Testing Markov property for an interval censored three-state progressive model

1. Introduction

We consider a progressive process on states $\{1, 2, 3\}$ with two possible transitions: $1 \to 2$ and $2 \to 3$. Let $S =$ the time of $1 \to 2$ transition, $V =$ the time in state 2,and $T = S + V$ the time of $2 \to 3$ transition. We assume that time of $1 \to 2$ transition is interval censored and time of $2 \to 3$ is either interval censored or right censored. The process may also be right censored in state 1. This model was first considered by De Gruttola and Lagakos (GL) (1989) in the context of studying the cohort of persons with hemophilia for whom both time of infection with the human immunodeficiency virus (HIV) (time of $0 \to 1$ transition) and the onset of the acquired immunodeficiency syndrome (Aids) (time of $1 \to 2$ transition) were interval censored. Following GL we refer to this process as being doubly interval censored (DIC) and to the data that it generates as DIC data.

Assuming independence of S and V, and discrete time, GL obtained the nonparametric maximum likelihood estimators of the distributions of S and V. Frydman (1992) estimated DIC process nonparametrically under a continuos time nonhomogeneous Markov assumption. Both estimations assumed that periodic examination times were independent of the underlying process.

This motivated the consideration of the test of Markov property of DIC model and interval-censored illness-death model. The following are the main developmens in the draft.

1 Finding the NPMLE $\widehat{F}(s,t)$ of the joint distribution of S and T. Maathuis (2005)

developed the algorithm for obtaining the support rectangles and their probabilities for interval censored (IC) bivariate data. Both DIC data and IC bivariate data are in the form of rectangles. But DIC data may also have right censored observations in state 1. Thus, Maathuis' algorithm can be used to obtain the support rectangles for the DIC data but in general, due to the presence of right censored observations in state 1, cannot be used for estimating the probabilites in DIC data, the support may have to be modified and even if it does not need to be modified, in general Maathuis' algorithm cannot be used for estimating the probability masses over the of support rectangles. A new algorithm developed in Section 3 is used for such estimation.

2 Testing Markov property with.Kendall's coefficient of concordance. Betensky and Finkelstein (BF) (1999) extended Kendall's tau proposed by Oakes (1982) for testing independence of *right censored* bivariate durations to testing independence of bivariate *interval censored* durations. I further extended Kendall's tau to testing Markov property of the DIC model. The main idea of the test rests on the formulation of Markov property as stating that given the present state of a Markov chain, its past and future histories are independent of each other. This formulation was used by Rodriguez-Girondo and Una-Alvarez (2012) to develop the test of Markov property for *right censored* illness-death model.

3 Testing Markov property with logrank statistics in Titman and Putter (TP) (2022)

We adapt the logrank statistic from TP to the DIC model. This involves the use of support rectangles and their estimated probabilities. In the propsed version of TP log rank statistic, an observation which includes a given support rectangle is weighted by the estimated probability of that support rectangle

The draft is organized as follows. In section 2 we describe the DIC data and discuss how

they differ from interval censored bivariate data. Section 3 derives the NPMLE of $F(s,t)$. The nonparametric test of Markov property of the DIC model based on Kendall's tau and the one based on the adapted TP log rank statistic are developed in Section 4. and in Section 4.3 respectively. Section 5 applies both tests of Markov property to Aids data. A comment about the extension of the development in this draft to an interval-censored illness-death model is in Section 6.

2. The data

There are three types of observations that arise from DIC model which are indicated by $\Delta \in \{0,1,2\}$. For the total of $m = 1,2,...,M$ observations, there are three types of observation. Type 1 observation makes $1 \to 2$ transition in an interval $(L_m, R_m]$ and $2 \to 3$ transition in an interval $(W_m, Z_m]$, and thus can be represented in the form of an observation rectangle denoted by O_m :

$$O_m = (L_m, R_m] \times (W_m, Z_m], Z_m < \infty, \Delta_m = 1, \tag{1}$$

where $0 \leq L_m < R_m \leq W_m < Z_m < \infty$. Type 2 observation differs from type 1 observation by having $Z_m = \infty$, and thus being right censored in state 2 at time W_m :

$$O_m = (L_m, R_m] \times (W_m, \infty], \Delta_m = 2 \tag{2}$$

Type 0 observation is right censored in state 1 at time L_m :

$$(L_m, \infty], \Delta_m = 0. \tag{3}$$

Throughout we assume that the observations on (S,T) are independent and that (S,T) are independent of censoring times (L,R,W,Z). Then the likelihood function of the m'th observation

is

$$\mathcal{L}_m = \begin{cases} F(R_m, Z_m) - F(L_m, Z_m) - F(R_m, W_m) + F(L_m, W_m), \Delta_m \in \{1, 2\} \\ 1 - F_1(L_m), \Delta_m = 0, \end{cases} \tag{4}$$

where $F(s,t) = P(S \leq s, T \leq t)$ is the joint distrbution of (S,T) and $F_1(s) = P(S \leq s)$ is the distribution function of S. Let $\mathcal{L}$ be the likelihood function of all observations. Then $\log \mathcal{L} = \sum_{m=1}^{M"'} \sum_{i=0}^{2} I\{\Delta_m = i\} \log \mathcal{L}_m$.

