Likelihood for Generally Coarsened Observations from Multistate or Counting Process Models

DANIEL COMMENGES and ANNE GÉGOUT-PETIT
INSERM, Université Victor Segalen Bordeaux 2

ABSTRACT. We consider first the mixed discrete-continuous scheme of observation in multistate models; this is a classical pattern in epidemiology because very often clinical status is assessed at discrete visit times while times of death or other events are observed exactly. A heuristic likelihood can be written for such models, at least for Markov models; however, a formal proof is not easy and has not been given yet. We present a general class of possibly non-Markov multistate models which can be represented naturally as multivariate counting processes. We give a rigorous derivation of the likelihood based on applying Jacod’s formula for the full likelihood and taking conditional expectation for the observed likelihood. A local description of the likelihood allows us to extend the result to a more general coarsening observation scheme proposed by Commenges & Gégout-Petit. The approach is illustrated by considering models for dementia, institutionalization and death.

Key words: coarsening, counting processes, dementia, interval-censoring, likelihood, Markov models, multistate models

1. Introduction

Multistate models have been proposed for a long time, in particular in biological applications. Until the 1970s, however, the attention was essentially focused on homogeneous Markov models. Non-homogeneous Markov models were studied by Fleming (1978), Aalen & Johansen (1978); semi-Markov models were also considered by Lagakos et al. (1978); the most studied model was the illness–death model (Andersen, 1988; Keiding, 1991). A thorough account of these developments can be found in Andersen et al. (1993) where the counting process theory is used to obtain rigorous results in this field; see also Hougaard (2000).

Another stream of research was started by Peto (1973) and Turnbull (1976) who tackled the problem of interval-censored observations in survival data analysis and gave the non-parametric maximum likelihood estimator; Frydman (1995a, b) extended this issue to the illness–death model in which transition towards illness could be interval-censored while time of death was exactly observed, while Wong & Yu (1999) studied a general case of multivariate interval-censored data. The penalized likelihood approach to this problem was proposed by Joly & Commenges (1999) with an application to acquired immunodeficiency syndrome (AIDS), and by Joly et al. (2002) and Commenges et al. (2004) with application to Alzheimer’s disease.

In the continuous time observation scheme it has been shown (Aalen, 1978; Borgan, 1984; Andersen et al., 1993) that the likelihood could be derived from Jacod’s formula (Jacod, 1975); also in this case the martingale theory yields simple and natural estimators (Andersen et al., 1993; Aalen et al., 2004). In the case where some transitions are observed in discrete time while others are observed in continuous time that we call the mixed discrete-continuous observation (MDCO) scheme, the natural martingale estimators are no longer available, so one has to return to likelihood-based methods. The above cited papers considering this case used heuristic likelihoods. Commenges (2003) derived the likelihood for this observation
scheme in a Markov illness–death model. The main aim of this paper was to rigorously derive the likelihood in a more general framework: (i) we consider a general class of multistate models which may have any number of states and are not necessarily Markov; (ii) we first consider the MDCO scheme and we extend to the so-called general coarsening model for processes (GCMP) proposed by Commenges & Gégout-Petit (2005). In most of this work we assume that the mechanism leading to incomplete data is ignorable (Gill et al., 1997); however, we prove ignorability in the special case when death produces a stochastic censoring of the other processes of the model. Our approach starts by remarking that most of the useful multistate models can be directly formulated in terms of multivariate counting process in a very natural way, an idea close to the ‘composable processes’ studied by Schweder (1970). Then Jacod’s formula can be applied to find the likelihood for continuous time observation. In the case where one or several of these processes is observed in discrete time the likelihood can be computed by taking a conditional expectation of the full likelihood.

In section 2, the heuristic likelihoods for diverse observation schemes are recalled. Section 3 develops a natural correspondence between the class of irreversible multistate (IM) processes and multivariate one-jump counting (OJC) processes. In section 4, the multivariate counting process representation is exploited to use Jacod’s formula for finding the likelihood in the MDCO scheme; then the result is extended to the GCMP scheme. This general modelling approach is illustrated in section 5 for describing a joint model for dementia, institutionalization and death, presented as a five-state model in Commenges & Joly (2004) and showing the benefits of the proposed approach to this case. Section 6 briefly concludes.

2. Heuristic likelihoods for multistate models

2.1. Notation

A multistate process \( X = (X_t) \) is a right-continuous process which can take a finite number of values \( \{0, 1, \ldots, K - 1\} \). The theory of Markov multistate models (or Markov chain models) is well established. The law of a Markov multistate process is defined by the transition probabilities between states \( h \) and \( j \) that we will denote by \( p_{hj}(s, t) = \Pr(X_t = j | X_s = h) \); transition intensities \( \lambda_{hj}(t) \), for \( h \neq j \), may be defined when the \( p_{hj}(s, t) \)s are continuous both in \( s \) and in \( t \) for all \( s \) and \( t \), as the following limits (if they exist):

\[
\lim_{\Delta t \to 0} \frac{p_{hj}(t, t + \Delta t)}{\Delta t}.
\]

It is reasonable in most applications in epidemiology to think that these limits exist; it is even reasonable to expect continuous and smooth transition intensities. We define \( \lambda_{hj}(t) = -\sum_{j \neq h} \lambda_{hj}(t) + \sum_{j \neq h} \lambda_{hj}(t) \) is the hazard function associated with the distribution of the sojourn time in state \( h \). When the \( \lambda_{hj}(t) \)s do not depend on \( t \), the Markov chain is said to be homogeneous. Transition probabilities and transition intensities are linked by the forward Kolmogorov differential equations, so that solving these equations transition probabilities can be expressed as a function of transition intensities; this solution can take the form of the product integral (see Andersen et al., 1993). For non-Markov multistate models one could define analogously transition probabilities \( p_{hj}(s, t) = \Pr(X_t = j | X_s = h, F_{s-}) \), where \( F_{s-} \) is the history before \( s \); similarly, transition intensities \( \lambda_{hj}(t; F_{s-}) \) could be defined. However, we are on a less firm ground because, contrarily to the Markov case, these quantities are random and the Kolmogorov equations have been given only for Markov processes.

In the remaining of this paper we shall consider observations for different schemes of observation, leading to incomplete data. We assume that the mechanism leading to incomplete data is ignorable: this means that we make a correct inference by using the likelihood
as if these schemes were deterministic. In particular, we can consider that the different times like $C$ and $v_l, l=0,\ldots,m$ involved in this mechanism, and which are defined below, are fixed. However, this raises a problem which will be discussed and solved in section 4.5.

