Likelihood for interval-censored observations from multi-state models

Daniel Commenges

LIKELIHOOD FOR INTERVAL-CENSORED OBSERVATIONS FROM MULTI-STATE MODELS

Daniel Commenges[1]

[1] INSERM Team of Biostatistics, ISPED, 146 rue Léo Saignat, Bordeaux, 33076, France

E-mail: daniel.commenges@isped.u-bordeaux2.fr

SUMMARY: We consider the mixed dicrete-continuous pattern of observation in a multi-state model; this is a classical pattern because very often clinical status is assessed at discrete visit times while time of death is observed exactly. The likelihood can easily be written heuristically for such models. However a formal proof is not easy in such observational patterns. We give a rigorous derivation of the likelihood for the illness-death model based on applying Jacod's formula to an observed bivariate counting process.

Key Words: multi-state models; illness-death; counting processes; ignorability; interval-censoring; Markov models.

1. INTRODUCTION

Multi-state models are a generalisation of survival and competing risks models. In epidemiology, multi-state models are used to represent the evolution of subjects through different statuses, generally including clinical statuses and death. Clinical statuses of subjects are often observed at a finite number of visits. This leads to interval-censored observations of times of transition from one state to another. A classical reference for multi-state models is Andersen et al. (1993). This book however essentially treats right-censored observations: building estimators by decomposing the observed processes and equating to zero the martingale term is very elegant in that case but this does not work for interval-censored observations.

One first issue is whether the mechanism leading to these incomplete observations is ignorable. If this is the case, the likelihood can be written heuristically in terms of both transition probabilities and transition intensities. In homogeneous Markov models, transition probabilities can be expressed simply in terms of transition intensities but this is not the case in more general multi-state models. In addition, inference in homogeneous Markov models is easy because these are parametric models. Non-parametric approaches to non-homogeneous Markov models may follow two paths: one is the completely non-parametric approach and can be seen as a generalisation of the Peto-Turnbull approach (Turnbull, 1976); the other implies a restriction to smooth intensities models. In particular, the penalized likelihood method has been applied to this problem. A review of this topic can be found in Commenges (2002). However all these approaches are based on likelihoods which have been given only heuristically. In the complex setting of observations from multi-state models involving a mixed pattern of continuous and dicrete time observations it is important to have a rigorous derivation of the likelihood.

In section 2 we describe the possible patterns of observation from multi-state models, especially those which are relevant in epidemiology, and then we give the heuristic formulas for the likelihood. We begin section 3 by describing the theoretical basis of likelihood, Jacod's formula for the likelihood ratio for a counting process and a way to apply it to incomplete observations; we give a rigorous derivation of the likelihood for the illness-death model, based on a representation of this model by a bivariate counting process and applying Jacod's formula to an observed bivariate counting process.

2. GENERALITIES ON INFERENCE

2.1 PATTERNS OF OBSERVATION

Generally we will represent the status of a subject i by a stochastic process X_i; $X_i(t)$ can take a finite number of values $\{0, 1, \ldots, K\}$ and we can make more or less stringent assumptions on the process, for instance, time homogeneity, Markov or semi-Markov properties. Multi-State processes are characterized by transition intensities or transition probabilities between states h and j that we will denote respectively by $\alpha_{hj}(t; \mathcal{F}_{t-})$ and $p_{hj}(s,t) = P(X(t) = j | X(s) = h, \mathcal{F}_{s-})$, where $\mathcal{F}_{s-}$ is the history before s; for Markov processes the history can be ignored.

We may consider that the state of the process i is observed at only a finite number of times $V_0^i, V_1^i, \ldots, V_m^i$. This typically happens in cohort studies where fixed visit times have been planned. In such cases the exact times of transitions are not known; it is only known that they occurred during a particular interval; these observations are said to be interval-censored. It is also possible that the state of the process is not exactly observed but it is known that it belongs to a subset of $\{0, 1, \ldots, K\}$.