To maximize $\log \mathcal{L}$ we first find the support of the nonparametric maximum likelihood (NPMLE) of $F(s,t)$ from iid observation rectangles in (1)-(2). To do so we make the connection between bivariate interval censored (IC) observations and our doubly IC observations. A bivariate IC observation consists of a pair of random variables (X,Y) in the form of a rectangle $(X_1, X_2] \times (Y_1, Y_2]$, with $0 \leq X_1 < X_2 \leq \infty$ and $0 \leq Y_1 < Y_2 \leq \infty$. A special case of a bivariate IC observation corresponds to a rectangle, $(X_1, \infty] \times (Y_1, Y_2]$, which cannot occur for doubly IC observation in (1)-(2) due to the $R \leq W$ constraint. In an example given in Betensky and Finkelstein (BF) (1999), (X,Y) are the times to the occurrences of two different infections in an Aids patient. However any observation rectangle described in (1-(2) can be a bivariate IC observation. That is, a doubly IC observation is a spcial case of a bivariate IC observations.

3. The NPML estimation of F(s,t)

3.0.1 Characterization of the support of the NPML of F(s,t)

Maathuis (2005) characterized the support of the NPMLE $\widehat{F}(x,y)$ of the joint distribution of a bivariate random variable (X,Y) from the iid observation rectangles. According to the discussion above the same characterization applies to the support of the NPMLE $\widehat{F}(s,t)$, the

joint distribution of a random vriable (S, T). The support of iid observation rectangles, $O_m = (L_m, R_m] \times (W_m, Z_m], 1 \leq m \leq M'$, where M' denotes the number of observations in the data, consists of special disjoint rectangles, called maximal intersection (MI) rectangles, contained in the observation rectangles, where

Definition 1 *$\mathcal{R}_j$ is the MI rectangle if and only if it is a finite intersection of the $O'_m s$ such that for each m, $\mathcal{R}_j \cap O_m = \mathcal{R}_j$ or $\mathcal{R}_j \cap O_m = \emptyset$.*

Figure 1 shows a simple example of four observation rectangles and the corresponding three MI rectangles which are marked by the slanted lines. Maathuis' algorithm in the MLECens package also provides probabilities of the MI rectangles, which we cannot use because we assume that there are right censored observation in state 1. We denote the MI rectangles by $\mathcal{R}_j \equiv (l_j, r_j] \times (w_j, z_j],\ 1 \leq j \leq J \leq M'$, where $l_j < r_j \leq w_j < z_j \leq \infty$, and obtain their probabilities as follows. Let $L_{\max}$ be the largest right censored time in state 1. When $L_{\max} \geq r_{\max} = \max_{1 \leq j \leq J}(r_j)$ F_1, the distribution of S, increases also on $(L_{\max}, \infty]$ and thus we set $f_\infty = F_1(\infty) - F_1(L_{\max})$. If for some j and $m, M' < m \leq M$, $l_j < L_m < r_j$, we modify $(l_j, r_j]$ to be $(L_m, r_j]$ and the corresponding modified MI rectangle is $(L_m, r_j] \times (w_j, z_j]$.

3.1 Likelihood function

Let $\mathcal{R}_j \equiv (l_j, r_j] \times (w_j, z_j],\ 1 \leq j \leq J$, where $l_j < r_j \leq w_j < z_j \leq \infty$, be the J, possibly modified, MI rectangles obtained from the M' observation rectangles and $M - M'$ right censored observations, and let $p_j = P(\mathcal{R}_j) = F(r_j, z_j) - F(r_j, w_j) - F(l_j, z_j) + F(l_j, w_j)$ be the probability of the occurrence of an $\mathcal{R}_j$ MI rectangle. We use the following algorithm for assigning probability masses to MI rectangles and to f_∞..

We express the likelihood factors in (4) in terms of $p = (p_j, 1 \le j \le J)$ and f_∞. For $1 \le m \le M$, we have

$$\mathcal{L}_m = \begin{cases} \sum_{j=1}^J \alpha_j^m p_j, \Delta_m \in \{1,2\} \\ \sum_{j=1}^J \alpha_j^m p_j + f_\infty, \Delta_m = 0, \end{cases} \tag{5}$$

where, for $1 \le j \le J$, α_j^m is an indicator function defined by

$$\alpha_j^m = \begin{cases} 1 \text{ if } (l_j, r_j] \times (w_j, z_j] \subseteq (L_m, R_m] \times (W_m, Z_m],\ \Delta_m \in \{1,2\} \\ 1 \text{ if } (l_j, r_j] \subseteq (L_m, \infty], \Delta_m = 0 \end{cases} \tag{6}$$

We want to maximize $\log \mathcal{L} = \log \mathcal{L}_m$ with respect to (p, f_∞) subject to the conditions: $p_j \ge 0, (1 \le j \le J), f_\infty \ge 0$ and $\sum_{j=1}^J p_j + f_\infty = 1$. Define the Lagrangian

$$H \equiv H(p, f_\infty) = \log \mathcal{L} - a(\sum_{j=1}^J p_j + f_\infty - 1) + \sum_{j=1}^J b_j p_j + d f_\infty,$$

where $a, d, b_j, 1 \le j \le J$, are constants, and

$$b_j p_j = d f_\infty = 0, \sum_{j=1}^J p_j + f_\infty - 1 = 0, a, d, b_j \ge 0. \tag{7}$$

The maximum likelihood estimator (MLE) $(\hat{p}, \widehat{f}_\infty)$ of (p, f_∞) satisfies the following system of equations

$$\frac{\partial H}{\partial p_j} = 0, 1 \le j \le J;\ \frac{\partial H}{\partial f_\infty} = 0 \tag{8}$$

and conditions in (7).