### 2.2. Likelihood for continuous time observations in Markov models

Consider the case where process $X$ is continuously observed from $v_0$ to $C$. We observe that transitions have occurred at (exactly) times $T^{(1)}<T^{(2)}<\cdots<T^{(M)}$. With the convention $T^{(0)}=v_0$, the value of the likelihood, conditional on $X_{v_0}$, on the event $\{M=m\}$ and $\{X_{T^{(r)}}=x_r, r=0,\ldots,m\}$ is:

$$L = \prod_{r=1}^{m} p_{x_{r-1}, x_r}(T^{(r-1)}, T^{(r)}) \cdot x_{x_{r-1}, x_r}(T^{(r)}) \cdot p_{x_m, x_0}(T^{(m)}, C).$$

The probability that no transition happens between $T^{(r-1)}$ and $T^{(r)}$ given $X_{T^{(r-1)}}=x_{r-1}$ can easily be computed as

$$p_{x_{r-1}, x_r}(T^{(r-1)}, T^{(r)}) = \exp \int_{T^{(r-1)}}^{T^{(r)}} x_{x_{r-1}, x_r}(u) \, du.$$

### 2.3. Likelihood for discrete time observations in Markov models

Consider now the case where $X$ is observed at discrete times $v_0, v_1,\ldots, v_m$. In this case, we observe a vector of random variables $(X_{v_0},\ldots, X_{v_m})$ and it is easy to derive the value of the likelihood, conditional on $X_{v_0}$, on the event $\{X_{v_r}=x_r, r=0,\ldots,m\}$:

$$L = \prod_{r=0}^{m-1} p_{x_r, x_{r+1}}(v_r, v_{r+1}).$$

### 2.4. Likelihood for mixed discrete-continuous time observations in Markov models

The most common case in applications is that some transitions are observed in discrete times and others in continuous time. A classical example is the irreversible illness-death model, a model with the three states ‘health’, ‘illness’, ‘death’, respectively, labelled 0, 1, 2; it is often the case that transition towards the illness state is observed in discrete time while transitions towards death is observed in continuous time (Frydman, 1995a; Joly et al., 2002). 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 $\delta = 1_{\{T \leq C\}}$. If the subject starts in state ‘health’, has never been observed in the ‘illness’ state and was last seen at visit $M$ (at time $v_M$) the likelihood, conditional on $X_{v_0}=0$, on the event $\{M=m\}$ is:

$$L = p_{00}(v_0, v_m)[p_{00}(v_m, \tilde{T})x_{02}(\tilde{T})^\delta + p_{01}(v_m, \tilde{T})x_{12}(\tilde{T})^\delta];$$

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

$$L = p_{00}(v_0, v_{l-1})p_{01}(v_{l-1}, v_l)p_{11}(v_l, \tilde{T})x_{12}(\tilde{T})^\delta.$$

These formulae can be extended rather easily to models with more than three states in the case where there is one absorbing state, and transitions towards the absorbing state are observed in continuous time while transitions towards other states are observed in discrete
time (Commenges, 2002). In the case where transitions towards one state are observed in discrete time while transitions towards other states are observed in continuous time the formula becomes cumbersome: see the example of the dementia–institution–death model of Commenges & Joly (2004). These formulae are heuristic. The same type of formulae can be written for non-Markov model using random versions of the transition probabilities and intensities and hoping to be able to compute transition probabilities in term of transition intensities; see Joly & Commenges (1999) for an example of a semi-Markov model.

3. Multistate models as multivariate counting processes

The main motivation for representing multistate models as multivariate counting processes is the availability of a formula giving the likelihood ratio for such processes (not necessarily Markov) observed in continuous time on \([0, C]\) (Jacod, 1975). It will be shown in the next section that this can be the basis of a rigorous derivation of the likelihood also in the MDCO and in the more general GCMP schemes. It was shown (Borgan, 1984; Andersen et al., 1993) that a multistate model \((X, \alpha)\), where \(\alpha = (\alpha_{jh}); 0 \leq h, j \leq K - 1\), can be represented by the multivariate counting process \((N, \lambda)\) where \(N = (N_{hj}, h \neq j, 0 \leq h, j \leq K - 1)\) and the \(N_{hj}\) count the number of transitions from state \(h\) to state \(j\); the intensity of \(N\) is \(\lambda = (\lambda_{jh}, h \neq j, 0 \leq h, j \leq K - 1)\) and \(\lambda_{jh} = Y_h(t)\alpha_{jh}(t)\), where \(Y_h(t) = 1_{\{X_\tau = h\}}\).

For IM models a more parsimonious representation is possible. It is often possible to formulate an epidemiological problem directly in term of counting processes rather than using the counting process representation as a mathematical device. Most multistate models are in fact used for jointly modelling several events. For instance, the illness–death model is used to jointly model onset of disease and death. So we can directly model the problem by considering a bivariate counting process \(N = (N_1, N_2)\), where \(N_1\) counts the onset of the disease and \(N_2\) counts the occurrence of death. It can be seen that the multistate process can be retrieved by \(X_t = \min(2N_2 + N_{1t}, 2)\). The processes \(N\) and \(X\) generate the same filtration. However, note that for a fixed \(t\), the random variables \(X_t\) and \(N_t\) do not generate the same \(\sigma\)-field in general because the event \(\{X_t = 2\}\) (subject in the ‘death’ state at \(t\)) is the same as \(\{N_2 = 1\}\); however, if we know \(N_t\), we also know \(N_{1t}\), that is, we know whether the subject passed through the state ‘illness’ before \(t\) or not. Note that the representation based on the basic events of interest is more economical than the Borgen representation (a bivariate rather than three-variate process). Similarly the dementia–institution–death model, which is a five-state model (see section 5), can be represented by a three-variate process counting onset of dementia, institutionalization and death; in this model there are eight possible transitions so that the Borgen representation would entail an eight-variate process.

