{T_1}(a)) + \log \mathrm{P}\{C = k\} \right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_2 &= \sum_{i=1}^{n} \mathbf{1}\{\delta_i = 2\} \log \mathrm{P}\{T_1 = y_{1i}, T_2 > y_{2i}, C - y_{1i} = y_{2i}\} \\
&= \sum_{k=0}^{K} \sum_{l=0}^{K} n_{2kl} (\log \mathrm{P}\{T_2 > l | T_1 = k\} + \log \mathrm{P}\{T_1 = k\} + \log \mathrm{P}\{C = k + l\}) \\
&= \sum_{k=0}^{K} \sum_{l=0}^{K} n_{2kl} (\log \prod_{b=0}^{l} (1 - \lambda_{T_2|T_1=k}(b)) + \log \lambda_{T_1}(k) + \log S_{T_1}(k) + \log \mathrm{P}\{C = k + l\}) \\
&= \sum_{k=0}^{K} \sum_{l=0}^{K} n_{2kl} \left(\sum_{b=0}^{l} \log(1 - \lambda_{T_2|T_1=k}(b)) + \log \lambda_{T_1}(k) \right. \\
&+ \left. \sum_{a=0}^{k-1} \log(1 - \lambda_{T_1}(a)) + \log \mathrm{P}\{C = k + l\} \right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_3 &= \sum_{i=1}^{n} \mathbf{1}\{\delta_i = 3\} \log \mathrm{P}\{T_1 = y_{1i}, T_2 = y_{2i}, T_3 > y_{3i}, C - (y_{1i} + y_{2i}) = y_{3i}\} \\
&= \sum_{k=0}^{K} \sum_{l=0}^{K} \sum_{m=0}^{K} n_{3klm} (\log \mathrm{P}\{T_3 > m | T_1 = k, T_2 = l\} + \log \mathrm{P}\{T_2 = l | T_1 = k\} + \log \mathrm{P}\{T_1 = k\} \\
&+ \log \mathrm{P}\{C = k + l + m\}) \\
&= \sum_{k=0}^{K} \sum_{l=0}^{K} \sum_{m=0}^{K} n_{3klm} (\log \prod_{c=0}^{m} (1 - \lambda_{T_3|T_1=k,T_2=l}(c)) + \log \lambda_{T_2|T_1=k}(l) + \log S_{T_2|T_1=k}(l) \\
&+ \log \lambda_{T_1}(k) + \log S_{T_1}(k) + \log \mathrm{P}\{C = k + l + m\})
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=0}^{K}\sum_{l=0}^{K}\sum_{m=0}^{K} n_{3klm}\left(\sum_{c=0}^{m}\log(1-\lambda_{T_3|T_1=k,T_2=l}(c)) + \log\lambda_{T_2|T_1=k}(l)\right.\\
&+ \sum_{b=0}^{l-1}(1-\lambda_{T_2|T_1=k}(b)) + \log\lambda_{T_1}(k)\\
&+ \left.\sum_{a=0}^{k-1}\log(1-\lambda_{T_1}(a)) + \log\mathrm{P}\{C=k+l+m\}\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_4 &= \sum_{i=1}^{n}\mathbf{1}\{\delta_i=4\}\log\mathrm{P}\{T_1=y_{1i},T_2=y_{2i},T_3=y_{3i},C\geq y_{1i}+y_{2i}+y_{3i}\}\\
&= \sum_{k=0}^{K}\sum_{l=0}^{K}\sum_{m=0}^{K} n_{4klm}\log\mathrm{P}\{T_3=m|T_1=k,T_2=l\} + \log\mathrm{P}\{T_2=l|T_1=k\}\\
&+ \log\mathrm{P}\{T_1=k\} + \log\mathrm{P}\{C\geq k+l+m\})\\
&= \sum_{k=0}^{K}\sum_{l=0}^{K}\sum_{m=0}^{K} n_{4klm}(\log\lambda_{T_3|T_1=k,T_2=l}(m) + \log S_{T_3|T_1=k,T_2=l}(m))\\
&+ \log\lambda_{T_2|T_1=k}(l) + \log S_{T_2|T_1=k}(l)\\
&+ \log\lambda_{T_1}(k) + \log S_{T_1}(k) + \log\mathrm{P}\{C\leq k+l+m\})\\
&= \sum_{k=0}^{K}\sum_{l=0}^{K}\sum_{m=0}^{K} n_{4klm}\left(\log\lambda_{T_3|T_1=k,T_2=l}(m) + \sum_{c=0}^{m-1}\log(1-\lambda_{T_3|T_1=k,T_2=l}(c))\right.\\
&+ \log\lambda_{T_2|T_1=k}(l) + \sum_{b=0}^{l-1}(1-\lambda_{T_2|T_1=k}(b))\\
&+ \left.\log\lambda_{T_1}(k) + \sum_{a=0}^{k-1}\log(1-\lambda_{T_1}(a)) + \log\mathrm{P}\{C\geq k+l+m\}\right)
\end{aligned}
$$

We can rewrite $\mathcal{L}$ grouping terms that only depend on λ_{T_1} or on $\lambda_{T_2|T_1=k}$ for $k=0,\ldots,K$ or on $\lambda_{T_3|T_1=k,T_2=l}$ for $k,l=0,\ldots,K$ or on C. Let denote by $\mathcal{L}_{T_1},\mathcal{L}_{T_2|T_1},\mathcal{L}_{T_3|T_1,T_2}$ and $\mathcal{L}_C$, the new terms, respectively, and we get

$$
\mathcal{L} = \mathcal{L}_{T_1} + \mathcal{L}_{T_2|T_1} + \mathcal{L}_{T_3|T_1,T_2} + \mathcal{L}_C \tag{18}
$$

Maximizing (18) for a fix time $t=0,\ldots,K$ we obtain

$$
\begin{aligned}
\hat{\lambda}_{T_1}(t) &= \frac{d_{1t}}{r_{1t}}\\
\hat{\lambda}_{T_2|T_1=k}(t) &= \frac{d_{2kt}}{r_{2kt}}, \quad k=0,\ldots,K
\end{aligned}
$$

$$\hat{\lambda}_{T_3|T_1=k,T_2=l}(t) \;=\; \frac{d_{3klt}}{r_{3klt}}, \quad k,l=0,\ldots,K \tag{19}$$

where $d_{\cdot}$ and $r_{\cdot}$ denote the number of events and the number of individuals at risk at time t, respectively, for the conditioned samples $T_1, T_2|T_1 = k$ and $T_3|T_1 = k, T_2 = l$.

Replacing estimators (19) in (16) and (17) we obtain estimates for $S_{T_3|T_1}(c|a)$ and $S_{T_1}(a)$, and consequently for $S_{13}(a,c)$.

3.2 Burke's extension to two consecutive gap times

In what follows we propose an estimator for the bivariate distribution function for two successive failure times. This estimator extends Burke's ideas (1988). In this subsection we are in the same conditions and using the same notation as in subsection 2.3.

In particular recall that we are assuming that the follow-up time is subject to independent right censoring by C, which implies that T_1 and $T_1 + T_2$ are independent of C but T_2 is subject to right censoring by $C - T_1$. This assumption can be equivalently expressed as either one of the following three assumptions:
Assumption A1: $\text{Prob}\{T_1 < t_1 | C = c\} = F_1(t_1)$ for all t_1 and c
Assumption A2: $\text{Prob}\{C > u + v | T_1 = u, T_2 = v\} = \text{Prob}\{C > u + v\}$
Assumption A3: $d\text{Prob}\{T_1 \le u, T_2 \le v | C \ge T_1 + T_2\} = dF(u,v)$

Define the subdistribution function

$$\tilde{F}(t_1,t_2) = \text{Prob}\{Y_1 \le t_1, Y_2 \le t_2, \delta = 3\} = \text{Prob}\{Y_1 \le t_1, Y_2 \le t_2, D_1 = D_2 = 1\}$$

which can be naturally estimated from the empirical subdistribution function

$$F_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{\delta_i = 3\}\mathbf{1}\{Y_{1i} \le x, Y_{2i} \le y\}$$

Proposition 1 *If A2 holds, the subdistribution function $\tilde{F}$ can be expressed as the convolution of F and G_1:*

$$\tilde{F}(x,y) = \int_0^x \int_0^y G_1(u+v)dF(u,v)$$

Proof 1

$$\begin{aligned}
\tilde{F}(x,y) &= \text{Prob}\{Y_1 \leq x, Y_2 \leq y, \delta = 3\} = \text{Prob}\{T_1 \leq x, T_2 \leq y, T_1 \leq C_1, T_2 \leq C_2\} \\
&= \text{Prob}\{T_1 \leq x, T_2 \leq y, T_1 + T_2 \leq C_1\} \\
&= \int_0^x \int_0^y \text{Prob}\{C_1 \geq u + v | T_1 = u, T_2 = v\} dF(u,v) \\
&= \int_0^x \int_0^y G_1(u+v) dF(u,v)
\end{aligned}$$

where last equality follows from A2.