Theorem 2 *Assume that $f_\infty > 0$, then the explicit form of equations in (8) subject to (7) is*

$$M p_j = \sum_{m=1}^{M'} I(\Delta_m = 1) \alpha_j^m p_j / \mathcal{L}_m, 1 \le j \le J, \tag{9}$$

$$M f_\infty = \sum_{m=1}^{M'} I(\Delta_m = 2) f_\infty / \mathcal{L}_m, \tag{10}$$

where $\mathcal{L}_m$ is defined in (5)-(6).

Proof. Proof is given in Appendix A. ■

The equations (9-10) have a natural interpretation. The left hand side of (9) is the expected number of observations that make $1 \to 2$ transition in $(l_j, r_j]$ and $2 \to 3$ transition in $(w_j, z_j]$ if $z_j < \infty$ or right censored if $z_j = \infty$, whereas the m'th term in the summation on the right hand side is the probability of an m'th observation making $1 \to 2$ transition in $(l_j, r_j]$ and $2 \to 3$ transition in $(w_j, z_j]$ conditional on the available information about that observation. Similarly, the left hand side of (10) is the expected number of type 2 observations making the $1 \to 2$ transition in $(L_{mx}, \infty]$ and the m'th term in the summation on the right hand side is the probability that the m'th observation makes the $1 \to 2$ transition in $(L_{mx}, \infty]$ conditional on the available information about that observation. All expectations and conditional probabilities are evaluated under the true parameter values (p, f_∞). We note that equations (9-10) can also be interpreted as the self-consistency equations, see Efron (1967).

We use (9-10) to obtain the maximum likelihood estimates of p and f_∞. We substitute an initial guess $p^0 = (p_j^0, 1 \le j \le J)$, f_∞^0 for p and f_∞ into the right hand sides of (9-10) to obtain the updated values p^1 and f_∞^1 from the left hand sides of (9-10.) We iterate in this way until the convergence criterion is satisfied: we stop at the i'th iteration if for $1 \le j \le J, |p_j^i - p_j^{i-1}| < \epsilon$ and $|f_\infty^i - f_\infty^{i-1}| < \epsilon$, where ϵ is a small number. In our experience, if $\epsilon = 10^{-3}$ then it takes about 30 iterations and if $\epsilon = 10^{-5}$, about 300 iterations for the algorithm to converge. The natural choice of initial values is $p_j^0 = 1/(J+1), 1 \le j \le J$ and $f_\infty^0 = 1/(J+1)$.

4. Test of Markov property

4.1 Background

The tests of Markov property for doubly interval-censored data is based on extension of Kendall's coefficient of concordance, referred to below as Kendall's τ. Kendall's population τ is a measure of association between two random variables (X, Y) from a bivariate population and it involves the notion of concordance and discordance of two independent pairs of observations $(X_i, Y_i), (X_j, Y_j)$. Two independent pairs $(X_i, Y_i), (X_j, Y_j)$ are said to be concordant if $(X_j - X_i)(Y_j - Y_i) > 0$ and discordant if $(X_j - X_i)(Y_j - Y_i) < 0$. Kendall's population tau is defined as $\tau = \pi_c - \pi_d$, where, $\pi_c = P[(X_j - X_i)(Y_j - Y_i) > 0]$ is the probability of concordance and $\pi_d = P[(X_j - X_i)(Y_j - Y_i) < 0]$ is the probability of discordance. τ is equal to zero if X and Y are independent random variables.

One estimates Kendall's τ from a random sample of pairs by counting the number of concordant and discordant pairs. The concordant pairs get a score of 1, discordant a score of -1 and the pairs in which there is an equality among either variable get a score of 0. Kendall's τ is the average of the scores over all pairs of observations. It estimates the difference between the probabilities of concordance and discordance.

Oakes (1982) proposed an extension of Kendall's tau to right censored bivariate observations and Betensky and Finkelstein (BF) (1999) extended it to bivariate IC observations. We in turn apply Kendall's τ to DIC observations. As discussed in BF, the difficulty with the direct application of Kendall's tau to bivariate IC data is that if a pair of observation rectangles overlaps it is not possible to decide on the ordering of the bivariate pair of times. This may also happen when the observation rectangles do not overlap. BF referred to such observation

rectangles as incomparable. An example of concordant, discordant and incomparable pairs of rectangles is in Figure 2. BF proposed an imputation method for comparison of incomparable rectangles. The variant of the imputation proposed in BF and the extension of Kendall's tau to doubly interval-censored data are discussed below.