Let us consider the more general problem of jointly modelling the onset of \(p\) types of events, each type occurring just once. This can be represented by a \(p\)-variate counting process \(N\), each \(N_j\) making at most one jump. It is possible to construct a multistate process \(W\) such that \(W_t\) generates the same \(\sigma\)-field as \(N_t\). A possibility is: \(W_t = N_p2^{p-1} + N_{p-1,1}2^{p-2} + \cdots N_{1t}\), that we denote by \(W_t = N_pN_{p-1,1}\cdots N_{1t}\) (this is the representation of \(W_t\) in base 2); \(W_t\) can take \(2^p\) integer values in the set \(\{0, 1, \ldots, 2^p - 1\}\). Consider the important case where \(N_p\) counts death; it is common in the multistate representation to consider that deceased subjects are in the same state, that is, we may construct a more compact multistate model \(X\) defined by \(X_t = \min(W_t, 2^{p-1})\); \(X_t\) can take \(2^{p-1} + 1\) values. For example, the dementia–institution–death model (where \(p = 3\)) has five states rather than eight states. We have exactly the same number of non-zero transition intensities for \(W\) and \(X\) and they are equal; we simply have to rename them. More specifically we have \(\alpha_{jh}^W(\cdot) = \alpha_{jh}(\cdot)\) for \(0 \leq h < j \leq 2^{p-1}\) and \(\alpha_{jh}^W(\cdot) = \alpha_{h2^{p-1}}(\cdot)\) for \(0 \leq h < 2^{p-1} \leq j\).
Theorem 1
Let \( N = (N_1, \ldots, N_p) \) be a counting process with \( N_j \leq 1; \ j = 1, \ldots, p; \ t \geq 0 \) and \( p > 1 \). Consider the multistate process \( W = (W_t) \) defined by \( W_t = N_1, \ldots, N_t \) in base 2. If, in a given probability measure, \( W \) is Markov with continuous transition intensities \( \lambda^{W}_j(\cdot); 0 \leq h, \ j \leq 2^p - 1, \) the generating counting process \( N_j, \ j = 1 \ldots, p, \) have intensities given by:

\[
\lambda_j(t) = 1_{\{N_{jt} = 0\}} \sum_{k_1 = 0}^{1} \cdots \sum_{k_p = 0}^{1} \prod_{l = 1}^{p} 1_{\{N_{lt} = k_l\}} \lambda^{W}_{k_p \ldots k_{j-1} 0 k_{j-1} \ldots k_1} (t); \ t \geq 0, \tag{4}
\]

where \( k_p \ldots k_{j-1} 0 k_{j-1} \ldots k_1 \) and \( k_p \ldots k_{j-1} 1 k_{j-1} \ldots k_1 \) are base 2 representations of integers.

Proof. Lemma 3.3 of Aalen (1978) gives an expression of the càdlàg modification \((\lambda_j(t+))\) of the caglåd process \((\lambda_j(t))\) by:

\[
\lambda_j(t+) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P \left[ N_{j(t+\delta)} - N_j - 1 \mid \mathcal{F}_t \right] = \lambda_j(t).
\]

This is easy to apply this theorem for finding the intensities of the \( N_j \)'s as a function of the transition intensities of \( W \) defined as above by first finding the intensities in terms of the transition intensities of \( W \) and then renaming the transition intensities.

Example. The illness–death model. Consider a bivariate counting process \( N = (N_1, N_2) \), where \( N_1 \) counts the onset of the disease and \( N_2 \) counts the occurrence of death; thus we have \( p = 2 \). We can form the process \((W_t)\) defined by \( W_t = N_2, N_{1t} \). This process has four states but states 2 and 3 have the same biological meaning (‘dead’) so that they are generally grouped to obtain the conventional illness–death model \((X_t)\) defined by \( X_t = \min(W_t, 2) \). If \( W \) is Markov, applying formula (4) yields for the intensities of \( N_1 \) and \( N_2 \) respectively:

\[
\lambda_1(t) = 1_{\{N_{1t} = 0\}} 1_{\{N_{2t} = 0\}} \lambda^{W}_{01}(t)
\]
\[
\lambda_2(t) = 1_{\{N_{2t} = 0\}} \left[ 1_{\{N_{1t} = 0\}} \lambda^{W}_{02}(t) + 1_{\{N_{1t} = 1\}} \lambda^{W}_{12}(t) \right].
\]

In terms of the transition intensities of \( X \) we obtain:

\[
\lambda_1(t) = 1_{\{N_{1t} = 0\}} 1_{\{N_{2t} = 0\}} \lambda_{01}(t)
\]
\[
\lambda_2(t) = 1_{\{N_{2t} = 0\}} \left[ 1_{\{N_{1t} = 0\}} \lambda_{02}(t) + 1_{\{N_{1t} = 1\}} \lambda_{12}(t) \right].
\]
The same approach can be applied if some counting processes take more than two values even if the trick of the representation of the state value in base 2 does not work. A process with more than one jump can represent progression of a disease: for instance, the progressive model of Joly & Commenges (1999) can be represented by a counting process making a jump at human immunodeficiency virus (HIV) infection and another jump at onset of AIDS. Alternatively we can represent progression of the disease by two 0–1 counting processes, one counting HIV infection, the other counting onset of AIDS, with the intensity of the latter being equal to zero if the former has not jumped. The 10-state model proposed by Alioum et al. (2005) can be represented by a process representing progression of the HIV disease which can make two jumps and three 0–1-processes counting HIV diagnosis, inclusion in a cohort and death or alternatively by five 0–1-processes if we represent progression of the disease by the two 0–1 counting processes as discussed above. A problem appears if we make the base-2 construction for such 0–1 processes. The construction leads to ‘phantom’ states which are not relevant, for instance, the state AIDS without HIV infection; however, we can still represent the relevant multistate model by putting the transition leading to this state uniformly equal to zero. Theorem 1 can then still be applied.

We can formalize the relationship between the class of IM processes and OJC processes by saying that they are equivalent. This concept of equivalence between classes of processes is based on the following equivalence relation between processes.

**Definition 1**

Two processes are informationally equivalent if they generate the same filtration.

For each IM process we can find at least one informationally equivalent OJC process; this is obvious from the Borgan representation. Inversely for each OJC process we can find an informationally equivalent IM process (using the base 2 representation). We can also define canonical processes, that is, simple representants of an equivalence class; this can be based on a notion of minimal representation: within the IM class the canonical process is the process with the smallest number of states; within the OJC class, this is the process of lowest dimension. For instance, for the illness–death process, the canonical IM is the three-state process $X$ (rather than the four-state process $W$) and the canonical OJC process is the bivariate ‘basic’ counting process (rather than the three-dimensional process obtained form the Borgan representation).

In the next section, we will derive the likelihood for the IM–OJC class of processes from increasingly complex schemes of observation and show in some examples that the rigorously derived likelihoods are the same as the heuristic ones.