Proposition 2 *If A3 holds, the bivariate distribution function F for the random vector (T_1, T_2) can be expressed in terms of the subdistribution function $\tilde{F}$ and the survival for C_1: G_1:*

$$F(x,y) = \int_0^x \int_0^y \{G_1(u+v)\}^{-1} d\tilde{F}(u,v) \tag{20}$$

Proof 2

$$\begin{aligned}
F(x,y) &= \int_0^x \int_0^y dF(u,v) = \int_0^x \int_0^y d\text{Prob}\{T_1 \leq u, T_2 \leq v | C_1 \geq T_1 + T_2\} \\
&= \int_0^x \int_0^y \frac{d\text{Prob}\{T_1 \leq u, T_2 \leq v, C_1 \geq T_1 + T_2\}}{\text{Prob}\{C_1 \geq u + v\}} \qquad (21) \\
&= \int_0^x \int_0^y \frac{d\tilde{F}(u,v)}{G_1(u+v)}
\end{aligned}$$

where last equality follows from A3.

The subdistribution function $\tilde{F}(x,y)$ can be naturally estimated from the empirical subdistribution function

$$F_n(x,y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\delta_i = 3\} \mathbf{1}\{Y_{1i} \leq x, Y_{2i} \leq y\}.$$

Since $C_1 = C$ is the elapsed time from randomization to closing the study, it is observed for all individuals and therefore G_1 can be estimated by the corresponding empirical survival function

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{C_{1i} \leq x\}.$$

The estimation of F is now accomplished plugging $F_n(u,v)$ and $G_n(u+v)$ into the integral (21).

4 Multivariate Survival Data

Multivariate survival data refers to a sequence of survival random variables, $T_1, T_2, T_3, \ldots, T_m$, collected on the same individual. In this section we continue the approach, and extend the notation, taken in Sections 2 and 3 where the random variables are sequentially observed and hence subject to the dependent censoring caused by the date in which the follow–up is closed.

4.1 Notation

The multivariate distribution of $(T_1, T_2, T_3, \ldots, T_m)$ will be characterized by the marginals and the conditional laws. Denote by f_j, S_j and λ_j, $j = 1, \ldots, m$ the marginal density, survival and hazard function of T_j. Denote by $S_{T_j|T_{j-1},\ldots,T_1}$ and $\lambda_{T_j|T_{j-1},\ldots,T_1}$, $j = 2, \ldots, m$ the conditional survival and hazard function of the random variable T_j conditioned to $T_{j-1}, \ldots, T_1$. We use the symbol $[\quad]$ for the joint laws.

As in most clinical trials the survival data is subject to right censoring. We will assume that closing the study is the only cause of censoring, and thus independent of all other survival and covariate information. As in Section 2, denote by C the elapsed time from randomization to closing the study and by G_1 the corresponding survival function. Note that C will be considered itself a random variable due to the large period of recruitment and that C acts on the sum $T_1 + \ldots + T_m$. Censoring due to lost-of-follow-up or withdraw is non considered here. However, if it does occur, it will most probably be informative.

Again, as discussed in Section 2, subsequent times T_j $(j = 2, \ldots, m)$, depend on their previous times in the sense that T_j can be only observed if $T_1 + \ldots + T_{j-1} \leq C$.

We introduce the following censoring variable

$$\delta = \begin{cases} 1 & \text{if} \quad C < T_1 \\ 2 & \text{if} \quad T_1 \leq C < T_1 + T_2 \\ 3 & \text{if} \quad T_1 + T_2 \leq C < T_1 + T_2 + T_3 \\ \ldots & \text{if} \quad \ldots \\ m & \text{if} \quad T_1 + T_2 + \ldots + T_{m-1} \leq C < T_1 + T_2 + \ldots + T_m \\ m+1 & \text{if} \quad T_1 + T_2 + \ldots + T_m \leq C. \end{cases}$$

That is, a given individual can only belong to one of the $m+1$ stages defined by δ. In Stage 1 ($\delta = 1$): all the durations are right–censored. In Stage 2

($\delta = 2$): T_1 is observed while T_2 is right–censored by $C - T_1$. In Stage 3 ($\delta = 3$): T_1 and T_2 are observed while T_3 is right–censored by $C - (T_1 + T_2)$. Finally, in Stage $m + 1$ ($\delta = m + 1$): $T_1, \ldots, T_m$ are observed.

The observed random variables for each individual are $Y_1 = \min(T_1, C)$, $Y_2 = \min(T_2, C - T_1)\mathbf{1}\{T_1 \leq C\}$, $Y_3 = \min(T_3, C - (T_1 + T_2))\mathbf{1}\{T_1 + T_2 \leq C\}, \ldots, Y_m = \min(T_m, C - (T_1 + \ldots + T_{m-1}))\mathbf{1}\{T_1 + \ldots + T_{m-1} \leq C\}$ together with δ. Thus the observed information for a given individual is summarized by $(y_1, y_2, \ldots, y_m, \delta)$.

We will assume that we have a sample and that the observable data are $\{(y_{1i}, y_{2i}, \ldots, y_{mi}, \delta_i), i = 1, \ldots, n\}$.

To have a clear picture of the implications of the censoring mechanism we define auxiliary censoring random variables $C_1 = C$, $C_2 = C - T_1$, $C_j = C - \sum_{l=1}^{j-1} T_l$, $j = 3, \ldots, m$ and describe the joint distribution for $(T_1, T_2, \ldots, T_m, C_1, C_2 \ldots, C_m)$.

Proposition 3

$$[T_1, \ldots, T_m, C_1, \ldots, C_m] \;=\; [T_1, \ldots, T_m, C]\Delta_m \ldots \Delta_3\Delta_2\Delta_1$$

where

$$\Delta_j = \Delta_j(c_j|t_1, \ldots, t_{j-1}) = \begin{cases} 1 & \text{if} \quad c_j = C - (t_1 + \ldots + t_{j-1}) \\ 0 & \text{if} \quad \text{otherwise.} \end{cases}$$

Proof 3 *Straightforward multiplication rules allow to factorize the joint density:*

$$\begin{aligned} [T_1, \ldots, T_m, C_1, \ldots, C_m] &= [T_m|T_1, \ldots T_{m-1}, C_1, \ldots, C_m][T_1, \ldots T_{m-1}, C_1, \ldots, C_m] \\ &= [T_m|T_1, \ldots T_{m-1}, C_1, \ldots, C_m][C_m|T_1, \ldots T_{m-1}, C_1, \ldots, C_{m-1}] \\ &\ldots \; [C_3|T_1, T_2, C_1, C_2][T_2|T_1, C_1, C_2][C_2|T_1, C_1][T_1|C_1][C_1] \end{aligned}$$

Since $C_1 = C$ is known and C_j is function of $T_1, \ldots, T_{j-1}$, last expression becomes proportional to the joint density of $[T_1, \ldots, T_m]$ and $[C]$. Indeed,

$$\begin{aligned} [T_1, \ldots, T_m, C_1, \ldots, C_m] &= [T_m|T_1, \ldots T_{m-1}, C][C_m|T_1, \ldots T_{m-1}] \\ &\quad \ldots [C_3|T_1, T_2][T_2|T_1, C][C_2|T_1][T_1|C][C] \\ &= [T_m|T_1, \ldots T_{m-1}, C][C - \sum_{l=1}^{m-1} T_l|T_1, \ldots T_{m-1}] \\ &\ldots \; [C - (T_1 + T_2)|T_1, T_2][T_2|T_1, C][C - T_1|T_1][T_1][C] \\ &= [T_m|T_1, \ldots T_{m-1}, C] \ldots [T_3|T_1, T_2, C][T_2|T_1, C][T_1|C][C]\Delta_m \ldots \Delta_3\Delta_2\Delta_1 \\ &= [T_1, \ldots, T_m, C]\Delta_m \ldots \Delta_3\Delta_2\Delta_1. \end{aligned}$$