4.2 Testing Markov property of DIC model using Kendall's tau

The Markov assumption has been widely used in multi-state modeling. It states that given the present state of a Markov chain, its future evolution does not depend on its past history. An alternative way of stating Markov property is: given a present state of a Markov chain (X_t), its past and future histories are independent.

We propose a test of Markov property for DIC process based on the alternative statement of Markov property. For DIC model, the statement takes the form: S and T are independent random variables conditionally on $X_t = 2$. One can then carry out the test of independence of S and T at different values values of t for which $X_t = 2$. When there is no censoring, for every selected value of t with $X_t = 2$, there is a subsample of observations $(S_i, T_i), 1 \leq i \leq M_t$, for which $S_i < t < T_i$. This subsample can then be used to test the independence of S and T using the standard Kendall's tau.

For DIC data, the support of $\widehat{F}(s,t)$ consists of the MI rectangles $\mathcal{R}_j, 1 \leq j \leq J$, with the probabilities $\widehat{p}_j = P(\mathcal{R}_j)$, $1 \leq j \leq J$ obtained from Maathuis' algorithm or, if $f_\infty > 0$, the probabilities are obtained from the algorithm in Theorem 1. For a given t, we select the subsample satisfying $S < t < T$ through the following procedure. In general an observation rectangle, O_m may contain a number of MI rectangles. O_m will be chosen for time-t subsample if it contains an MI rectangle, $R = (l, r] \times (w, z)$, for which $r ¡ t \leq w$.Mtim echo and if one

of them is time -t rectangel, thatwe have $\widehat{P}(O_m) = \sum_{j:\mathcal{R}_j \subset O_m} \widehat{p}_j$. If among ths to be in the time t-subsample if it contains Such a rectangle may contain other MI rectangles. In general , and in partculat and impute an MI rectangle by drawing it from the multinomial distribution with values and the probabilities given by $\{j, \widehat{p}_j/\widehat{P}(O_m), j : \mathcal{R}_j \subset O_m\}$. An imputed rectangle, $(l, r] \times (w, z)]$, is selected for the time t subsample if it satisfies $r¡t \leq w$, and we refer to it as time t-MI rectangle. In the k'th imputation, $1 \leq k \leq K$, where K is the number of imputations, we assign a score $\psi_{ij,t}^{(k)} = 1, 0$ or -1 to the of (i,j)'th pair of the observation rectangles if a pair of imputed time t-MI rectangles is concordant, discordant, or incomparable, respectively. Let $M_t^{(k)}$ be the size of the time t subsample in the k'th imputation. Then $\hat{\tau}_{t,k}$, Kendall's tau for the k'th imputation at time t, is

$$\hat{\tau}_{t,k} = \binom{M_t^{(k)}}{2}^{-1} \sum_i \sum_{j>i} \psi_{ij,t}^{(k)},$$

where $\binom{M_t^{(k)}}{2}$ is the total number of pairs so that $\hat{\tau}_{t,k}$ is the average of $\psi_{ij,t}^{(k)}$ over all pairs. As described in Oakes (1982), the mean and variance of $\hat{\tau}_{t,k}$ under $H_0 : \tau_{t,k} = 0$, are

$$\begin{aligned} \mu_{\hat{\tau}_{t,k}} &= 0 \\ \sigma^2_{\widehat{\tau}_{t,k}} &= \frac{2E[(\psi_{ij,t}^{(k)})^2]}{M_t^{(k)}(M_t^{(k)} - 1)} + \frac{4(M_t^{(k)} - 2)E(\psi_{ij,t}^{(k)}\psi_{ik,t}^{(k)})}{M_t^{(k)}(M_t^{(k)} - 1)}, \end{aligned}$$

where $E(\psi_{ij}^{(k)})^2$ and $E(\psi_{ij}^{(k)}\psi_{ik}^{(k)})$ can be estimated from the data. For a total of K imputations, the estimator of τ_t is $\hat{\tau}_t = \sum_{k=1}^K \hat{\tau}_{t,k}/K$. As in Oakes (1982), the asymptotic normality of $\hat{\tau}_k$ follows from results in Hoeffding (1948) and its variance is.

$$\sigma^2_{\hat{\tau}_t} = \sum_{k=1}^K \frac{\sigma^2_{\hat{\tau}_{t,k}}}{K} + (1 + \frac{1}{K}) \sum_{k=1}^K \frac{(\hat{\tau}_{t,k} - \hat{\tau}_t)^2}{K - 1}. \tag{11}$$

The first term in above is the within imputation variance and the second is the between-imputation variance, see Rubin (1987). We use the statistic $\hat{\tau}_t/\sqrt{\sigma^2_{\widehat{\tau}_t}}$, which has the standard normal distribution under $H_0 : \tau_t = 0$, to conduct the test of Markov property. One increases

the number of imputations until the estimation results stabilize. Note that for a given t, the subsample size $M_t^{(k)}$ may differ from imputation to imputation. Since one needs a reasonably large sample for testing Markov property at a given time t, time t is selected as appropriate for the test if the average sample size $\sum_{k=1}^{K} M_t^{(k)}/K$ from K imputations is at least around 50 observations. Further note that the rejection of Markov property for any value of t would suggest that the underlying process is not Markovian.