### 4. Derivation of the likelihood from Jacod’s formula

#### 4.1. Likelihood for counting processes observed in continuous time

From now on we adopt a somewhat more rigorous probabilistic formalism. We assume a probability space $(\Omega, \mathcal{F}, P)$ is given and we consider a counting process $N$ defined on this space. Jacod (1975) has given the likelihood ratio for observation of a counting process $N$ on $[0, C]$, that is, relative to the $\sigma$-field $\mathcal{F}_C = \mathcal{F}_0 \vee \mathcal{N}_C$ where $\mathcal{N}_C = \sigma(N_j, 0 \leq u \leq C; j = 1, \ldots, p)$. Aalen (1978), based on results of Jacod & Memin (1976), gave a simple form of the likelihood ratio in the case of absolutely continuous compensators by taking a reference probability under which the $N_j$s are independent Poisson processes with intensity 1. Here, we
consider a multivariate counting process $N$ with components $N_j$ which are 0–1 counting processes. For such processes it is more attractive to take a reference probability $P_0$ under which the $N_j$s are independent with intensities $\lambda_j^0(t)=1_{(N_j=0)}$; equivalently the $T_j$s are independent with exponential distributions with unit parameter. The likelihood ratio for a probability $P_0$ (with $P_0<<P_0$) relative to $P_0$ is:

$$
\mathcal{L}^0_{\mathcal{F}_C} = \mathcal{L}^0_{\mathcal{F}_0} \left[ \prod_{j=1}^{N_C} \lambda_j^0(T_j) \right] \exp[-\Lambda_j^0(C)] \prod_{j=1}^p e^{T_j^0},
$$

(7)

where for each $r \in \{1, \ldots, N_C\}$, $J_r$ is the unique $j$ such that $\Delta N_{T(r)}=1$; $N_r=\sum_{j=1}^p N_{j_r}$, $\Lambda_j^0(t)=\sum_{j=1}^p \Lambda_j^0(t)$, $\Lambda_j^0(t)=\int_0^t \lambda_j^0(u) \, du$ and $\lambda_j^0(t)$ is the intensity of $N_j$ under $P_0$. This formula allows us to compute the likelihood for any multivariate model once we have written it as a multivariate counting process. Within the OJC class, we denote $T_j$ the jump time of $N_j$; the likelihood can then be written (in a more manageable form for applications) in terms of $\bar{T}_j = \min(T_j, C)$ and $\delta_j = 1_{[T_j \leq C]}$ as:

$$
\mathcal{L}^0_{\bar{T}_C} = \mathcal{L}^0_{\mathcal{F}_0} \left[ \prod_{j=1}^p \tilde{\lambda}_j^0(\bar{T}_j) \right] \exp[-\tilde{\Lambda}_j^0(C)] \prod_{j=1}^p e^{\tilde{T}_j^0}.
$$

(8)

The term $\mathcal{L}^0_{\mathcal{F}_0}$ appears when there is some information about the process of interest in the initial $\sigma$-field at time 0; from now on we suppose that it is not the case and so $\mathcal{L}^0_{\mathcal{F}_0} = 1$ and this term disappears. We may have to compute conditional likelihoods, in particular likelihoods conditional on $\mathcal{F}_0$, as in section 2. The conditional likelihood is simply $\mathcal{L}^0_{\mathcal{F}_C | \mathcal{F}_0} = \mathcal{L}^0_{\mathcal{C}} / \mathcal{L}^0_{\mathcal{F}_0}$.

In particular, the likelihood $\mathcal{L}^0_{\mathcal{F}_C}$ on $\{N_0=0\}$ is equal to $\exp[-\Lambda^0(v_0) + p v_0]$ so that

$$
\mathcal{L}^0_{\mathcal{F}_C | N_0=0} = \left[ \prod_{j=1}^p \tilde{\lambda}_j^0(\bar{T}_j) \right] \exp[-(\Lambda_0^0(C)-\Lambda^0(v_0))] \prod_{j=1}^p e^{\tilde{T}_j^0 - v_0}
$$

(in this formula the $\bar{T}_j$ are the jump times possibly observed after $v_0$).

It is interesting to make the link with the heuristic likelihood expressed in terms of transition probabilities and intensities. Dropping the multiplicative factor $\prod_{j=1}^p e^{\tilde{T}_j^0 - v_0}$ (which does not depend on $\theta$) and rearranging the product, the likelihood (7) can be written:

$$
\mathcal{L}^0_{\mathcal{C}} = \prod_{r=1}^{N_C} \exp[-\Lambda_r(T_r) - \Lambda_r(T_{r-1})]\tilde{\lambda}_r(T_r) \exp[-\Lambda_r(C) - \Lambda_r(T_{N_r,C})],
$$

still with the convention $T_{(0)}=v_0$. We note that the number of jumps called $M$ in section 2.2 is precisely $N_C$; on $\{X_{T(r)}=x_r, r=1, \ldots, m\}$ we have $\tilde{\tilde{\lambda}}_r(T_r) = x_{r-1} x_r$ and

$$
\exp[-\Lambda_r(T_r) - \Lambda_r(T_{r-1})] = \exp\left[ - \int_{T_{r-1}}^{T_r} \tilde{\tilde{\lambda}}_r(u) \, du \right] = \exp\left[ - \int_{T_{r-1}}^{T_r} \sum_{h \neq x_{r-1}} x_{x_{r-1}h}(u) \, du \right].
$$

The latter term is equal to $p_{x_{r-1}x_{r-1}}(T_{r-1}, T_r)$ so that we retrieve the expression given in 2.2.

It must be noted that in general the $\tilde{\tilde{\lambda}}_r(t)$s and $\Lambda_r^0(t)$s may depend on what has been observed before $t$; to be more explicit we can write $\tilde{\tilde{\lambda}}_r(t; \bar{T}_l \wedge t, l=1, \ldots, p)$ and $\Lambda_j^0(C; \bar{T}_l, l=1, \ldots, p)$.  

The likelihood can be written as:

$$\mathcal{L}_C^\theta = f^\theta_C(\tilde{T}_1, \ldots, \tilde{T}_p) \prod_{j=1}^p e^{\tilde{T}_j},$$

(9)

where

$$f^\theta_C(s_1, \ldots, s_p) = \prod_{j=1}^p [\ell^\theta_j(s_j; s_j, l = 1, \ldots, p)]^{\mathbb{1}_{\{s_j < \cdot\}}} \exp[-\Lambda^\theta(C; s_j \wedge C, l = 1, \ldots, p)],$$

is the part of the likelihood which depends on $\theta$. Note that in this expression of the likelihood we have got rid of the $\delta_j$, thus simplifying the notation for the developments of the next sections; in this expression it is considered that if $\tilde{T}_j = C$ the observation is right censored; the case $\{ T_j = C \}$ does not make problem because this event has probability zero with our assumptions and the likelihood is defined almost everywhere.