4.2 Likelihood Function

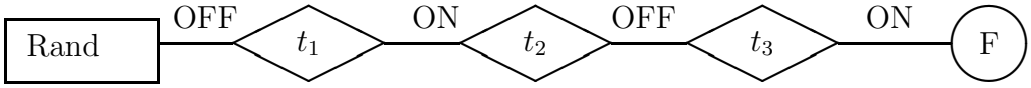
If we are interested in the joint behaviour of $(T_1, \ldots, T_m)$, the contribution to the likelihood of a given subject in terms of the observables $(y_1, y_2, \ldots, y_{m+1}, \delta)$ can be one of the following:

$$\begin{aligned}
P_1 &= \text{Prob}\{T_1 > y_1, C = y_1\} \\
P_2 &= \text{Prob}\{T_1 = y_1, T_2 > y_2, C - y_1 = y_2\} \\
P_3 &= \text{Prob}\{T_1 = y_1, T_2 = y_2, T_3 > y_3, C - (y_1 + y_2) = y_3\} \\
\ldots &\quad \ldots \\
P_m &= \text{Prob}\{T_1 = y_1, \ldots, T_{m-1} = y_{m-1}, T_m > y_m, C - (y_1 + \ldots + y_{m-1}) = y_m\} \\
P_{m+1} &= \text{Prob}\{T_1 = y_1, \ldots, T_m = y_m, C > y_1 + \ldots + y_m\}.
\end{aligned}$$

For instance, an individual in Stage 4 has reinitiated treatment for the second time before the study ends and has $T_1 = t_1 < C$, $T_2 = t_2 < C - t_1$, $T_3 = t_3 < C - (t_1 + t_2)$, $T_4 > C - (t_1 + t_2 + t_3)$ and $\delta = 4$ (See Figure 1). For this individual

$$P_4 = \text{Prob}\{T_1 = y_1, T_2 = y_2, T_3 = y_3, T_4 > C - (y_1 + y_2 + y_3)\}$$

Figure 1: Stage 4: $T_1 = t_1 < C$, $T_2 = t_2 < C - t_1$, $T_3 = t_3 < C - (t_1 + t_2)$, $T_4 > C - (t_1 + t_2 + t_3)$ and $\delta = 4$



The likelihood for the observable data for a given individual is the product of the probability of the $m + 1$ different situations. Note that every subject will only contribute to one of these $m + 1$ factors.

$$Lik = P_1^{\mathbf{1}\{\delta=1\}} P_2^{\mathbf{1}\{\delta=2\}} P_3^{\mathbf{1}\{\delta=3\}} \ldots P_{m+1}^{\mathbf{1}\{\delta=m+1\}} \tag{22}$$

Each term in (22) can be factorized following the usual rules. That is, for any $1 \leq j \leq m$

$$P_j = \text{Prob}\{T_1 = y_1, \ldots, T_{j-1} = y_{j-1}, T_j > y_j, C - (y_1 + \ldots + y_{j-1}) = y_j\}$$

$$
\begin{aligned}
&= \text{Prob}\{T_j > y_j | T_1 = y_1, \ldots, T_{j-1} = y_{j-1}\} \\
&\cdot \text{Prob}\{T_1 = y_1, \ldots, T_{j-1} = y_{j-1}, C = y_1 + \ldots + y_{j-1} + y_j\}
\end{aligned}
$$

where in last equality $C - (y_1 + \ldots + y_{j-1}) = y_j$ can be omitted as it is proved in Proposition 3. Completing the factorization we have

$$
\begin{aligned}
P_j &= \text{Prob}\{T_j > y_j | T_1 = y_1, \ldots, T_{j-1} = y_{j-1}\}\text{Prob}\{T_{j-1} = y_{j-1} | T_1 = y_1, \ldots, T_{j-2} = y_{j-2}\} \\
&\cdot \ldots \text{Prob}\{T_2 = y_2 | T_1 = y_1\}\text{Prob}\{T_1 = y_1\}\text{Prob}\{C = y_1 + \ldots + y_j\} \\
&= S_{T_j|T_1,\ldots,T_{j-1}}(y_j^+ | y_1, \ldots, y_{j-1})\lambda_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{j-1} | y_1, \ldots, y_{j-2})S_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{j-1} | y_1, \ldots, y_{j-2}) \\
&\cdot \ldots \lambda_{T_2|T_1}(y_2|y_1)S_{T_2|T_1}(y_2|y_1)\lambda_{T_1}(y_1)S_{T_1}(y_1)\text{Prob}\{C = y_1 + \ldots + y_j\}
\end{aligned}
$$

where last expression is given in terms of the survivals and the conditional hazards. Note that an individual whose $\delta = j$, will contribute to the estimation of the following functions: $S_{T_j|T_1,\ldots,T_{j-1}}$, $\lambda_{T_{j-1}|T_1,\ldots,T_{j-2}}$, $S_{T_{j-1}|T_1,\ldots,T_{j-2}}$

In order to derive the likelihood function for the observable sample $\{(y_{1i}, y_{2i}, \ldots, y_{m+1i}, \delta_i), i = 1, \ldots, n\}$, we introduce some straightforward notation: for any $1 \leq j \leq m$, the probability P_j for the i^{th} individual is given by:

$$
\begin{aligned}
P_{ji} &= S_{T_j|T_1,\ldots,T_{j-1}}(y_{ji}^+ | y_{1i}, \ldots, y_{(j-1),i})\lambda_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{(j-1),i} | y_{1i}, \ldots, y_{(j-2),i}) \\
&\cdot S_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{(j-1),i} | y_{1i}, \ldots, y_{(j-2),i}) \\
&\cdot \ldots \lambda_{T_2|T_1}(y_{2i}|y_{1i})S_{T_2|T_1}(y_{2i}|y_{1i})\lambda_{T_1}(y_{1i})S_{T_1}(y_{1i})\text{Prob}\{C = y_{1i} + \ldots + y_{ji}\}.
\end{aligned}
$$

The likelihood for the sample can be rewritten as

$$
Lik(sample) = \prod_{i=1}^{n} P_{1i}^{\mathbf{1}\{\delta_i=1\}} P_{2i}^{\mathbf{1}\{\delta_i=2\}} P_{3i}^{\mathbf{1}\{\delta_i=3\}} \ldots P_{m+1,i}^{\mathbf{1}\{\delta_i=m+1\}}, \qquad (23)
$$

and collecting terms

$$
\begin{aligned}
Lik(sample) &= \prod_{i=1}^{n}\prod_{j=1}^{m}\{S_{T_j|T_1,\ldots,T_{j-1}}(y_{ji}^+ | y_{1i}, \ldots, y_{(j-1),i})\lambda_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{(j-1),i} | y_{1i}, \ldots, y_{(j-2),i}) \\
&\cdot S_{T_{j-1}|T_1,\ldots,T_{j-2}}(y_{(j-1),i} | y_{1i}, \ldots, y_{(j-2),i}) \\
&\cdot \ldots \lambda_{T_2|T_1}(y_{2i}|y_{1i})S_{T_2|T_1}(y_{2i}|y_{1i})\lambda_{T_1}(y_{1i})S_{T_1}(y_{1i})\text{Prob}\{C = y_{1i} + \ldots + y_{ji}\}\}^{\mathbf{1}\{\delta_i=j\}}.
\end{aligned}
$$

Last expression can be grouped differently in terms of the different conditional laws.

4.3 Modelling the time dependencies via Partial Likelihood

The expressions corresponding to P_j will depend on the assumptions we are making concerning the renewal times $T_1, T_2, \ldots, T_m$. We assume now that these times hold a second order Markov property and that their hazards are proportional. Under these assumptions a simplified likelihood function is given as well as a proposal for a partial likelihood.

The assumption of a second order Markov implies that the conditional distribution of T_j given $T_1, \ldots, T_{j-1}$ only depends on T_{j-2} and T_{j-1}.That is, the conditional distributions satisfy the following equalities: $[T_j|T_1, \ldots, T_{j-1}] = [T_j|T_{j-2}, T_{j-1}]$, $j = 3, \ldots, m$.

In order to characterize the joint density of $(T_1, T_2, \ldots, T_m, C_1, C_2 \ldots, C_m)$, we only need the conditional density function of T_2 given T_1, named g_2 and the conditional densities of T_j given T_{j-2} and T_{j-1}, denoted by g_j, for $j = 3, \ldots, m$. Denote by G_2 and G_j the corresponding distribution functions.