The most common pattern of observation is in fact a mixing of discrete and continuous time observations. This is because most multi-state models include states which represent clinical status and one state which represents death: most often clinical status is observed at discrete times (visits) while the (nearly) exact time of death can be retrieved. This is the case in the study of dementia by Joly et al. (2002) where an irreversible illness-death model (see Figure 1) was used and dementia was assessed only at planned visits. Note that in the irreversible model no transition from state 1 to state 0 is possible, which is well adapted to modelling dementia, considered as an irreversible clinical condition.

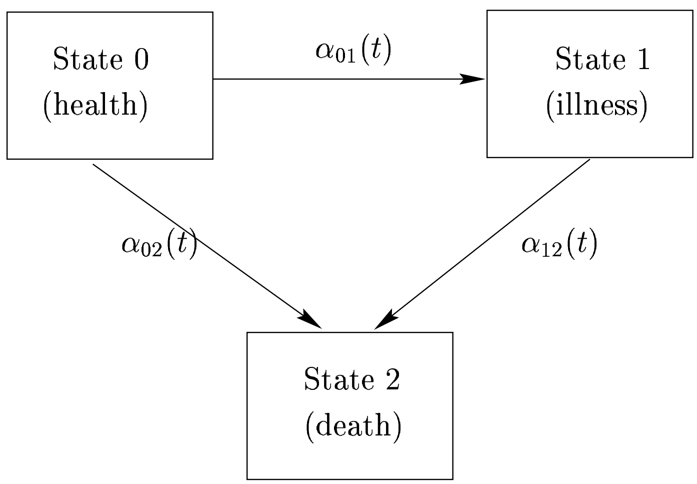

Figure 1: Illness-death model

In all cases we should have a model describing the way the data have been observed. For writing reasonably simple likelihoods, there must be some kind of independence of the mechanisms leading to incomplete observations relative to the process itself. A simple likelihood can be written if the observation times are fixed. More realistically, the observation process should be considered as random and intervene in the likelihood.

The mechanism leading to incomplete data will be said to be ignorable if the likelihood treating the observation process as non-random leads to the same inference as the full likelihood. An instance where this works is the case of observation processes completely independent of the processes of interest X_i. A general approach for representing the observation of a process X_i is to consider a process R_i which takes value 1 at t if $X_i(t)$ is observed, 0 otherwise. R_i must satisfy certain independence properties relatively to X_i in order to be ignorable; in that case one can write the likelihood as if R_i was fixed. In the remaining of this paper we will assume that this is the case that the mechanism leading to incomplete observation is ignorable: we shall write the likelihood as if the discrete observation times and the right censoring variable were fixed.

2.2 INFERENCE

The first interesting fact to be noted is that with continuous observation times, the inference problem in a multi-state model can be decoupled into several survival problems; with discrete-time observation (leading to interval-censoring), this is no longer possible. The likelihood for the whole observation of the trajectory must be written as in Joly and Commenges (1999); Joly et al. (2002) gave an example of the bias that occurs when one tries to treat interval-censored observation from an illness-death model as a survival problem.

We shall give the likelihood for interval-censored observations of a single process X taken at $V_0, V_1, \ldots, V_m$, (treating the V_j as fixed); for sake of simplicity we drop the index i. If we have a sample of size n the processes X and the observation times should be indexed by i; assuming the independence of the processes (the histories of the "subjects") the likelihood is the product of the individual likelihoods. For sake of simplicity we will also restrict to Markov models. So, for purely discrete-time observations this individual likelihood is as follows:

$$\mathcal{L} = \prod_{r=0}^{m-1} p_{X(V_r),X(V_{r+1})}(V_r, V_{r+1}),$$

where $p_{hj}(s,t) = P(X(t) = j | X(s) = h)$.

Variants of this likelihood can be written in cases of mixing of continuous and discrete-time observations. We give the likelihood when the process is observed at discrete times but time of transition towards one absorbing state, representing generally death, is exactly observed or right-censored, a common model and observational pattern in epidemiology. Denote by K this absorbing state. Observations of X are taken at $V_0, V_1, \ldots, V_L$ and the vital status is observed until C ($C \geq V_L$); here V_L is the last visit time of an alive subject. Let us call $\tilde{T}$ the follow-up time that is $\tilde{T} = \min(T, C)$, where T is the time of death; we observe $\tilde{T}$ and $\delta = I\{T \leq C\}$. For continuous intensities model the likelihood can be written:

$$\mathcal{L} = \Big[\prod_{r=0}^{L-1} p_{X(V_r),X(V_{r+1})}(V_r,V_{r+1})\Big] \sum_{j \neq K} p_{X(V_L),j}(V_L,\tilde{T})\alpha_{j,K}(\tilde{T})^{\delta}.$$