### 4.2. Likelihood for the MDCO scheme: the case when only one component is partially observed

In this section, we shall compute the likelihood for the following scheme of observation of $N: N_1$ is observed at discrete times $v_0, \ldots, v_m$ while $N_j$, $j = 2, \ldots, p$ are observed in continuous time on $[0, C]$. The observation in this scheme is represented by the $\sigma$-field $O = \sigma(N_{tv}, l = 0, \ldots, m) \vee G_C$, where $G_C = \sigma(N_{mu}, j = 2, \ldots, p; 0 \leq u \leq C)$. We obviously have $O \subset \mathcal{F}_C$ so that the observed likelihood can be expressed as: $\mathcal{L}_O^\theta = E_0[\mathcal{L}_C^\theta \mid O]$ where $E_0$ means that we take expectation under $P_0$ defined in the previous subsection. We note from formula (9) that $\mathcal{L}_C^\theta = f^\theta_C(\tilde{T}, \Gamma)g(\Gamma)e^{\tilde{T}}$, where $\Gamma = (\tilde{T}_2, \ldots, \tilde{T}_p)$, $g(\Gamma) = \prod_{j=2}^p e^{\tilde{T}_j}$ and $f^\theta_C(\tilde{T}, \Gamma)$ is a shortcut for $f^\theta_C(\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_p)$; note that $\Gamma$ is a $G_C$-measurable random variable. From independence between the $T_j$s under $P_0$, we have independence between $\sigma(N_{tv}, l = 1, \ldots, m)$ and $G_C$ and the computation of the conditional expectation can be performed using the disintegration theorem (Kallenberg, 2001) which yields:

$$\mathcal{L}_O^\theta = g(\Gamma) \int f^\theta(\sigma, \Gamma)e^{\sigma}v(\sigma)\,ds,$$

where $v(\cdot)$ is a regular version of the law of $T_1$ given $O$ which, by independence equals a version of the law of $T_1$ given $\sigma(N_{tv}, l = 0, \ldots, m)$. We decompose $\mathcal{L}_O^\theta$ on atoms (see a definition in section 4.4) of $\sigma(N_{tv}, l = 0, \ldots, m)$ as:

$$\mathcal{L}_O^\theta = \sum_{l=1}^m 1_{\{v_l-1 < T_1 \leq v_l\}} E_0[\mathcal{L}_C^\theta \mid O] + 1_{\{T_1 > v_m\}} E_0[\mathcal{L}_C^\theta \mid O],$$

(10)

with the convention $v_0 = 0$. Using the fact that

$$v(\sigma) = \frac{1_{\sigma < \cdot \wedge v_l}(\sigma)e^{-\sigma}}{(e^{-\sigma} - e^{-v_l})} \,ds$$

gives the law of $T_1$ given $\{T_1 \in (v_l-1, v_l)\}$, we obtain for the first terms:

$$1_{\{v_l-1 < T_1 \leq v_l\}} E_0[\mathcal{L}_C^\theta \mid O] = 1_{\{v_l-1 < T_1 \leq v_l\}} g(\Gamma)E_0[f^\theta_C(T_1, \Gamma) \mid T_1 \in [v_l-1, v_l]]$$

$$= 1_{\{v_l-1 < T_1 \leq v_l\}} \frac{g(\Gamma)}{e^{-v_l} - e^{-v_{l-1}}} \int_{v_l-1}^{v_l} f^\theta_C(s, \Gamma) \,ds.$$  

(11)
We consider now the last term of (10). \( v(ds) = 1_{(t_m, +\infty)}(s)e^{-\gamma s}e^{\omega s}ds \) is a regular version of the law of \( T_1 \) given \( (T_1 > t_m) \) and we can write

\[
1_{\{T_1 > t_m\}}E_0[\mathcal{L}_{\mathcal{F}_C}^0 | \Omega] = 1_{\{T_1 > t_m\}}g(\Gamma)E_0[f_0^\theta(T_1, \Gamma) | T_1 > t_m]
\]

\[
= 1_{\{t_m < T_1\}}g(\Gamma)e^{\omega m} \left[ \int_{t_m}^C f_0^\theta(s, \Gamma) ds + \int_{C}^{+\infty} f_0^\theta(C, \Gamma) e^{-\gamma s} ds \right]
\]

\[
= 1_{\{T_1 > t_m\}}g(\Gamma)e^{\omega m} \left[ \int_{t_m}^C f_0^\theta(s, \Gamma) ds + f_0^\theta(C, \Gamma) \right].
\]

Combining equations (10), (11) and (12) we have proved the following.

**Lemma 1**

For \( N_1 \) observed at discrete times \( v_0, \ldots, v_m \) and \( N_j \) (2 \( \leq j \leq p \)), observed in continuous time the likelihood is given by

\[
\mathcal{L}_{\mathcal{F}_C}^0 = \sum_{j=1}^m 1_{\{v_{j-1} < T_1 \leq v_j\}} \frac{g(\Gamma)}{e^{-v_{j-1}} - e^{-v_j}} \int_{v_{j-1}}^{v_j} f_0^\theta(s, \Gamma) ds + 1_{\{T_1 > t_m\}}g(\Gamma)e^{\omega m} \left[ \int_{t_m}^C f_0^\theta(s, \Gamma) ds + f_0^\theta(C, \Gamma) \right].
\]

**(13)**

**Example. The Markov illness-death model.** Let us apply the formula to a Markov illness-death model; for brevity we shall compute the likelihood only on the event \( \{v_{j-1} < T_1 \leq v_j\} \cap \{T_2 > C\} \) (i.e. in epidemiological language, when a subject has been first seen ill at \( v_j \) and was still alive at \( C \)). On this event \( N_C = 1 \) and in this case: \( f_0^\theta(T_1, \Gamma) = \lambda_1(T_1)e^{-\Lambda_1(C)} \) (here \( \Gamma = T_2 \) and \( T_2 = C \) on this event). We decompose the cumulative total intensity as: \( \Lambda_1(C) = \Lambda_1(v_{j-1}) + \Lambda_1(T_1) + \Lambda_1(v_j) + \Lambda_1(v_j + C) \). We have in the Markov model [from formulae (5) and (6)] that \( \Lambda_1(v_{j-1}) = A_{00},(0, v_{j-1}), \Lambda_1(T_1) = A_{00}(v_{j-1}, T_1), A_0(v_j) = A_{12}(v_j, C) \), where \( A_{ij}(a,b) = \int_a^b \omega(s) ds \). So that, taking out of the integral the terms which do not depend on \( s \) we obtain:

\[
\int_{v_{j-1}}^{v_j} f_0^\theta(s, \Gamma) ds = e^{-A_{00}(0, v_{j-1})} \left[ \int_{v_{j-1}}^{v_j} e^{-A_{00}(v_{j-1}, u)} \omega_1(s) e^{-A_{12}(v_j, C)} ds \right] e^{-A_{12}(v_j, C)}.
\]

This is a product of three terms in which we recognize \( p_{00}(0, v_{j-1}), p_{01}(v_{j-1}, v_j) \) and \( p_{11}(v_j, \tilde{T}) \) (Commenges et al., 2004). Conditioning on \( \{N_m = 0\} \) we have to divide by \( e^{-A_{00}(0, v_m)} \) and we retrieve formula (3).