The contributions given by (22) are therefore proportional to

$$\begin{aligned}
P_1 &= S_1(C) \\
P_2 &= \left(\int_{C-t_1}^{\infty} g_2(t_2|t_1)dt_2\right) f_1(t_1) = G_2(C - t_1|t_1)f_1(t_1) \\
\ldots &= \ldots \\
P_m &= \left((\int_{C-(t_1+\ldots+t_{m-1})}^{\infty} g_m(t_m|t_{m-1}, t_{m-2})dt_m\right) g_{m-1}(t_{m-1}|t_{m-2}, t_{m-3}) \ldots g_2(t_2|t_1)f_1(t_1) \\
P_{m+1} &= g_m(t_m|t_{m-1}, t_{m-2}) \ldots g_2(t_2|t_1)f_1(t_1).
\end{aligned}$$

Then the likelihood for a given individual reads as

$$\begin{aligned}
Lik &= S_1(C)^{1-\delta_1} \left(G_2(C - t_1|t_1)f_1(t_1)\right)^{\delta_1(1-\delta_2)} \times \ldots \times \\
&\times \left(G_m(C - (t_1 + \ldots t_{m-1})|t_{m-1}, t_{m-2})g_{m-1}(t_{m-1}|t_{m-2}, t_{m-3}) \ldots g_2(t_2|t_1)f_1(t_1)\right)^{\delta_1 \ldots \delta_{m-1}(1-\delta_m)} \\
&\times \left(g_m(t_m|t_{m-1}, t_{m-2}) \ldots g_2(t_2|t_1)f_1(t_1)\right)^{\delta_1 \ldots \delta_m},
\end{aligned}$$

and rearranging terms

$$\begin{aligned}
Lik &= S_1(C)^{1-\delta_1} f_1(t_1)^{\delta_1} G_2(C - t_1|t_1)^{\delta_1(1-\delta_2)} g_2(t_2|t_1)^{\delta_1\delta_2} \ldots \\
&\quad G_m(C - (t_1 + \ldots t_{m-1})|t_{m-1}, t_{m-2})^{\delta_1 \ldots \delta_{m-1}(1-\delta_m)} g_m(t_m|t_{m-1}, t_{m-2})^{\delta_1 \ldots \delta_m} \\
&= S_1(C)^{1-\delta_1} f_1(t_1)^{\delta_1} \prod_{j=2}^{m} G_j(C - \sum_{l=1}^{j-1} t_j|t_{j-1}, t_{j-2})^{\delta_1 \ldots \delta_{j-1}(1-\delta_j)} g_j(t_j|t_{j-1}, t_{j-2})^{\delta_1 \ldots \delta_j}
\end{aligned}$$

which is equivalent to

$$Lik = S_1(y_1)^{1-\delta_1} f_1(y_1)^{\delta_1} \prod_{j=2}^{m} G_j(y_j|y_{j-1}, y_{j-2})^{\delta_1 \ldots \delta_{j-1}(1-\delta_j)} g_j(y_j|y_{j-1}, y_{j-2})^{\delta_1 \ldots \delta_j} \tag{24}$$

if we use the compact notation $(y_1, \delta_1, y_2, \delta_2, \ldots, y_m, \delta_m)$.

The likelihood is now rewritten in terms of the corresponding hazard and cumulative hazard functions. Denote by $h_1, h_2, \ldots, h_m$ and by $H_1, H_2, \ldots, H_m$ the hazards and cumulative hazards. Since

$$\begin{aligned} f_1(t_1) &= h_1(t_1)S_1(t_1), \quad g_2(t_2|t_1) = h_2(t_2|t_1)G_2(t_2|t_1) \\ g_j(t_j|t_{j-1}, t_{j-2}) &= h_j(t_j|t_{j-1}, t_{j-2})G_j(t_j|t_{j-1}, t_{j-2}), \quad j = 2, \ldots, m \\ S_1(t_1) &= \exp\{-H_1(t_1)\}, G_2(t_2|t_1) = \exp\{-H_2(t_2|t_1)\} \\ G_j(t_j|t_{j-1}, t_{j-2}) &= \exp\{-H_j(t_j|t_{j-1}, t_{j-2})\}, \quad j = 2, \ldots, m \end{aligned}$$

The likelihood function given by equation (24) becomes after substitution

$$\begin{aligned} Lik &= h_1(y_1)^{\delta_1} \exp\{-H_1(y_1)\} \\ &\times \prod_{j=2}^{m} h_j(y_j|y_{j-1}, y_{j-2})^{\delta_1 \ldots \delta_j} \exp\{-H_j(y_j|y_{j-1}, y_{j-2})\}^{\delta_1 \ldots \delta_{j-1}}. \end{aligned}$$

If instead of assuming a second order Markov property we would assume that renewal times are independent, then the likelihood function is equal to

$$Lik = S_1(y_1)^{1-\delta_1} f_1(y_1)^{\delta_1} \prod_{j=2}^{m} S_j(y_j)^{\delta_1 \ldots \delta_{j-1}(1-\delta_j)} f_j(y_j)^{\delta_1 \ldots \delta_j}$$

Suppose now that we have a proportional hazards model for each time T_j conditioned to T_{j-1} and T_{j-2}:

$$\begin{aligned} h_1(t_1) &= \lambda_{01}(t_1) \\ h_2(t_2|t_1) &= \lambda_{02}(t_2)e^{\beta_{21}t_1} \\ h_j(t_j|t_{j-1}, t_{j-2}) &= \lambda_{0j}(t_j)e^{\beta_{j1}t_{j-1} + \beta_{j2}t_{j-2}} \end{aligned}$$

Assuming that $\lambda_{01}(t) = \ldots = \lambda_{0m}(t) = \lambda_0(t)$ and denoting by $\Lambda_0(t)$ the baseline cumulative hazard function, (25) becomes equal to

$$\begin{aligned}
Lik &= (\lambda_0(y_1))^{\delta_1} \exp\{-\Lambda_0(y_1)\} \\
&\times \left(\lambda_0(y_2)e^{\beta_{21}y_1}\right)^{\delta_1\delta_2} \left(\exp\{-\Lambda_0(y_2)e^{\beta_{21}y_1}\}\right)^{\delta_1} \\
&\times \prod_{j=3}^{m} \left(\lambda_0(y_j)e^{\beta_{j1}y_{j-1}+\beta_{j2}y_{j-2}}\right)^{\delta_1\ldots\delta_j} \left(\exp\{-\Lambda_0(y_j)e^{\beta_{j1}y_{j-1}+\beta_{j2}y_{j-2}}\}\right)^{\delta_1\ldots\delta_{j-1}} \\
&= \prod_{j=1}^{m} (\lambda_0(y_j))^{\delta_1\ldots\delta_j} \exp\{-\Lambda_0(y_1)\} \left(\exp\{-\Lambda_0(y_2)e^{\beta_{21}y_1}\}\right)^{\delta_1} \\
&\times \prod_{j=3}^{m} \left(\exp\{-\Lambda_0(y_j)e^{\beta_{j1}y_{j-1}+\beta_{j2}y_{j-2}}\}\right)^{\delta_1\ldots\delta_{j-1}} \qquad (25) \\
&\times \prod_{j=2}^{m} e^{y_{j-1}\delta_1\ldots\delta_j\beta_{j1}} \prod_{j=3}^{m} e^{y_{j-2}\delta_1\ldots\delta_j\beta_{j2}}
\end{aligned}$$

If we define $\beta_{22} = 0$ (we can also choose $\beta_{11} = \beta_{12} = 0$) then we have a more compact notation since we can write last expression (25) as

$$\begin{aligned}
Lik &= \prod_{j=1}^{m} (\lambda_0(y_j))^{\delta_1\ldots\delta_j} \exp\{-\Lambda_0(y_1)\} \prod_{j=2}^{m} \left(\exp\{-\Lambda_0(y_j) \prod_{l=1}^{2} e^{\beta_{jl}y_{j-l}}\}\right)^{\delta_1\ldots\delta_{j-1}} \\
&\quad \prod_{j=2}^{m} \prod_{l=1}^{2} e^{y_{j-l}\delta_1\ldots\delta_j\beta_{jl}}.
\end{aligned}$$

The likelihood function for the observed sample is then a product over the n individuals of (26).

A partial likelihood is constructed via the product over every time T_j, $j = 2, \ldots, m$ of the following quantities:

$$\prod_{i=1}^{n} \left(\frac{e^{\beta_{j1}t^i_{j-1}+\beta_{j2}t^i_{j-2}}}{\sum_{i*=1}^{n} J^{i*}_j(t^i_j) e^{\beta_{j1}t^{i*}_{j-1}+\beta_{j2}t^{i*}_{j-2}}} \right)^{\delta^i_1\ldots\delta^i_j} \qquad (26)$$

where $J^{i*}_j(t) = \mathbf{1}\{T^{i*}_j \geq t\}$. Each term in equation (26) is the relative *contribution-hazard* of a subject i that has failed at time t^i_j with respect to all the other subjects at risk at time t^i_j.