4.3 Testing Markov property of DIC model using Titman-Putter log rank statistic.

Titman and Putter (2022) proposed a new test of Markov property in which the null hypothesis says that the process is Markov and is tested against general alternatives. More precisely, if $X = (X_u, u \geq 0)$ is the process under consideration, $H_0 : \alpha_{lm}(t|X_s = j) = \alpha_{lm}(t|X_s \neq j)$, where $\alpha_{lm}(t|X(s) = j)$ and $\alpha_{lm}(t|X(s) \neq j)$, $t > s$, are the intensities of the $l \to m$ transition conditional on $X_s = j$ and $X_s \neq j$ respectively. The statistic given in Titman and Putter for testing H_0 is

$$\sum_{i=1}^{n} \int_s^{\tau} \left\{ \delta_i^{(j)}(s) - \frac{\sum_k \delta_k^{(j)}(s) Y_{kl}(t)}{\sum_k Y_{kl}(t)} \right\} dN_i^{lm}(t),$$

where in $\delta_i^{(j)}(s) = I(X_i(s) = j)$, i refers to the i'th subject and $I(\cdot)$ is an indicator function. Further, $Y_{kl}(t) = I(X_k(t-) = l)Y_k(t)$ is the at risk indicator of $l \to m$ transition for subject k and τ is the maximum time of follow-up. Finally, $N_i(t)$ counts the number of $l \to m$ transitions by time t and $dN_i(t) = 1$ if the i'th subject makes the $l \to m$ transition at time t and is equal to 0 otherwise. To show that above statistics is of log rank type we rewrite it as follows

$$\int_s^\tau \left\{ \sum_{i=1}^n \delta_i^{(j)}(s) dN_i^{lm}(t) - \sum_k \delta_k^{(j)}(s) Y_{kl}(t) \frac{\sum_{i=1}^n dN_i^{lm}(t)}{\sum_k Y_{kl}(t)} \right\}, \tag{12}$$

where now the first term in parentheses is the total number of subjects that were in state j at time s and made $l \to m$ transition at time t, $\sum_{i=1}^n dN_i(t) / \sum_k Y_{kl}(t)$ is the estimated intensity of $l \to m$ transition at time t based on all subjects, and $\sum_k \delta_k^{(j)}(s) Y_{kl}(t)$ is the total number of subjects who were in state j at time s and are at risk of $l \to m$ transition at time t. Under H_0 the left and right terms should be close to each other in statistical terms.

In DIC model the only choice for states j, l, and m, is $j = 1, l = 2$, and $m = 3$. The null hypothesis for testing Markov property of the DIC model is $H_0 : \alpha_{23}(t|X(s) = 1) = \alpha_{23}(t|X(s) = 2)$, $t > s$. We adapt **??**) to DIC model as follows.

We define a p*roper* MI rectangle as the one with a finite time of $2 \to 3$ transition. The set of such rectangles together with their estimated probabilities is: $P_U = \{\mathcal{R}_u \equiv (l_u, r_u] \times (w_u, z_u], z_u < \infty, \widehat{p}_u = P(\mathcal{R}_u),$ where $z_u < \infty$ signifies that a MI rectangle is proper, U is the number of proper MI rectangles and $\widehat{p}_u = P(\mathcal{R}_u)$ is the estimated probability of u'th proper MI rectangle. The improper MI rectangles have right censored times of $2 \to 3$ transition and form a set $I_E = \{\mathcal{R}_u \equiv (l_u, r_u] \times (w_u, \infty], \widehat{p}_u = P(\mathcal{R}_u), 1 \le u \le E\}$.

For a given rectangle in P_U we find observation rectangles which include it. These are clearly observation rectangles of type 1 with $Z_i < \infty$, see (**??**), and we refer to them as proper observations or proper observation rectangles. We note that each rectangle in P_U may belong to one or more proper observations. Conversely, a proper observation may include more than one rectangle from MIR_U set. However, as can be seen from comparing information in Tables 1 and 2, the converse does not happen in case of Aids data. Therefore for now, we will assume that each proper observation contains only one MI rectangle from P_U. However, an improper

observation rectangle may include a number of improper MI rectangles. Thus, as in first part of the paper, we will use imputation to compute the test statistic..

Next we reduce each MI rectangle in P_U to (r_u, z_u), $1 \leq u \leq U$, thus assuming that $1 \to 2$ and $2 \to 3$ transitions occur at the endpoints of their respective intervals and denote its probability by $\widehat{p}_u = \widehat{p}(r_u, z_u)$. The set of reduced proper MI rectangles together with their probabilities is $R_U = \{(r_u, z_u), z_u < \infty, \widehat{p}_u = \widehat{p}(r_u, z_u), 1 \leq u \leq U \}$. Consequently each proper observation rectangle which includes the u'th proper MI rectangle is also reduced to (r_u, z_u). We refer to such an observation as a reduced proper observation. We also define the set of reduced improper MI rectangles as $R_E = \{(r_u, w_u), \widehat{p}_u = \widehat{p}(r_u, w_u), 1 \leq u \leq E\}$ where w_u is the u'th censoring time.