This likelihood can be understood intuitively as the "probability" of the observed trajectory but it is not so easy to prove that this is really the likelihood, as we shall see in the next section. For this likelihood to be useful, it must be expressed in term of the transition intensities which are the basic parameters of the model; so we must be able to express the transition probabilities in term of the transition intensities. This is particularly easy in the homogeneous Markov model. In other models it generally requires the computation of integrals.

Let us now specialize these formulas to the illness-death model, a model with the three states "health", "illness", "death" respectively labelled $0, 1, 2$. If the subject starts in state "health", has never been observed in the "illness" state and was last seen at visit L (at time V_L) the likelihood is:

$$\mathcal{L} = p_{00}(V_0, V_L)[p_{00}(V_L, \tilde{T})\alpha_{02}(\tilde{T})^{\delta} + p_{01}(V_L, \tilde{T})\alpha_{12}(\tilde{T})^{\delta}]; \tag{1}$$

if the subject has been observed in the illness state for the first time at V_J then the likelihood is:

$$\mathcal{L} = p_{00}(V_0, V_{J-1})p_{01}(V_{J-1}, V_J)p_{11}(V_J, \tilde{T})\alpha_{12}(\tilde{T})^{\delta}. \tag{2}$$

This equations are valid for the reversible as well as for the irreversible illness-death model. In Markov models, the transition probabilities are linked to the transition intensities by the Kolmogorov differential equations. For the irreversible illness-death model, to which we shall specialize from now on, the forward Kolmogorov equation gives:

$$\frac{dp_{00}}{dt}(s,t) \quad = \quad -p_{00}(s,t)[\alpha_{01}(t) + \alpha_{02}(t)]$$

$$
\begin{aligned}
\frac{dp_{11}}{dt}(s,t) &= -p_{11}(s,t)\alpha_{12}(t) \qquad (3)\\
\frac{dp_{01}}{dt}(s,t) &= p_{00}(s,t)\alpha_{01}(t) - p_{01}(s,t)\alpha_{12}(t).
\end{aligned}
$$

The solution of these equations are:

$$p_{00}(s,t) = e^{-A_{01}(s,t)-A_{02}(s,t)}$$

$$p_{11}(s,t) = e^{-A_{12}(s,t)}$$

$$p_{01}(s,t) = \int_s^t p_{00}(s,u)\alpha_{01}(u)p_{11}(u,t)du,$$

where $A_{hj}(s,t) = \int_s^t \alpha_{hj}(u)du$. These equations have been given for general compensators in Andersen et al. (1993).

Inference can be based on maximising the likelihood. If a parametric model is chosen, modified Newton-Raphson algorithms (such as the Marquardt algorithm) can be used for the maximisation (the simplest parametric model is the homogeneous Markov model, followed by the piece-wise homogeneous Markov model). Non-parametric approaches can take two paths: one is the unconstrained non-parametric approach in the spirit of Turnbull (1976) and this was developped by Frydman (1995), another one uses smoothing, for instance through penalized likelihood such as in Joly and Commenges (1999). In the former path the EM algorithm is attractive, in the latter the Marquard algorithm achieves a good speed of convergence. All the above approaches are based on the likelihood which has been derived heuristically. In complex problems such as the one at hand, it is important to have a rigourous derivation of the likelihood; this is the purpose of the next section.