The partial likelihood would then be defined as

$$\mathcal{L} \;=\; \prod_{j=2}^{m}\prod_{i=1}^{n}\left(\frac{e^{\beta_{j1}t^i_{j-1}+\beta_{j2}t^i_{j-2}}}{\sum_{i*=1}^{n} J^{i*}_j(t^i_j)e^{\beta_{j1}t^{i*}_{j-1}+\beta_{j2}t^{i*}_{j-2}}}\right)^{\delta^i_1\ldots\delta^i_j}. \tag{27}$$

5 Bayesian approach for modelling trends and time dependencies

There are several ways of jointly modelling the longitudinal data and the survival time. The most common methods consist on specifying a linear mixed-effects model for the covariate process and a Cox proportional hazards model for the survival. Generalization to multiple longitudinal data is however complicated by the need to model the joint relationship among all the covariates and the difficulty derived of high-dimensional integration in the likelihood (Song, Davidian and Tsiatis, 2002).

An interesting issue to investigate when analyzing recurrent events is whether or not exists a trend between the consecutive observed times. In the Tibet clinical trial this is an important question which needs an answer. Observing a trend of longer and longer times without treatment, would represent a very good behavior of the interruption treatment associated with a strengthening of the immunologic system. On the contrary, a trend of smaller and smaller times without treatment, or larger and larger times with treatment, would be possibly associated with a deterioration of the immunologic system.

To investigate the trend between consecutive times we propose two alternatives. First, we model the time dependencies via indicators within a proportional hazards model. Then, the observed times are used to model the dependency between consecutive times. In both cases, after assuming a specific model for T_1, inferences on the parameters of interest could be performed either by a maximum likelihood method or through the posterior distribution following a Bayesian approach.

We have chosen a Bayesian approach as an interesting alternative to maximum likelihood methods since it provides a direct probabilistic interpretation of the posterior distribution and allows the incorporation of prior beliefs about the distribution function. Furthermore, nowadays, the implementation of the Bayesian approach is much more easily-approached because of the Markov chain Monte Carlo algorithms and their implementation with programs such as Winbugs.

We describe both approaches assuming a Weibull model for T_1.

5.1 Modelling time dependencies via indicators

Following the notation from Section 4, let $\lambda_1(t)$ the hazard function of T_1 and, for $j = 2, \ldots, m$, $\lambda_{j|j-1,\ldots,1}$ the conditional hazard of T_j given all previous times $T_{j-1}, \ldots, T_1$. We define recurrently these conditional hazards as

$$\begin{aligned}
&\lambda_{2|1}(t|t_1) = \lambda_1(t) \cdot e^{\theta_1} \\
&\lambda_{3|2,1}(t|t_1, t_2) = \lambda_{2|1}(t|t_1) \cdot e^{\theta_2} \\
&\ldots \\
&\lambda_{j|j-1,\ldots,1}(t|t_1, \ldots, t_{j-1}) = \lambda_{j-1|j-2,\ldots,1}(t|t_1, \ldots, t_{j-2}) \cdot e^{\theta_{j-1}},
\end{aligned}$$

which is equivalent to

$$\lambda_{j|j-1,\ldots,1}(t|t_1, \ldots, t_{j-1}) = \lambda_1(t) \cdot e^{\theta_1 + \ldots + \theta_{j-1}}; \quad j = 2, \ldots, m. \tag{28}$$

Note that $e^{\theta_1} = \lambda_2(t|t_1)/\lambda_1(t)$ is the risk factor between T_2 and T_1 and $e^{\theta_2} = \lambda_3(t|t_1, t_2)/\lambda_2(t|t_1)$ is the risk factor between T_3 and T_2. In general, e^{θ_k} is the risk factor between consecutive times T_k and T_{k+1}. A positive value $\theta_k > 0$ means that times T_{k+1} tend to be smaller than times T_k.

The likelihood function for this model is given in equation (22). The different contributions into the likelihood for a given individual are:

$$\begin{aligned}
P_1 &= \text{Prob}\{T_1 > F\} = S_1(y_1) \\
P_2 &= \text{Prob}\{T_1 = t_1, T_2 > F - t_1\} = f_1(y_1) S_{2|1}(y_2) = \lambda_1(y_1) S_1(y_1) (S_1(y_2))^{\exp\{\theta_1\}} \\
P_3 &= \text{Prob}\{T_1 = t_1, T_2 = t_2, T_3 > F - (t_1 + t_2)\} = f_1(y_1) f_{2|1}(y_2) S_{3|2,1}(y_3) = \\
&= \lambda_1(y_1) S_1(y_1) \lambda_{2|1}(y_2) S_{2|1}(y_2) S_{3|2,1}(y_3) = \\
&= \lambda_1(y_1) S_1(y_1) \lambda_1(y_2) e^{\theta_1} (S_1(y_2))^{\exp\{\theta_1\}} (S_1(y_3))^{\exp\{\theta_1 + \theta_2\}} \\
\ldots &= \ldots \\
P_m &= \text{Prob}\{T_1 = t_1, \ldots, T_{m-1} = t_{m-1}, T_m > F - (t_1 + \ldots + t_{m-1})\} \\
&= e^{(m-2)\theta_1 + (m-3)\theta_2 + \ldots + \theta_{m-2}} \lambda_1(y_1) \lambda_1(y_2) \ldots \lambda_1(y_{m-1}) \cdot \\
&\cdot S_1(y_1) (S_1(y_2))^{\exp\{\theta_1\}} (S_1(y_3))^{\exp\{\theta_1 + \theta_2\}} \ldots (S_1(y_m))^{\exp\{\theta_1 + \ldots + \theta_{m-1}\}} \\
P_{m+1} &= \text{Prob}\{T_1 = t_1, \ldots, T_m = t_m\} =
\end{aligned}$$

$$= e^{(m-1)\theta_1+(m-2)\theta_2+\ldots+\theta_{m-1}}\lambda_1(y_1)\lambda_1(y_2)\ldots\lambda_1(y_m)\cdot$$

$$\cdot S_1(y_1)(S_1(y_2))^{\exp\{\theta_1\}}(S_1(y_3))^{\exp\{\theta_1+\theta_2\}}\ldots(S_1(y_m))^{\exp\{\theta_1+\ldots+\theta_{m-1}\}},$$

where $f_{j|j-1,\ldots,1}$ and $S_{j|j-1,\ldots,1}$ are the conditional density and survival functions, respectively, of T_j given all previous times $T_{j-1},\ldots,T_1$, for $j=2,\ldots,m$.

We assume a Weibull distribution for T_1 with the following parameterization (the one used by the Winbugs):

$$\lambda_1(t)=\gamma\lambda t^{\gamma-1}$$

We denote this by $T_1\sim\text{Weib}(\gamma,\lambda)$ where γ is the shape parameter and λ is the scale parameter.

Under modelization (28) the conditional hazards have the form

$$\lambda_{j|j-1,\ldots,1}(t|t_1,\ldots,t_{j-1})=\lambda_1(t)\cdot e^{\theta_1+\ldots+\theta_{j-1}}=\gamma\lambda t^{\gamma-1}\cdot e^{\theta_1+\ldots+\theta_{j-1}}$$

which correspond to a Weibull distribution with shape parameter γ and scale parameter $\lambda e^{\theta_1+\ldots+\theta_{j-1}}$.

Therefore, to obtain the posterior distributions of γ, λ and θ_j, $j\in\{1,\ldots,m-1\}$ using the Winbugs program we only need to specify the following conditional distributions, indicating that the last time T_m is always right censored:

$T_1\sim\text{Weib}(\gamma,\lambda)$

$T_2|T_1\sim\text{Weib}(\gamma,\lambda e^{\theta_1})$

$T_3|T_2,T_1\sim\text{Weib}(\gamma,\lambda e^{\theta_1+\theta_2})$

...

$T_m|T_{m-1},\ldots,T_2,T_1\sim\text{Weib}(\gamma,\lambda e^{\theta_1+\theta_2+\ldots+\theta_{m-1}})$ right censored at $F-(t_1+\ldots+t_{m-1})$

and prior distributions for γ, λ and θ_j, $j\in\{1,\ldots,m-1\}$.