**4.3. Likelihood for the MDCO scheme: the case when several components are partially observed**

In this section, we shall first compute the likelihood for the following scheme of observation of \( N \): \( N_1 \) is observed at discrete times \( v_0, \ldots, v_m \), and \( N_2 \) is observed at discrete times \( v_0, \ldots, v_m \), while \( N_j \), \( j = 3, \ldots, p \), are observed in continuous time on \([0, C]\) (see section 4.5 for the problem raised by the randomness of the number of observations). The observation in this scheme is represented by the \( \sigma \)-field \( \mathcal{O} = \sigma(N_{1|l}, N_{2|l}, 0 \leq l \leq m_1, 0 \leq l_2 \leq m_2) \cap \mathcal{G}_2^C \), where \( \mathcal{G}_2^C = \sigma(N_{j|l}, j = 3, \ldots, p, 0 \leq u \leq C) \). Using (9) we write \( \mathcal{L}_{\mathcal{F}_C}^0 = f_0^\theta(\tilde{T}_1, \tilde{T}_2, \Gamma^2)g(\Gamma^2)e^{\tilde{\epsilon}_1} + \tilde{T}_2 \), where \( \Gamma^2 \) is a \( \mathcal{G}_2^C \)-measurable random variable. If we note \( \mathcal{O}' = \sigma(N_{1|l}, l_1 = 0, \ldots, m_1) \cap \mathcal{G}_C \) (which we denoted \( O \) in the previous subsection), we obviously have \( O \subset O' \subset \mathcal{F}_C \) so that the observed likelihood can be expressed as \( \mathcal{L}_{\mathcal{F}_C}^0 = E_0[E_0[\mathcal{L}_{\mathcal{F}_C}^0 | O' | O] \]. Lemma 1 gives an expression of \( E_0[\mathcal{L}_{\mathcal{F}_C}^0 | O'] \) and using the same properties as above and the convention \( v_0 = 0 \) and \( v_2 = 0 \), we can write:

Here, we develop an approach which is closer to the statistician’s point of view and which is

\[ \mathcal{L}_O^\theta = \sum_{l_1 = 1}^{m_1} \sum_{l_2 = 1}^{m_2} \sum_{k_1 = 1}^{t_{l_1} + 1} \mathbb{E} \left[ \prod_{l = 1}^{m_2} \left( \int_{l_1 - 1}^{l_1} e^{-l_1 T_{l_2}} \mathbb{P}(T_{l_2} = l_2 | \mathcal{F}_l) \, ds \right) \right] \]

\[ + \sum_{l_1 = 1}^{m_1} \sum_{l_2 = 1}^{m_2} \sum_{k_2 = 1}^{t_{l_2} + 1} \mathbb{E} \left[ \prod_{l = 1}^{m_2} \left( \int_{l_1 - 1}^{l_1} e^{-l_1 T_{l_2}} \mathbb{P}(T_{l_2} = l_2 | \mathcal{F}_l) \, ds \right) \right] \]

We compute each term of the right hand of this equality using the law and the independence of the \( T_i \)s under \( \mathbb{P} \) as in the previous subsection. Dropping the terms which not depend on \( \theta \) (as \( g(\Gamma^2) \)) we have the following lemma.

**Lemma 2**

For \( N_1 \) observed at discrete times \( v_0^1, \ldots, v_{m_1}^1 \) and \( N_2 \) observed at discrete times \( v_0^2, \ldots, v_{m_2}^2 \) and \( N_j (3 \leq j \leq p) \), observed in continuous time the likelihood is given by

\[ \mathcal{L}_O^\theta = \sum_{l_1 = 1}^{m_1} \sum_{l_2 = 1}^{m_2} \sum_{k_1 = 1}^{t_{l_1} + 1} \mathbb{E} \left[ \prod_{l = 1}^{m_2} \left( \int_{l_1 - 1}^{l_1} e^{-l_1 T_{l_2}} \mathbb{P}(T_{l_2} = l_2 | \mathcal{F}_l) \, ds \right) \right] \]

We can prove by induction a formula for the likelihood when \( k \) components of the processes are partially observed. The high number of possibilities for the realization of the \( (T_i)_{1 \leq i \leq k} \) in the intervals \([v_i^j; v_i^{j-1}] \) or \([v_m^j; +\infty[ \) makes the formula complicated. Lemma in the next subsection makes it unnecessary to give such a cumbersome formula.

**4.4. General formula of the likelihood via its local representation**

Here, we develop an approach which is closer to the statistician’s point of view and which is

more general. We begin with a lemma.

**Lemma 3**

Consider two observation schemes of \( N \) yielding observed \( \sigma \)-fields \( \mathcal{O} \) and \( \tilde{\mathcal{O}} \); consider an event \( A \) such that \( \mathbb{P}_0(A) > 0 \), \( A \in \mathcal{O} \cap \tilde{\mathcal{O}} \) and \( A \cap \mathcal{O} = A \cap \tilde{\mathcal{O}} \); then \( \mathcal{L}_O^0 = \mathcal{L}_{\tilde{\mathcal{O}}}^0 \) a.s. on \( A \).
Proof. Remember that \( L^0_\mathcal{O} = E_\mathcal{O}[L^0_F | \mathcal{O}] \) and \( L^0_{\tilde{\mathcal{O}}} = E_\mathcal{O}[L^0_F | \tilde{\mathcal{O}}] \). Direct application of a lemma of local equality of conditional expectations (Kallenberg, 2001, lemma 2, ch. 6) gives the result.

For instance, consider the case where \( N_1 \) is observed at discrete times \( v_0, \ldots, v_m \) and \( N_j, j > 1 \) are observed in continuous time as in section 4.1; take \( A = \{ T_i \in (v_{j-1}, v_j] \} \). We have

\[
L^0_{\mathcal{O}} = \frac{g(\Gamma)}{e^{-v_{j-1}} - e^{-v_j}} \int_{v_{j-1}}^{v_j} f^0(s, \Gamma) \, ds,
\]
a.s. on \( A \) (this can be seen by multiplying both sides of equation (13) by \( 1_A \) which is equal to 1 on \( A \)). From a statistician’s point of view we can say that if \( A \) happens the likelihood takes that particular form which does not depend on the other values of the observation times \( (v_i, f_i \neq l, l - 1) \). So it is obvious either directly or using lemma 3 that if we consider another scheme of observation where \( N_1 \) is observed at \( w_1 = v_{j-1} \) and \( w_2 = v_j \) leading to the likelihood \( L^0_{\tilde{\mathcal{O}}} \), we have \( L^0_\mathcal{O} = L^0_{\tilde{\mathcal{O}}} \) a.s. on \( A \).