3. RIGOROUS DERIVATION OF LIKELIHOOD FOR ILLNESS-DEATH

3.1 GENERALITY ON LIKELIHOOD

Consider a measurable space $(\Omega, \mathcal{F})$ and a family of measures P^θ absolutely continuous relatively to a dominant measure P^0. The likelihood ratio is defined by:

$$\mathcal{L}_{\mathcal{F}}(\theta) = \frac{dP^\theta}{dP^0}_{|\mathcal{F}}$$

where $\frac{dP^\theta}{dP^0}_{|\mathcal{F}}$ is the Radon-Nikodym derivative of P^θ relatively to P^0. Recall that $\frac{dP^\theta}{dP^0}_{|\mathcal{F}}$ is the $\mathcal{F}$-measurable random variable such that

$$P^\theta(F) = \int_F \frac{dP^\theta}{dP^0} dP^0, F \in \mathcal{F}$$

For instance, the likelihood ratio corresponding to the observation of a random variable X (that is to the σ-algebra $X = \sigma(X)$) can be written

$$\mathcal{L}_{\mathcal{X}}(\theta) = \frac{f_X^\theta(X)}{f_X^0(X)},$$

where $f_X^\theta(.)$ is the density of the law of X relatively to a given measure: for instance, for a continuous variable, $f_X^\theta(.)$ is the probability density function. Since the denominator does not depend on θ, inference can be based only on $f_X^\theta(X)$, which is the form of the likelihood which appears in statistical papers. It is sometimes overlooked that the likelihood is a random variable, being a composition of the probability density function and the random variable X itself.

When dealing with complex problems such as inference based on incomplete observations of processes, such a simplification is not available and it is necessary to return to more fundamental theory. We are especially interested here in writing the likelihood for interval-censored observations from an illness-death model. We shall see that an illness-death model can be described as a bivariate counting process. We could find the likelihood for interval-censored observation of a unidimensional counting process relatively easily, for instance by considering that we have interval-censored observation of a random variable which represents the time of jump. However for a multivariate process this becomes much more difficult.

Consider the case of multivariate (or marked) point processes: $N = (N_h, h = 1, 2, \ldots)$. Denote $N_. = \sum N_h$ and $\Lambda_. = \sum \Lambda_h$, where Λ_h are the compensators of N_h (that is $N_h - \Lambda_h$ are martingales and Λ_h are increasing predictable processes); when the compensators are continuous we define intensities λ_h by $\Lambda_h = \int \lambda_h$. Consider also two probability measures $\tilde{P}$ and P with $\tilde{P} << P$. Jacod (1975) has given the formula for the likelihood ratio of the process N; this formula is presented in

Andersen et al. (1993) in term of product-integral, and supposing there is no information at time 0 it takes the form:

$$\frac{d\tilde{P}}{dP} = \prod_{t \le C} \prod_{h} \left(\frac{d\tilde{\Lambda}_h}{d\Lambda_h}(t) \right)^{\Delta N_h(t)} \frac{\prod_{t \le C: \Delta N_.(t) \neq 1} (1 - d\tilde{\Lambda}_.(t))}{\prod_{t \le C: \Delta N_.(t) \neq 1} (1 - d\Lambda_.(t))}$$

This is the likelihood ratio for the sigma-algebra $\mathcal{N} = \sigma(N(t), t \geq 0)$ with compensators relative to the filtration $\mathcal{N}_t = \sigma(N(u), u \geq 0, u \leq t)$; thus we cannot directly apply the formula because we do not observe $\mathcal{N}$ but $\mathcal{O} \subset \mathcal{N}$.

There are two strategies for applying this formula to our incomplete observation problem:

- Take the conditional expectation: $E[\frac{d\tilde{P}}{dP}|\mathcal{O}]$
- Apply the formula not on N but on an observed process

As an example of the latter consider the one-dimensional (so $h = 1$) process $N^{\mathcal{O}}(t) = N(l(t))$, where $l(t) = \sup(u \leq t : R(u) = 1)$. By definition this process is observed: $\mathcal{O} = \sigma(N^{\mathcal{O}}(t), t > 0)$, so that we can apply Jacod's formula. Consider the case of purely interval-censored data: $R(t) = 1$ for $t = V_0, V_1, \dots, V_m$, $R(t) = 0$ otherwise. Then $N^{\mathcal{O}}$ has a discrete compensator with jumps at $V_0, V_1, \dots, V_m$

$$\Delta\Lambda^{\mathcal{O}}(V_j) = P[N^{\mathcal{O}}(V_j) = 1 | N^{\mathcal{O}}(V_{j-1}) = 0] I_{\{N^{\mathcal{O}}(V_{j-1})=0\}}$$

It is easy to see that by applying Jacod's formula we get the expected result for the likelihood (expressed in term of the survival function S of the jump time):

$$\mathcal{L} = d\tilde{P} = \tilde{S}(V_{J-1}) - \tilde{S}(V_J),$$

where the random variable J is defined as $N^{\mathcal{O}}(V_J) - N^{\mathcal{O}}(V_{J-1}) = 1$; in this formula we have dropped the denominator which does not depend on the parameters.