The following code corresponds to a Winbugs program to analyze the first consecutive 4 times assuming diffuse Gamma prior distributions for both γ and λ and diffuse normal distributions for $\theta_j,\ j \in \{1, \ldots, m-1\}$.

```
model weibull1_4;
    {
        for(i in 1:n4){ # individuals with observed T_1, T_2, T_3 and T_4
            t1[i] ~ dweib(gam,lambda);
            t2[i] ~ dweib(gam,lambda2);
            t3[i] ~ dweib(gam,lambda3);
            t4[i] ~ dweib(gam,lambda4);
        }
        for(i in n4+1:n3){ # individuals with right censored T_4
            t1[i] ~ dweib(gam,lambda);
            t2[i] ~ dweib(gam,lambda2);
            t3[i] ~ dweib(gam,lambda3);
            t4[i] ~ dweib(gam,lambda4) I(c4[i],);
        }
        for(i in n3+1:n2){ # individuals with right censored T_3
            t1[i] ~ dweib(gam,lambda);
            t2[i] ~ dweib(gam,lambda2);
            t3[i] ~ dweib(gam,lambda3) I(c3[i],);
        }
        for(i in n2+1:n1){  # individuals with right censored T_2
            t1[i] ~ dweib(gam,lambda);
            t2[i] ~ dweib(gam,lambda2) I(c2[i],);
        }
        for(i in n1+1:N){   # individuals with right censored T_1
            t1[i] ~ dweib(gam,lambda) I(c1[i],);
        }

        lambda2<-lambda*exp(theta1);
        lambda3<-lambda*exp(theta1+theta2);
        lambda4<-lambda*exp(theta1+theta2+theta3);

        theta1~ dnorm(alpha1,tau1);
        theta2 ~ dnorm(alpha2,tau2);
        theta3 ~ dnorm(alpha3,tau3);
```

```
        sumtheta12<-theta1+theta2;
        sumtheta123<-theta1+theta2+theta3;


        lambda ~ dgamma(1.0E-3, 1.0E-3);
        gam ~ dgamma(1.0E-3, 1.0E-3);

        alpha1 ~ dnorm(0, 1.0E-6);
        tau1~ dgamma(1.0E-3, 1.0E-3);
        alpha2 ~ dnorm(0, 1.0E-6);
        tau2~ dgamma(1.0E-3, 1.0E-3);
        alpha3 ~ dnorm(0, 1.0E-6);
        tau3~ dgamma(1.0E-3, 1.0E-3);
    }
#  Constants:
    list(N=96,n1=55,n2=28,n3=13,n4=7);
#  Initial values:
    list(alpha1=0, tau1=1, alpha2=0, tau2=1,alpha3=0, tau3=1,lambda=1, gam=1);
```

5.2 Modelling the time dependencies using durations

Consider a modelization similar to (28) but using the observed times to model the dependency between consecutive times.

As in previous sections, let $\lambda_1(t)$ be the hazard function of T_1 and, for $j = 2, \ldots, m$, $\lambda_{j|j-1,\ldots,1}$ the conditional hazard of T_j given all previous times $T_{j-1}, \ldots, T_1$. We define recurrently these conditional hazards as

$$\lambda_{2|1}(t|t_1) = \lambda_1(t) \cdot e^{\theta_1 t_1}$$

$$\lambda_{3|2,1}(t|t_1, t_2) = \lambda_{2|1}(t|t_1) \cdot e^{\theta_2 t_2}$$

$$\ldots$$

$$\lambda_{j|j-1,\ldots,1}(t|t_1, \ldots, t_{j-1}) = \lambda_{j-1|j-2,\ldots,1}(t|t_1, \ldots, t_{j-2}) \cdot e^{\theta_{j-1} t_{j-1}}$$

Which is equivalent to

$$\lambda_{j|j-1,\ldots,1}(t|t_1, \ldots, t_{j-1}) = \lambda_1(t) \cdot e^{\theta_1 t_1 + \ldots + \theta_{j-1} t_{j-1}}; \quad j = 2, \ldots, m. \tag{29}$$

As before, the likelihood function for this model is given in equation (22). The different contributions into the likelihood for a given individual are:

$$
\begin{aligned}
P_1 &= S_1(y_1) \\
P_2 &= \lambda_1(y_1)S_1(y_1)(S_1(y_2))^{\exp\{\theta_1 t_1\}} \\
P_3 &= \lambda_1(y_1)S_1(y_1)\lambda_1(y_2)e^{\theta_1 t_1}(S_1(y_2))^{\exp\{\theta_1 t_1\}}(S_1(y_3))^{\exp\{\theta_1 t_1+\theta_2 t_2\}} \\
\ldots &= \ldots \\
P_m &= e^{(m-2)\theta_1 t_1+(m-3)\theta_2 t_2+\ldots+\theta_{m-2}t_{m-2}}\lambda_1(y_1)\lambda_1(y_2)\ldots\lambda_1(y_{m-1})\cdot \\
&\quad \cdot S_1(y_1)(S_1(y_2))^{\exp\{\theta_1 t_1\}}(S_1(y_3))^{\exp\{\theta_1 t_1+\theta_2 t_2\}}\ldots(S_1(y_m))^{\exp\{\theta_1 t_1+\ldots+\theta_{m-1}t_{m-1}\}} \\
P_{m+1} &= e^{(m-1)\theta_1 t_1+(m-2)\theta_2 t_2+\ldots+\theta_{m-1}t_{m-1}}\lambda_1(y_1)\lambda_1(y_2)\ldots\lambda_1(y_m)\cdot \\
&\quad \cdot S_1(y_1)(S_1(y_2))^{\exp\{\theta_1 t_1\}}(S_1(y_3))^{\exp\{\theta_1 t_1+\theta_2 t_2\}}\ldots(S_1(y_m))^{\exp\{\theta_1 t_1+\ldots+\theta_{m-1}t_{m-1}\}},
\end{aligned}
$$

To analyze model (29) from a parametric Bayesian approach we need to specify each conditional distribution. As before, we assume a Weibull distribution for T_1:

$$\lambda_1(t) = \gamma\lambda t^{\gamma-1}$$

and

$$\lambda_{j|j-1,\ldots,1}(t|t_1,\ldots,t_{j-1}) = \lambda_1(t)\cdot e^{\theta_1+\ldots+\theta_{j-1}} = \gamma\lambda t^{\gamma-1}\cdot e^{\theta_1 t_1+\ldots+\theta_{j-1}t_{j-1}}$$

which corresponds to a Weibull distribution with shape parameter γ and scale parameter $\lambda e^{\theta_1 t_1+\ldots+\theta_{j-1}t_{j-1}}$.

To obtain the posterior distributions of γ, λ and $\theta_j,\ j\in\{1,\ldots,m-1\}$ using the Winbugs program we only need to specify the following conditional distributions, indicating that the last time T_m is right censored:

$T_1 \sim \text{Weib}(\gamma,\lambda)$

$T_2|T_1 \sim \text{Weib}(\gamma,\lambda e^{\theta_1 t_1})$

$T_3|T_2,T_1 \sim \text{Weib}(\gamma,\lambda e^{\theta_1 t_1+\theta_2 t_2})$

$\ldots$

$T_m|T_{m-1},\ldots,T_2,T_1 \sim \text{Weib}(\gamma,\lambda e^{\theta_1 t_1+\theta_2 t_2+\ldots+\theta_{m-1}t_{m-1}})$ right censored at $F-(t_1+\ldots+t_{m-1})$

and prior distributions for γ, λ and $\theta_j,\ j\in\{1,\ldots,m-1\}$.

6 Review of Multivariate Regression Survival Models

6.1 Introduction

Multivariate survival data arise when each subject in a given study may experience multiple events. Multiple events may be ordered or unordered. We restrict this paper to the analysis of ordered events which includes recurrent events of the same type, such as, repeated myocardial infarction attacks, and ordered events of different types, like infection with HIV and subsequent AIDS diagnosis. We review some of the contributions of the analysis of multivariate survival data when the scientific interest center on the effects of covariates on the risk of failure.

Models for multivariate survival time data can be classified as

1. parametric models which specify the structure of the dependence through the joint distribution of the multivariate data,

2. regression models based on models for the marginal hazard functions and

3. frailty models where the dependence is specified via random effects in the model.

Both the first and last approaches rely on parametric assumptions which are difficult to validate. For this reason in this paper we restrict our revision to the second category which includes the marginal approach due to Wei, Lin & Weissfeld (1989), the multiplicative intensity model by Andersen and Gill (1990) and the conditional approach by Prentice, Williams & Peterson (1981), referred hereafter as WLW, AG and PWP, respectively. All these methods are sometimes called marginal approaches. That is, on one hand, they model the marginal distribution of each failure time variable and, on the other, the regression parameters are estimated ignoring the dependence between related failure times with corrected variance. In this review we will use the term "marginal approach" only when referring to WLW.

The main differences between these three methods are based on the following three points: the assumption on the dependence structure among failures, the definition of the risk sets and the computation of the standard error of the proposed estimates.