This leads to an extension of the field of application of the formulae of the likelihood for incomplete data. A more general scheme of observation can be considered:

**Definition 2 (the deterministic GCMP)**

A deterministic GCMP is a scheme of observation for a multivariate process \( X = (X_1, \ldots, X_p) \) specified by a response function \( r(\cdot) = (r_1(\cdot), \ldots, r_p(\cdot)) \), where the \( r_j(\cdot) \)'s take values 0 or 1, such that \( X_{ij} \) is observed at time \( t \) if and only if \( r_j(t) = 1 \), for \( j = 1, \ldots, p \).

This general (deterministic) scheme applied to \( N \) allows each component \( N_j \) to be observed in continuous time over some windows and in discrete time over other windows. Within this scheme (assuming a family of equivalent probability measures), the likelihood is given by \( E_\mathcal{O}[L^0_F | \mathcal{O}] \), where \( \mathcal{O} = \sigma(r_j(t)N_j; 0 \leq t \leq C; j \in \{1, \ldots, p\}) \). Lemma 3 will help us to give a simple expression of this likelihood if we can find a finite class of events \( (A_k) \) which form a partition of \( \Omega \) and on which the likelihood is relatively easy to compute. If all the observed events had a positive (and bounded away from zero) probability, the class of atoms of \( \mathcal{O} \) would yield a natural finite partition of \( \Omega \) (a \( P \)-atom of a \( \sigma \)-field is a set \( A \) belonging to it, such that \( P(A) > 0 \) and if \( B \subset A \), then \( P(B) = 0 \) or \( P(A \cap \bar{B}) = 0 \)). For instance, this would be the case if all the components of \( N \) were observed in discrete time. As soon as one component may be observed in continuous time, there may still exist atoms, but we do not have a partition of \( \Omega \) with atoms. This leads us to define a class of pseudo-atoms which is a partition of \( \Omega \). For \( 0 \leq j \leq p \), we denote:

- \( v_0^j = 0 \) and \( \{ v_1^j, \ldots, v_{m_j}^j \} \) the finite set of times of discontinuities of the function \( r_j(t) \) on \([0, C] \) such that \( \{ v_0^j = 0 < v_1^j < \ldots < v_{m_j}^j \} \).
- If \( m_j > 0 \) we define \( D_j = \{ T_j \in (v_{l-1}^j, v_l^j) \} \cap \{ r_{j\mu} = 0, u \in (v_{l-1}^j, v_l^j) \} \), \( l \geq 1 \). For some \( l \), \( D_j \) is empty otherwise it is an atom of \( \mathcal{O}_j = \sigma(r_j(t)N_j; 0 \leq t \leq C) \).
- \( E_j = (\{ T_j > v_{m_j}^j \} \cap \{ r_{j\mu} = 0, u > v_{m_j}^j \}) \cup \{ T_j > C \} \) which also is either empty or an atom of \( \mathcal{O}_j \).
- \( C_j = \{ T_j \in r_{j\mu}^{-1}(1) \} \) where \( r_{j\mu}^{-1}(1) \) denotes the interior (in topological sense) of the \( \mathbb{R} \)-subset \( r_{j\mu}^{-1}(1) \) in \([0, v_{m_j}^j] \). On \( C_j \), the event \( T_j \) is exactly observed and \( C_j \) is the complementary set of \( \{(u_1 \leq u \leq u_\mu) \} \cup E_j \) in \( \Omega \).

\( \Omega \) can be partitioned into the disjoint sets \( D_j, l \geq 1 \), \( C_j \) and \( E_j \) for each \( j \) (there is no such \( D_j \) if \( N_j \) is observed in continuous time over \([0, C] \); also \( C_j \) is empty if \( N_j \) is only observed at discrete times). A finer partition can be obtained by the intersection of these \( p \) partitions.
Definition 3 (Pseudo-atoms)
In the deterministic GCMP framework, we call pseudo-atoms of \( \mathcal{O} \), a set \( A = \bigcap_{j=1}^p A_j \), where \( A_j = D_j \) for some \( l \), \( A_j = C_j \) or \( A_j = E_j \) and such that \( P_0(A) > 0 \).

Remark. If \( A \) is a pseudo-atom of \( \mathcal{O} \) we have \( A \in \mathcal{O} \). The class of pseudo-atoms \( (A_k) \) form a partition of \( \Omega : \Omega = \bigcup_k A_k \cup \{ \} \) where \( \forall k \neq k' \).

Example. The Illness-death model with hybrid observation scheme for illness: pseudo-atoms. Consider an illness–death model in which illness represents a complication of a disease. The occurrence of the complication (illness) is observed in continuous time during a sojourn in hospital from time 0 to time \( v_1 \). After the hospitalization, the complication is diagnosed at planned visits at times \( \{v_2, v_3, \ldots, v_m \leq C \} \). We suppose that if death occurs during the study (i.e. between \( [0, C] \)) its time can be retrieved exactly. So the process of observation is given by

\[
\begin{align*}
  r_1(t) &= I_{[0,v_1]}(t) + \sum_{i=2}^m I_{(t=v_i)} \\
  r_2(t) &= 1 \quad 0 \leq t \leq C
\end{align*}
\]

With this scheme, the times of discontinuities and the sets defined previously are:

- For the first component \( r_1(\cdot) \), \( m_1 = m \) and \( v_1 = v_k \). For the second one, \( m_2 = 0 \).
- \( D_{11} \) is empty and for \( 2 \leq l \leq m \), \( D_{u_l} = \{ T_1 \in (v_{l-1}, v_l) \} \). If the event \( D_{u_l} \) occurs for some \( l \geq 2 \) it means that the complication is diagnosed for the first time at the visit \( v_l \). As \( m_2 = 0 \) there is no \( D_{22} \).
- \( E_1 = \{ T_1 > v_m \} \) and \( E_2 = \{ T_2 > C \} \).
- \( C_1 = \{ T_1 \leq v_1 \} \). If \( C_1 \) occurs it means that the illness is diagnosed during the sojourn in hospital. We have \( C_2 = \{ T_2 \leq C \} \).