3.2 COUNTING PROCESS MODEL FOR ILLNESS-DEATH

Consider one counting process N_I for illness ($N_I(t) = 0$ if healthy at t, $N_I(t) = 1$ if subject became ill before t) with intensity λ_I and one for death N_D ($N_D(t) = 0$ if alive at t, $N_D(t) = 1$ if subject died before t) with intensity λ_D. Let us model the intensities (in the $\mathcal{N}_t$-filtration) as:

$$\lambda_I(t) = I_{\{N_I(t-)=0\}} I_{\{N_D(t-)=0\}} \alpha_{01}(t)$$

$$\lambda_D(t) = I_{\{N_D(t-)=0\}} [I_{\{N_I(t-)=0\}} \alpha_{02}(t) + I_{\{N_I(t-)=1\}} \alpha_{12}(t)] \tag{4}$$

If we define $X = N_I + N_D + N_D(1 - N_I)$, this defines a multi-state process taking values on $\{0, 1, 2\}$ and with transition intensities $\alpha_{01}(.)$, $\alpha_{02}(.)$ and $\alpha_{12}(.)$ between $(0, 1)$, $(0, 2)$ and $(1, 2)$ respectively; there is identity between this multi-state (illness-death) process and the bivariate counting process.

To N_D we associate a response process $R_D(t) = 1$, for all $t \leq C$; to N_I, we associate a response process $R_I(t) = 1$ for $t = V_0, \ldots, V_m$, $R_I(t) = 0$ otherwise. The observed process is $N^{\mathcal{O}} = (N_I^{\mathcal{O}}, N_D^{\mathcal{O}})$, with

$$N_I^{\mathcal{O}}(t) = N_I(l(t))$$

where $l(t) = \sup\{u \leq t : R_I(u) = 1\}$, and

$$N_D^{\mathcal{O}}(t) = N_D(t), \text{ for } t \leq C.$$

Jacod's formula can be applied if we know the compensator of $N^{\mathcal{O}}$ in the $\mathcal{O}_t$ filtration: although we observe N_D its compensator is not the same on $\mathcal{N}_t$ and on $\mathcal{O}_t$. Thus, we need compute the compensators of $N_I^{\mathcal{O}}$ and $N_D^{\mathcal{O}}$ in the $\mathcal{O}_t$-filtration. It is easy to see that $N_I^{\mathcal{O}}$ has a discrete compensator which is null everywhere except possibly at observation times $V_j, j = 0, \ldots, m$ where it is equal to :

$$\Delta\Lambda_I^{\mathcal{O}}(V_j) = P[N_I^{\mathcal{O}}(V_j) = 1 | N_I^{\mathcal{O}}(V_{j-1}) = 0, N_D^{\mathcal{O}}(V_j-) = 0] I_{\{N_I^{\mathcal{O}}(V_{j-1})=0\}} I_{\{N_D^{\mathcal{O}}(V_j-)=0\}}$$

It can be seen that $N_I^{\mathcal{O}}$ and $N_D^{\mathcal{O}}$ can be replaced by N_I and N_D and, reminding that N_I and N_D are not independent, we can write:

$$P[N_I(V_j) = 1 | N_I(V_{j-1}) = 0, N_D(V_j-) = 0] = \frac{p_{01}(V_{j-1}, V_j)}{p_{0.}(V_{j-1}, V_j)},$$

where $p_{0.}(.,.) = p_{00}(.,.) + p_{01}(.,.)$ (the probability of being still alive); of course the transition probabilities $p_{hj}(s,t)$ still have a meaning in terms of the bivariate counting process, for instance $p_{00}(s,t) = P[N_I(t = 0, N_D(t) = 0 | N_I(s) = 0, N_D(s) = 0)$.