6.2 Regression models for a sequence of survival times based on the marginal distributions

Let $U_1, \ldots, U_m$, such that $U_1 < \ldots < U_m$, be the times of m consecutive events. The times are measured from the same origin, for instance, from the beginning of the study. Let $Z_{ji}(t)$ be a vector of covariates for the i^{th} subject at time t. The multivariate distribution of $(U_1, U_2, \cdots, U_m)$ will be characterized by the marginals and the conditional laws. Denote by F_j, f_j, S_j and λ_j, $j = 1, \cdots, m$ the marginal distribution, density, survival and hazard function of U_j.

As in most clinical trials the survival data is subject to right censoring. We will assume that closing the study is the only cause of censoring, and thus independent of all other survival and covariate information. Denote by C the elapsed time from randomization to closing the study. Censoring indicators are defined as $\mu_j = \mathbf{1}\{U_j \leq C\}$, for $j = 1, \ldots, m$. Note that $\mu_j = 1$ implies $\mu_1 = \cdots \mu_{j-1} = 1$. Denote by $X_j = \min\{U_j, C\}$, $j = 1, \cdots, m$. In addition to censoring we also have to deal with missing data. The observation of subsequent times U_j, $(j = 2, \cdots, m)$ depends on their previous times in the sense that U_j is only observed if $U_{j-1} \leq C$. We define a missing data indicator, R_j , as $R_j = 1$ if U_j is observed or censored and $R_j = 0$ when U_j is missing. In this last case, we set, for notational convenience, $X_j = X_{j-1}$ and $\mu_j = 0$, though these cases will make no contribution to the estimations.

Suppose that there are n subjects. The observable data for subject i is given by $(X_{1i}, \mu_{1i}, Z_{1i}, R_{1i}, \cdots, X_{mi}, \mu_{mi}, Z_{mi}, R_{mi})$, for $i \in \{1, \ldots, n\}$.

The methods considered in this section are based on a likelihood function (see (31)) which is built taking into account: 1) the specification of the marginal distributions for U_j (see (30)) and 2) a working independence assumption concerning the failure times. Furthermore, inferences are performed using standard univariate methods to estimate the corresponding regression parameters assuming the U_j are independent. When the consecutive times are not independent it is necessary to develop specific variance estimation procedures.

Marginal distributions for each specific failure time U_j are assumed to follow a proportional hazards model,

$$\lambda_{ji}(t; Z_{ji}(t)) = \lambda_{0j}(t) \exp\{\beta_j Z_{ji}(t)\}, \ j = 1, \ldots, m, \tag{30}$$

with corresponding partial likelihood function given by

$$L_j(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta_j Z_{ji}(y_{ji})\}}{\sum_{k=1}^{n} H_{jk}(y_{ji}) \exp\{\beta_j Z_{jk}(y_{ji})\}} \right\}^{\mu_{ji} R_{ji}}$$

where $H_{ki}(t)$ is the risk-set indicator of individual i for the k^{th} failure.

Under the working independence assumption, that is, assuming that the consecutive times $U_1, \ldots, U_m$ are independent and independent censoring, the overall partial likelihood function can be factorized as the product of the marginal partial likelihoods, that is:

$$L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta_j Z_{ji}(y_{ji})\}}{\sum_{k=1}^{n} H_{jk}(y_{ji}) \exp\{\beta_j Z_{jk}(y_{ji})\}} \right\}^{\mu_{ji} R_{ji}}. \tag{31}$$

When $U_1, \ldots, U_m$ are not independent given covariates, expression (31) does not correspond to the correct partial likelihood function and is referred to as a pseudolikelihood function. However, if the marginal models are correctly specified, the estimator $\hat{\beta}$ which satisfies the pseudolikelihood score equation, $U(\beta) = 0$, is a consistent estimator of the parameter β (Lin, 1994).

As mentioned in the introduction the most common approaches following this working independence assumption are the marginal approach due to Wei, Lin & Weissfeld (WLW), the independent increment method by Andersen and Gill (AG) and the conditional approach by Prentice, Williams & Peterson (PWP). The three methods base the inferences on the pseudolikelihood function (31) and they differ in the way the risk-set indicators are defined, in the use of the individual's history data and in the use of different time scales.

6.2.1 Wei, Lin and Weissfield approach

The method proposed by Wei, Lin and Weissfield models the marginal distribution of each failure time with a Cox proportional hazards model. The regression parameters are estimated by maximizing the pseudolikelihood function (31) which is equivalent to the maximization of each failure-specific partial likelihood. Moreover an estimator of the standard errors of the regression parameter estimates is based on a sandwich estimator given in Wei, Lin and Weissfield paper.

This approach is similar to the competing risks approach because treats the ordered event as if they were unordered. That is, it is assumed that an

individual is at risk of event j during its follow-up irrespectively that the event $j-1$ has occurred or not. However WLW follows a competing risk approach. The risk-set indicator of individual k for the j^{th} failure time is defined as

$$H_{jk}(t) = I(X_{jk} \geq t).$$

WLW method does not use any information on the individual's history data and is only appropriate for analyzing total times, that is, times measured from the same origin.

6.2.2 Prentice, Williams and Peterson approach

The method proposed by Prentice, Williams and Peterson is called the conditional approach because it uses information on the individual's history. It is based on a stratified Proportional Hazards model (Cox, 1972) where the hazard at time t is assumed to depend on the covariate process up to time t and, the counting process defined by the number of failures prior to time t. The regression parameters are estimated by maximizing the pseudolikelihood function (31), as in WLW method, but now, the covariate process, $Z_{ki}(t)$, may contain information on the past individual's history. In particular, appropriate time dependent covariates may be used to capture the dependence between the sequence of survival times. Also the risk-set indicator contains information on the individual's past history since it is assumed that an individual cannot be at risk of event j until event $j-1$ has occurred. In this case, the risk-set indicator of individual k for the j^{th} failure time is defined as

$$H_{jk}(t) = I(X_{j-1,k} < t \leq X_{jk}).$$

This approach also allows to analyze the effect of covariates on the gap times between consecutive events. For this time scale, the hazard function (30) takes the form

$$\lambda_{ji}(t; Z_{ji}(t)) = \lambda_{0j}(t - t_{k-1,i}) \exp\{\beta_j Z_{ji}(t)\},\ j = 1, \ldots, m. \tag{32}$$

We define by $T_{ji} = X_{ji} - X_{j-1,i}$ the elapsed or gap time between the occurrence of event $j-1$ and event j. Gap time T_{ji} is right-censored when U_{ji} is. When U_{ji} is missing, $R_{ji} = 0$ and the value of the corresponding gap time T_{ji} is also equal to zero by definition because, as we mention before, in this case we set $X_{ji} = X_{j-1,i}$. However, these cases do not contribute to the

inferences. The corresponding pseudolikelihood function for analyzing gap times can be written as

$$L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta_j Z_{ji}(y_{ji})\}}{\sum_{k=1}^{n} H^*_{jk}(x_{ji}) \exp\{\beta_j Z_{jk}(y_{j-1,k} + x_{ji})\}} \right\}^{\mu_{ji} R_{ji}} \tag{33}$$

where the risk-set indicator of individual k for the j^{th} failure time is defined as

$$H^*_{jk}(t) = I(T_{jk} \geq t).$$

6.2.3 Andersen and Gill approach

The approach by Andersen and Gill is appropriate when one is interested in the overall effect of covariates for recurrences of the same nature. AG can be seen as a particular case of PWP where the marginal distributions are assumed to have identical baseline hazard functions:

$$\lambda_{ji}(t; Z_{ji}(t)) = \lambda_0(t) \exp\{\beta Z_{ji}(t)\}, \ j = 1, \ldots, m. \tag{34}$$

In this case the overall pseudolikelihood function is given by:

$$L(\beta) = \prod_{j=1}^{m} \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta Z_{ji}(y_{ji})\}}{\sum_{k=1}^{n} \sum_{l=1}^{m} H_{lk}(y_{ji}) \exp\{\beta Z_{lk}(y_{ji})\}} \right\}^{\mu_{ji} R_{ji}} \tag{35}$$

where, as in PWP, the risk-set indicator of individual k for the j^{th} failure time is defined as

$$H_{jk}(t) = I(X_{j-1,k} < t \leq X_{jk}).$$

A gap time version of AG model is also possible following similar steps as in PWP model.

6.2.4 Possible biased inferences

Each of the three methods described above may have potential biases if some hypotheses are violated.