Suppose that in the set R_U, there are $V \leq U$ different values of $z_u, 1 \leq u \leq U$, which we denote by $(t_1, t_2, ..., t_V)$. Then the set R_U becomes $R_V = \{\cup_{j=1}^{J_v}[(r_{v_j}, t_v)], 1 \leq v \leq V \}$, where for a given t_v, J_v =number of different times of $1 \to 2$ transition and r_{v_j} is the j'th time of $1 \to 2$ transition.

In (12) the integral is over the time t of $2 \to 3$ transition. In our setup, the integral becomes the sum over t_v, $1 \leq v \leq V$ and the test statistic becomeswhich we simplify to

$$\sum_{v=1}^{V} \left\{ \sum_{i=1}^{M} \delta_i^{(1)}(s) dN_i(t_v) - \left(\sum_k \delta_k^{(1)}(s) Y_k(t_v) \right) \frac{\sum_{i=1}^{M} dN_i(t_v)}{\sum_k Y_k(t_v)} \right\} \tag{13}$$

with the understanding that $N(t)$ is a counting process for $2 \to 3$ transition by time t, and $Y(t)$ is the number at risk for $2 \to 3$ transition at time t-. Here s has to be chosen chosen so that for some proper reduced observations, $X(s) = 1 \Leftrightarrow \delta_i^{(1)}(s) = 1$, and for some other proper reduced observations, $X(s) = 2 \Leftrightarrow \delta_i^{(1)}(s) = 0$. Otherwise, (13) would be equal to zero.

We next compute the quantities in (13) referring to patients instead of observations. For a

given t_v,

$$\sum_{i=1}^{M} \delta_i^{(1)}(s) dN_i(t_v) = \sum_{i=1}^{M} I(s \leq r_v^i < t_v) dN_i(t_v) \tag{14}$$

is the number of patients who were in state 1 at time s and made $2 \to 3$ transition at time t_v, while

$$\sum_{i=1}^{M} dN_i(t_v) \tag{15}$$

is the total number of patients who made $2 \to 3$ transition at time t_v. However, to compute $Y_k(t_v)$ for right censored observations in state 2, we use imputation procedure described in the first part of the paper. Let $\mathcal{R}_{i,k} \equiv (l_{i,k}, r_{i,k}] \times (w_{i.k}, \infty]$ be the imputed improper MI rectangle for the i'th improper observation rectangle in the k'th imputation. We reduce this rectangle to $(r_{i.k}, w_{i,k})$. Then

$$\begin{aligned} \sum_{i=1}^{M} Y_i(t_v) &= \sum_{i=1}^{M} \left[\sum_{p=0}^{V-v} I(r_{v+p}^i < t_v) dN_i(t_{v+p}) \right]^{\Delta_i=1} \\ &\quad + \sum_{i=1}^{M} \left[I(r_{i,k} < t_v < w_{i,k}) \right]^{\Delta_i=0} \end{aligned} \tag{16}$$

is the k'th imputation total number of patients who were at risk of $2 \to 3$ transition at time $t_v -$. Here $\Delta_i = 1$ indicates that the i'th patient made $1 \to 2$ transition at time r_{v+p}^i before time t_v and $2 \to 3$ transition at time $t_{v+p} \geq t_v$, thus contributing to the risk set at time t_v-. And $\Delta_i = 0$ indicates that the patient made $1 \to 2$ transition at time $r_{i.k}$ before time t_v and was right censored in state 2 at time $w_{i,k}$ greater than t_v again contributing to the risk set at time t_v-. Now

$$\begin{aligned} \sum_{i=1}^{M} \delta_i^{(1)}(s) Y_i(t_v) &= \sum_{i=1}^{M} \left[\sum_{p=0}^{V-v} I(s \leq r_{v+p}^i < t_v) dN_i(t_{v+p}) \right]^{\Delta_i=1} \\ &\quad + \sum_{i=1}^{M} \left[I(s \leq r_{i,k} < t_v < w_{i,k}) \right]^{\Delta_i=0}, \end{aligned} \tag{17}$$

is the k'th imputation total number of patients who were in state 1 at time s and later were at risk for $2 \to 3$ transition at time $t_v -$.

Note that $\sum_{i=1}^{M} Y_i(t_v) = \sum_{i=1}^{M} \delta_i^{(1)}(s) Y_i(t_v) + \sum_{i=1}^{M} \delta_i^{(2)}(s) Y_i(t_v)$, where $\delta_k^{(2)} = I(X(s) = 2)$ and

$$\sum_{i=1}^{M} \delta_i^{(2)}(s) Y_i(t_v) = \sum_{i=1}^{M} \left[\sum_{p=0}^{V-v} I(r_{v+p}^i < s < t_v) dN_k(t_{v+p}) \right]^{\Delta_i=1}$$
$$+ \sum_{i=1}^{M} \left[I(r_{i,k} < s < t_v < w_{i,k} \right]^{\Delta_i=0}. \tag{18}$$

For computation of (13) we need (14)-(17). For computation of the variance of (13) we will also need (18).