As for N_D, it is observed in continuous time so we have $N_D^{\mathcal{O}}(t) = N_D(t)$, for $t \le C$. However its compensator is not the same in the $\mathcal{N}_t$-filtration and in the $\mathcal{O}_t$-filtration: it is clear that the intensity given in formula (4) is not $\mathcal{O}_{t-}$-measurable. We may use the innovation theorem and compute the $\mathcal{O}_t$-intensity as:

$$\lambda_D^{\mathcal{O}}(t) = \mathrm{E}[\lambda_D(t)|\mathcal{O}_{t-}] = \mathrm{E}[I_{\{N_D(t-)=0\}}[I_{\{N_I(t-)=0\}}\alpha_{02}(t)+I_{\{N_I(t-)=1\}}\alpha_{12}(t)]|\mathcal{O}_{t-}].$$

In this formula, only $I_{\{N_I(t-)=0\}}$ is not $\mathcal{O}_{t-}$-measurable so the only problem is to compute

$$\mathrm{E}[I_{\{N_I(t-)=0\}}|\mathcal{O}_{t-}] = P[N_I(t-) = 0|\mathcal{O}_{t-}].$$

If $N_D(t-) = 1$ we can take any arbitrary value for this probability; if $N_I(l(t-)) = 1$, this probability is null. The only non-trivial quantity is

$$P[N_I(t-) = 0|N_D(t-) = 0, N_I(l(t-)) = 0] = \frac{p_{00}(l(t-), t-)}{p_{0.}(l(t-), t-)}.$$

Finally, the $\mathcal{O}_t$-intensity of N_D is

$$\lambda_D^{\mathcal{O}}(t) = I_{\{N_D(t-)=0\}}[I_{\{N_I(l(t-))=0\}}\bar{\alpha}_D(t) + I_{\{N_I(l(t-))=1\}}\alpha_{12}(t)],$$

where $\bar{\alpha}_D(t) = \frac{p_{00}(l(t-),t-)\alpha_{02}(t)+p_{01}(l(t-),t-)\alpha_{12}(t)}{p_{0.}(l(t-),t-)}$. This formula has a natural interpretation, the intensity being a weighing of the transition intensities from health and illness with the required probabilities conditional on what has been observed just before t; if the subject has been observed in the illness state, then the intensity is α_{12} (for an alive subject).

The likelihood ratio in Jacod's formula can be written as the product of three terms $\mathcal{L} = \mathcal{L}_I\mathcal{L}_D\mathcal{L}_.$. The first term is the contribution of observing a jump of N_I: it is equal to 1 if no jump has been observed and if a jump has been observed at V_J:

$$\mathcal{L}_I = \frac{\Delta\tilde{\Lambda}_I^{\mathcal{O}}(V_J)}{\Delta\Lambda_I^{\mathcal{O}}(V_J)} = \frac{\tilde{p}_{01}(V_{J-1}, V_J)p_{0.}(V_{J-1}, V_J)}{\tilde{p}_{0.}(V_{J-1}, V_J)p_{01}(V_{J-1}, V_J)}.$$

From now on we drop the denominator and the tilde and we will simply write: $\mathcal{L}_I = \frac{p_{01}(V_{J-1},V_J)}{p_{0.}(V_{J-1},V_J)}$

The second term is the contribution of observing a jump of N_D: it is equal to 1 if no jump has been observed; if a jump (that is death) has been observed at T,it is equal to $\lambda_D^{\mathcal{O}}(T)$. If the subject has been seen ill at V_J the contribution is $\mathcal{L}_D = \alpha_{12}(T)$; if not it is

$$\mathcal{L}_D = \bar{\alpha}_D(T) = \frac{p_{00}(l(T-), T-)\alpha_{02}(T) + p_{01}(l(T-), T-)\alpha_{12}(T)}{p_{0.}(l(T-), T-)}.$$