If the different consecutive survival times are conditionally independent given covariates, then data is missing at random:

$$R_{ji} = P(T_j \text{ missing} \mid U_1, \ldots, U_{j-1}, Z_{1i}, \ldots, Z_{ji}) = P(T_j \text{ missing} \mid Z_{1i}, \ldots, Z_{ji})$$

However, when an important covariate is not included in the model, then the consecutive survival times are not conditionally independent and therefore data is not missing at random. In this case the estimations performed with any of the above methods might be biased.

Thernau & Grambsch (2000) pointed out, using simulated data, that the omission of a covariate can cause bias on the estimated parameters of the other covariates included in the model. They establish a model with two covariates (treatment and a per-patient covariate) and only treatment is included in the model. Then they obtain a slight biased estimated parameter when the method used is the AG while for the marginal and conditional methods the estimation is seriously biased. They argue that in the conditional model bias is due to loss of balance in the later strata and for the marginal model, ignoring a covariate may violate the proportional hazards assumption. Finally their work includes a brief comparison of the quality of estimators when it is considered only the time until the first event. In this case when a covariate is omitted the estimator is biased and the magnitude of this bias is something between the AG and the conditional models.

Another source of biased inferences may be the violation of the proportional hazards assumption. It is necessary to check if this assumption is reliable. As we pointed out all of the three methods establishes a proportional hazards model for each specific failure time.

6.3 Computational aspects

The models described in section 6.2 can be fitted using both SAS and S-Plus system softwares. Each model requires an specific data set structure which reflects the different definitions of the risk-sets.

To describe the different data sets we consider the following example with three individuals: First subject experiences one event at time 12 and was followed until time 45. Second subject has two events at times, 8, 32 and a follow–up to time 58. Third subject was followed during 50 days and did not experienced any event.

The data set corresponding to this example is given in Tables 1 and 2. Table 1 structures the data as it is required for PWP and AG approaches, while Table 2 is appropriate for WLW approach.

The main differences in these tables is on the number of rows corresponding to a given individual. That is, if an individual experiences two events (see the individual with $id = 2$) then for the PWP model it defines three rows,

Table 1: PWP and AG data set structure for total time scale

id	start	stop	status	enum
1	0	12	1	1
1	12	45	0	2
2	0	8	1	1
2	8	32	0	2
2	32	58	0	3
3	0	50	0	1

Table 2: WLW data set structure for total time scale. Since individual 2 experiences two events, every subject is described with three lines

id	start	stop	status	enum
1	0	12	1	1
1	0	45	0	2
1	0	45	0	3
2	0	8	1	1
2	0	32	1	2
2	0	58	0	3
3	0	50	0	1
3	0	50	0	2
3	0	50	0	3

one for each of the three time intervals, but if an individual experiences only one event then it has one or two rows depending on the total follow-up time. However if the method applied is WLW, then all the individuals has the same number of rows defined by the maximum number of events that have the individual with the maximum number of events. In our example all individuals have three rows defined by two events and the follow–up time.

Table 1 will be substituted by Table 3 if we are interested in the gap time scale and the methods of PWP and AG are used.

Consider, in addition, a time dependent covariate $x(t)$ which represents the time of a specific intervention. $x(t)$ is equal to zero until the time of the intervention; at this time, and thereafter, $x(t)$ is set to one. The intervention

Table 3: PWP and AG data set structure for gap time scale

id	start	stop	status	enum
1	0	12	1	1
1	0	33	0	2
2	0	8	1	1
2	0	24	0	2
2	0	26	0	3
3	0	50	0	1

times for each individual are 25, 20 and 41, respectively. The data set corresponding to this example is in Table 4 for the methods PWP and AG and Table 5 for the WLW approach. These data sets are builded in a similar way than in the case of using time-dependent covariates in the standard survival analysis of a univariate time. That is, it is needed a record for each value of the time-dependent covariate.

Table 4: PWP and AG data set structure for total time scale

id	start	stop	status	enum	x
1	0	12	1	1	0
1	12	25	0	2	0
1	25	45	0	2	1
2	0	8	1	1	0
2	8	20	0	2	0
2	20	32	1	2	1
2	32	58	0	3	1
3	0	41	0	1	0
3	41	50	0	1	1

Table 5: WLW data set structure for total time scale

id	start	stop	status	enum	x
1	0	12	1	1	0
1	0	25	0	2	0
1	25	45	0	2	1
1	0	25	0	3	0
1	25	45	0	3	1
2	0	8	1	1	0
2	0	20	0	2	0
2	20	32	1	2	1
2	0	20	0	3	0
2	20	58	0	3	1
3	0	41	0	1	0
3	41	50	0	1	1
3	0	41	0	2	0
3	41	50	0	2	1
3	0	41	0	3	0
3	41	50	0	3	1

6.3.1 S-Plus and SAS codes

In this section we present the SAS and S-Plus codes to obtain the estimates coming from the methods described above. We point out that in both softwares these methods may be implemented using the standard procedures for the analysis of univariate survival data.

In S-Plus the statements are based on the `coxph` function and the differences among the AG, conditional and marginal models are as follows:

- `afit <- coxph(Surv(start,stop,status) ~ cov1 + cov2 + cluster(id), data=data1)`
- `cfit <- coxph(Surv(start,stop, status) ~ cov1 + cov2 + cluster(id) + strata(enum), data=data1)`
- `wfit <- coxph(Surv(time,status) ~ cov1 + cov2 + cluster(id) + strata(enum), data=data2)`

Using the SAS system software the code is based on the `phreg` procedure. In what follows we present the code to implement the AG, the conditional and the marginal methods:

- `proc phreg data=data1; model (start, stop) * status(0) = cov1 cov2; run;`
- `proc phreg data=data1; model stop*status(0)=cov1 cov2; strata enum; run;`
- `proc phreg data=data2; model stop*status(0)=cov1 cov2; strata enum; id id; run;`

  Standard errors have to be corrected using the sandwich formulae.

For the marginal model due to Wei, Lin & Weissfeld (1989) there is also a macro (WLW macro given, for instance, in Allison (1995)) for the SAS software, producing tests and partial likelihood estimates. Standard errors are already corrected. The macro incorporates a SAS/IML program (see *SAS/STAT Software: Changes and Enhancements, Release 6.10*)

References

[1] Andersen, P.K and Gill, R.D. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* 1982; **10**, 1100–1120.

[2] Burke, M.D. Estimation of a bivariate distribution function under random censorship. *Biometrika* 1988; **75**, 379–382.

[3] Campbell, G. Nonparametric bivariate estimation with randomly censored data. *Biometrika* 1981; **68**, 417–422.

[4] Campbell, G. and Foldes, A. Large–sample properties of nonparametric bivariate estimators with censored data. *Colloquia Mathematica Societatis Janos Bolyai* 1982; **32**, 103–121.

[5] Clayton, D.G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978; **65**, 141–151.

[6] Cox,D.R. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**, 187–220.

[7] Dabrowska, D.M. Kaplan-Meier estimate on the plane. *The Annals of Statistics* 1988; **16**, 1475–1489.

[8] Fan, J. Hsu, L. and Prentice, R. Dependence estimation over a finite bivariate failure time region. *Lifetime Data Analysis* 2000; **6**, 343–355.

[9] Gill, R.D. Multivariate survival analysis. *Theory of Probability and its Applications* 1992; **37**, 18–31.

[10] Hanley, J.A. and Parnes, M.N. Nonparametric estimation of a multivariate distribution in the presence of censoring. *Biometrics* 1983; **39**, 129–139.

[11] Lin, D.Y. and Ying, Z. A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* 1993; **80**, No.3, 573–581.

[12] Lin, D.Y., Sun, W and Ying, Z. Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* 1999; **86**, No.1, 59–70.

[13] Oakes, D. Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 1989; **84**, 487–493.

[14] Prentice, R.L. and Cai, J. Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* 1992; **79**, 495–512.

[15] Prentice, R.L., Williams, B.J. and Peterson, A.V. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**, 373–379.

[16] Satten, G.A. and Datta, S. The Kaplan-Meier Estimator as an Inverse–Probability–of–Censoring Weighted Average. *The American Statistician* 2001; **55**, No.3, 207–210.

[17] Song, X., Davidian, M. and Tsiatis, A.A. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* 2002; **3**, 511–528.

[18] Tsai,W.Y., Leurgans, S. and Crowley, J. Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics* 1986; **14**, 1351–1365.

[19] Visser, M. Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS. *Biometrika* 1996; **83**, 507–518.

[20] Wang, M.C. and Wells, M.T. Nonparametric Estimation of successive duration times under dependent censoring. *Biometrika* 1998 **85**, 561–572.

[21] Wei, L.J., Lin, D.Y. and Weissfeld, L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association* 1989; **84**, 1065–1073.