5. Application to Aids data

We consider Aids data set from Kim et. al (1993), which is an updated versions of the data set considered in GL and Frydman(1992). The data set arose from the study that involved a cohort of persons with hemophilia for whom the time of infection with HIV and the time of the onset of the Aids were both interval censored.

During the period from the beginning of 1978 to August 1988, 257 persons with Type A or B hemophilia have been treated at two hospitals in France. These persons were at the risk of HIV infection because they were treated with contaminated blood transfusions. They were divided into two groups of 104 heavily and 153 lightly treated with heavily treated group receiving more blood transfusions. The data were obtained from the periodically stored blood samples and is reproduced in Table 1 for heavily treated group. The MI rectangles for this data are shown in Table 2.

5.1 Results based on Kendall's tau

Note that in this dataset, $L_{\max} = 17 > r_{\max} = 16$, so the algorithm from Theorem 2 was used for estimation of the probability masses over the MI rectangles. Table 3 presents the results of the Markov property test at integer values of $t \in [13, 19]$. For each t, we did a 1000 imputations to select the subsample that satisfies $S \leq t < T$. The results show that we cannot reject Markov assumption at $\alpha = 0.05$ for any value of t. This conclusion suggests that a Markov chain could be an appropriate model for describing the progression of a disease from (only hemophilia) to (hemophilia with HIV) and finally to (hemophilia with Aids symptoms present).

5.2 Results based on extension of Titman and Putter's test statistics.

The MI rectangles for Aids data were obtained after 8 right-censored observations in state 1 were removed from the data. There are 25 MI rectangles of which 17 are proper MI rectangles and 8 are not proper. Among 17 proper rectangles there are 11 different times of $2\rightarrow$ 3 transition. These are 12, 13,15,16,17,18,19,20,21,22,23. The computation of test statistic for $t_1 = 12$

6. Extension to an interval censored illness-death model

For an interval censored illness-death model, Markov property test is based on the same idea of conditional independence between S and T, where S and T are defined in the same way as for a progressive model. What is new is that there is a direct $1 \rightarrow 3$ transition, the time of which is interval censored. This transition has to be taken into account when deriving the NPMLE of the joint distribution of S and T. I can provide the details.

References

Betensky, R. A. and Finkelstein, D. (1999) An extension of Kendall's coefficient of concordance to bivariate interval censored data. *Statistics in Medicine* **18**, 3101-3109.

Commenges D. (2002) Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research.* **11(2):**167-182

De Gruttola, V. and Lagakos, S. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics* **45**, 1-11.

Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the 5th Berkeley Symposium* (Volume **4**), 831-853. Berkeley: University of California Press.

Frydman, H. (1992). A nonparametric estimation procedure for a periodically observed three state Markov process, with application to AIDS. *Journal of the Royal Statistical Society,* Series B **54**, 853-66.

Frydman, H. (1995). Nonparametric estimation of a Markov "illness-death" process from interval-censored observations, with application to diabetes survival data. *Biometrika* **82**, 773-89.

Frydman, H.and Szarek, M. (2009). Nonparametric Estimation in a Markov "illness-death" process from interval censored observations with missing intermediate transition status. *Biometrics.***65**, 143-151.

Gomez, G. and Calle, M. L. (1999). Nonparametric estimation with doubly censored data. *Journal of Applied Statistics* **26,** 45-58.

Grüger, J., Kay, R. and Schumacher, M. (1991). The validity of inferences based on incomplete of observations in disease state models. *Biometrics* **47**, 595-605

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19,** 293-325**.**

Harezlak, Gao and Hui (2003). An illness–death stochastic model in the analysis of longitudinal dementia data. *Statistics in Medicine* **22**, 1465–1475.

Joly, P., D. Commenges, D., Helmer, C., Letenneur, L. (2002) A penalized likelihood approach for an illness–death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3**, 433–443

Kay, R. (1986) A Markov model for analyzing cancer markers and disease states in survival studies. *Biometrics* **42**, 855-865.

Kendall, M.and Gibbons, J. D. (1990).*Rank Correlation Methods* (5th edition). Edward Arnold.

Kim, Y. M., De Gruttola, V.G., Lagakos, S.W. (1993) Analyzing Doubly Censored Data with Covariates, with Application to AIDS. *Biometrics* **49**, 13-22.

Maathuis, M. H. (2005) Reduction Algorithm for the NPMLE for the Distribution Function of Bivariate Interval-Censored Data. *Journal of Computational and Graphical Statistics* **14**, 352-362.

Oakes, D. (1982) A concordance test for independence in the presence of right censoring. *Biometrics* **38,** 451-455.

Scholz, S., Strub, J.-R., Gerds, T. (2007). Clinical proof of facing composite Signum in conical-crown-retained dentures. Results on an average observation period of 3 years. *Deutsche Zahnärztliche Zeitschrift* **62**, 176-179.

Pan, W. (2001) A multiple imputation approach to regression analysis for doubly censored data with applications to AIDS studies. *Biometrics* **57**, 1245-1250.

Peto, R. (1973) Experimental survival curves for interval-censored data. *Applied Statistics* **22**, 86-91.