The last term of the formula, the product integral over times where no jump happened, is the product of a dicrete and a continuous part: $\mathcal{L}_{.}\mathcal{L}_{.I}\mathcal{L}_{.D}$. The discrete part $\mathcal{L}_{.I}$ comes from the discrete compensator $\Lambda_I^{\mathcal{O}}$ and if a subject has been seen ill for the first time at V_J is a simple product:

$$\mathcal{L}_{.I} = \prod_{j=1}^{J-1}(1 - \Delta\Lambda_I^{\mathcal{O}}(V_j)) = \frac{p_{00}(V_0, V_{J-1})}{p_{0.}(V_0, V_{J-1})};$$

the product stops at V_{J-1} because there is a jump at V_J and the compensator is constant after V_J; if the subject is never seen ill, the product goes until the last visit time. Finally the continuous part of the product integral is

$$\mathcal{L}_{.D} = \prod_{t \le \tilde{T}}(1 - d\Lambda_D^{\mathcal{O}}(t)) = e^{-\int_{V_0}^{\tilde{T}} \lambda_D^{\mathcal{O}}(t)dt}.$$

On $V_{j-1} < t < V_j$, where $N_I(V_{j-1}) = 0$ and $N_D(t-) = 0$ we have using the Kolmogorov equations (3)

$$\lambda_D^{\mathcal{O}}(t) = \bar{\alpha}_D(t) = -\frac{d \log p_{0.}(V_{j-1}, t)}{dt}.$$

Thus for a subject who has not been seen ill we have:

$$\mathcal{L}_{.D} = e^{-\int_{V_0}^{\tilde{T}} \bar{\alpha}_D(t)dt} = p_{0.}(V_0, \tilde{T}),$$

and for a subject seen ill at V_J:

$$\mathcal{L}_{.D} = e^{-\int_{V_0}^{V_J} \bar{\alpha}_D(t)dt - \int_{V_J}^{\tilde{T}} \alpha_{12}(t)dt} = p_{0.}(V_0, V_J)p_{11}(V_J, \tilde{T}).$$

Finally for a subject not seen ill, calling $V_L = l(\tilde{T})$ the last visit time, we have

$$\mathcal{L}_{.I}\mathcal{L}_{.D} = p_{00}(V_0, V_L)p_{0.}(V_L, \tilde{T}).$$

Thus the likelihood is:

$$\mathcal{L} = p_{00}(V_0, V_L)p_{0.}(V_L, \tilde{T})\bar{\alpha}_D(\tilde{T})^{\delta},$$

where $\alpha_D(\tilde{T}) = \frac{p_{00}(V_L,T)\alpha_{02}(\tilde{T})+p_{01}(V_L,\tilde{T})\alpha_{12}}{p_{0.}(V_L,\tilde{T})}$, which is identical to (1).

For a subject seen ill at V_J, writing the likelihood as $\mathcal{L} = \mathcal{L}_{.I}\mathcal{L}_I\mathcal{L}_{.D}\mathcal{L}_D$ we have:

$$\mathcal{L} = \frac{p_{00}(V_0, V_{J-1})}{p_{0.}(V_0, V_{J-1})}\frac{p_{01}(V_{J-1}, V_J)}{p_{0.}(V_{J-1}, V_J)}p_{0.}(V_0, V_J)p_{11}(V_J, \tilde{T})\alpha_{12}(\tilde{T})^{\delta},$$

which is identical to (2).

Thus we have proved that the heuristic way of deriving the likelihood gives the correct result for the illness-death model with the mixed discrete-continuous time observation pattern.

5. REFERENCES

Andersen, PK, Borgan Ø, Gill RD and Keiding N. (1993) *Statistical Models Based on Counting Processes.* New-York: Springer-Verlag.

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

Frydman, H. (1995). "Non-parametric estimation of a Markov "illness-death model" process from interval-censored observations, with application to diabetes survival data". *Biometrika* 82, 773-789.

Jacod, J. (1975). "Multivariate point processes: predictable projection; Radon-Nikodym derivative, representation of martingales. *Z. Wahrsheinlichkeitsth.* 31, 235-253.

Joly, P. and Commenges, D. (1999). "A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to AIDS". *Biometrics* 55, 887-890.

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

Turnbull, B. W. (1976). "The empirical distribution function with arbitrarily grouped, censored and truncated data". *Journal of the Royal Statistical Society, Series B* 38, 290-5.