Rodriguez-Girondo, M. and de Una-Alvarez, J. (2012) A nonparametric test for Markovianity in the illness-death model. *Statistics in Medicine* **31,** 4416-4427.

Rubin, D..(ed.) (1987) Multiple imputation for nonresponse in survey, New York: John Wiley & Sons, 1-23.

Sun, J., Lim, H. J., Zhao, X.(2004) An Independence test for doubly censored failure time data. *Biometrical Journal* **46,** 503-511**.**

Sun, J. (2006) *The Statistical Analysis of Interval-censored Failure Time Data* (Statistics for Biology and Health), Springer.

7. Appendix A: Proof of Theorem 1

The equations in (8) are

$$\frac{\partial \log L}{\partial p_j} + b_j - a = 0, 1 \leq j \leq J \tag{19}$$

$$\frac{\partial \log L}{\partial f_\infty} + d - a = 0. \tag{20}$$

We multiply the j'th equation in (19) by p_j and add the resulting J equations to get

$$\sum_{j=1}^{J} p_j \frac{\partial \log L}{\partial p_j} + \sum_{j=1}^{J} b_j p_j - a \sum_{j=1}^{J} p_j = 0, \tag{21}$$

and we multiply (20) by f_∞

$$f_\infty \frac{\partial \log L}{\partial f_\infty} + d f_\infty - a f_\infty = 0. \tag{22}$$

Now adding equations (21)-(22) and taking into account the constraints in (7) we obtain

$$\sum_{j=1}^{J} p_j \frac{\partial \log L}{\partial p_j} + f_\infty \frac{\partial \log L}{\partial f_\infty} - a = 0. \tag{23}$$

Using (5) we have

$$\sum_{j=1}^{J} p_j \frac{\partial \log L}{\partial p_j} = \sum_{j=1}^{J} \sum_{m=1}^{M'} (\Delta_m \in \{1,2,3\}) \frac{\alpha_j^m p_j}{L_m}, \tag{24}$$

and

$$f_\infty \frac{\partial \log L}{\partial f_\infty} = \sum_{m=1}^{M'} I(\Delta_m = 3) \frac{f_\infty}{L_m}. \tag{25}$$

We next add (24) and (25) after first interchanging the summations in (24)

$$\begin{aligned}\sum_{m=1}^{M'} I(\Delta_m \in \{1,2\}) \sum_{j=1}^{J} \frac{\alpha_j^m p_j}{L_m} + \sum_{m=1}^{M'} I(\Delta_m = 3) \frac{1}{L_m} \left(\sum_{j=1}^{J} \alpha_j^m p_j + f_\infty \right) \\ = \sum_{m=1}^{M'} I(\Delta_m \in \{1,2\}) + \sum_{m=1}^{M'} I(\Delta_m = 3) = M',\end{aligned}$$

which when compared with equation (23) shows that $a = M'$. Multiplying equation (19) by p_j gives

$$p_j = \frac{1}{M'} \sum_{m=1}^{M'} I(\Delta_m \in \{1,2,3\}) \alpha_j^m p_j / L_m, 1 \leq j \leq J,$$

and multiplying (20) by f_∞ gives

$$f_\infty = \frac{1}{M'} \sum_{m=1}^{M'} I(\Delta_m = 3) f_\infty / L_m,$$

which are equations in the statement of Theorem 1.

Table 2: The MI rectangles for AIDS data

j	l_j	r_j	w_j	z_j	$\hat{p}_j$	j	l_j	r_j	w_j	z_j	$\hat{p}_j$
1	5	7	23	∞	0.0437	14	10	12	16	17	0.0104
2	5	7	13	13	0.0208	15	11	12	23	∞	0.0379
3	3	7	17	17	0.0104	16	9	12	22	22	0.0104
4	1	7	16	16	0.0104	17	10	12	23	23	0.0104
5	5	7	11	12	0.0208	18	12	13	23	∞	0.1682
6	7	9	19	19	0.0104	19	12	13	20	20	0.0104
7	8	9	23	∞	0.0235	20	13	14	17	18	0.0312
8	9	10	23	∞	0.0639	21	13	14	20	21	0.0208
9	9	10	15	15	0.0156	22	13	14	19	20	0.0104
10	10	11	23	∞	0.1649	23	14	15	23	∞	0.1548
11	10	11	15	16	0.0312	24	14	15	23	23	0.0104
12	9	11	17	18	0.0312	25	15	16	23	∞	0.0619
13	10	11	14	15	0.0156						

Table 3: The Markov property test results for heavily treated group ($K = 1000$)

t	$\hat{\tau}$	s.d. ($\hat{\tau}$)	$\frac{\hat{\tau}}{s.d.(\hat{\tau})}$	Average number of observations
11	0.082	0.062	1.320	42
12	0.082	0.054	1.522	49
13	0.090	0.048	1.892	64
14	0.012	0.045	0.265	69
15	0.033	0.030	1.088	80
16	0.028	0.029	0.972	77
17	0.000	0.017	0.024	77
18	0.000	0.017	0.024	75
19	-0.012	0.014	-0.854	75