Note that \( \Omega = C_1 \cup (\bigcup_{l=1}^m D_{u_l}) \cup E_1 \) gives a partition of \( \Omega \) and \( \Omega = C_2 \cup E_2 \) gives another one and the pseudo-atoms are the intersection of sets of these two partitions of positive probability. Now we can give the likelihood by describing it on any pseudo-atom.

Theorem 2
Consider the deterministic GCMP scheme specified by \( r_j(\cdot), j = 1, \ldots, p \), and a pseudo-atom of \( \mathcal{O} \), \( A = \bigcap_{j \in L} \{ T_i \in (v_{i1}, v_{i2}) \} \cap_{j \in L'} \{ T_{l_r} > v_{m_l} \} \cap_{l \in L''} \{ T_l \in r_l^{-1}(1) \} \) where \( L \) and \( L' \) are disjoint subsets of \( \{1, \ldots, p\} \). On \( A \) the likelihood \( L_0^0 \) is equal to that of an MDCO scheme where \( N_l \) is observed at times \( v_{i1} \) and \( v_{i2} \) for \( l \in L \), at \( v_{m_l} \) for \( l' \in L' \) and in continuous time for \( l'' \in L'' = L \cap L' \). Without loss of generality, we can suppose that \( L'' = \{ k + 1, \ldots, p \} \). On \( A \) this likelihood is equal to:

\[
L_0^0 = \frac{e^{\sum_{l \in L} v_
u t_{l\nu}}}{\prod_{l \in L} (e^{-v_{l1}} - e^{-v_{l2}})} \left[ \prod_{l \in L} \left[ \prod_{j \in L} 1_{(v_{i1} < v_{i2})} \prod_{l_r \in L'} 1_{(v_{m_l} < v_{l1})} \right] f_{l1}^0(s_1, \ldots, s_k, \Gamma_k) \prod_{l \in L''} ds_l \right. \\
+ \sum_{l' \in L'} \left[ \prod_{l \in L} 1_{(v_{i1} < v_{i2})} \prod_{l' \in L'} 1_{(v_{m_l} < v_{l1})} \right] f_{l1}^0(s_1, \ldots, C_{l1}, \ldots, s_k, \Gamma_k) \prod_{l' \neq i} ds_{l'} \\
+ \sum_{(v_{i1}, v_{i2}) \in L''^2} \left[ \prod_{l \in L} 1_{(v_{i1} < v_{i2})} \prod_{l' \in L' \setminus i} 1_{(v_{m_l} < v_{l1})} \right] f_{l1}^0(s_1, \ldots, C_{l1}, \ldots, C_{l2}, \ldots, s_k, \Gamma_k) \prod_{l' \neq i} ds_{l'} \\
+ \ldots \right].
\]
The likelihood for general coarsening

Proof. Let \( \hat{\Omega} \) the observed \( \sigma \)-field in the MDCO scheme of theorem 2. We have \( A \in \Omega \), \( A \in \hat{\Omega} \) and \( A \cap \Omega = A \cap \hat{\Omega} \). Lemma 3 gives us \( L^0_\hat{\Omega} = L^0_\Omega \) a.s. on \( A \). The value of the likelihood on this event can be derived using the technique of the preceding subsection.

Example. The Illness–death model with hybrid observation scheme for illness: likelihood. The partition of \( \Omega \) given by the class of pseudo-atoms is \( \{ C_1 \cap C_2; D_{11} \cap C_2, l = 2, \ldots , m; E_1 \cap C_2; C_1 \cap E_2; D_{11} \cap E_2, l = 2, \ldots , m; E_1 \cap E_2 \} \). In this case, we have

\[
\begin{align*}
&f^0(s_1, s_2) = \lambda^0(s_1; s_1, s_2 \wedge s_1) \lambda^2(s_2; s_1 \wedge s_2, s_2) 1_{[t < c]} \exp[-\Lambda^0(C; s_1 \wedge C, s_2 \wedge C)].
\end{align*}
\]

The likelihood is easy to write on each of the pseudo-atoms.

• On \( \{ C_1 \cap C_2 \} \), all the components are observed in continuous time (\( L \) and \( L' \) are empty):

\[
L^0_\hat{\Omega} = L^0_{\mathcal{F}_C};
\]

• On \( D_{11} \cap C_2 \), \( L = \{ 1 \} \) and \( L'' = \{ 2 \} \),

\[
L^0_\hat{\Omega} = \frac{1}{(e^{-v_{l-1}} - e^{-v_l})} \int_{v_{l-1}}^{v_l} f^0(s_1, T_2) \, ds_1;
\]

• On \( E_1 \cap C_2 \), \( L = \{ 2 \} \),

\[
L^0_\hat{\Omega} = e^{\lambda^0} \left[ f^0_{\mathcal{F}_C}(s_1, T_2) \, ds_1 + f^0_{\mathcal{F}_C}(T_2) \right];
\]

• On \( C_1 \cap E_2 \), all the components are observed in continuous time on \( [0, C] : L^0_\hat{\Omega} = L^0_{\mathcal{F}_C}; \)

• On \( D_{11} \cap E_2 \), \( L = \{ 1 \} \) and \( L'' = \{ 2 \} \),

\[
L^0_\hat{\Omega} = \frac{1}{(e^{-v_{l-1}} - e^{-v_l})} \int_{v_{l-1}}^{v_l} f^0(s_1, C) \, ds_1;
\]

• On \( E_1 \cap E_2 \), one finds

\[
L^0_\hat{\Omega} = e^{\lambda^0} \left[ f^0_{\mathcal{F}_C}(s_1, C) \, ds_1 + f^0_{\mathcal{F}_C}(C, C) \right].
\]

4.5. Ignorability of the stochastic censoring by death

In many examples of interest in epidemiology, one of the processes, say \( N_p \), counts death, and the observation of \( N_1, \ldots , N_{p-1} \) may be right censored by the time of death \( T_p \) (in addition to other types of coarsening). For instance, suppose it has been planned to visit a subject at discrete visit times \( v_1^1, \ldots , v_m^1 \leq C \) to observe the process \( N_1 \) (representing an illness status), the last visit time is necessarily random: it is \( v_{M_1} \), where \( M_1 = \max_{j} \{ f:v_{l} < T_p \} \). So for treating this problem we have to resort to the stochastic GCMP: this is the same definition as the deterministic GCMP except that the response process is stochastic and will be denoted by \( R \). It is assumed that \( R \) is observed. In this problem we have to consider the \( \sigma \)-field \( \mathcal{F} = \mathcal{R} \cap N_C \) and the observed \( \sigma \)-field is \( \mathcal{O} = \sigma( R_j, R_p N_p, 0 \leq t \leq C; j = 1, \ldots , p \). Consider the case where the event \( \{ R = r \} \) has a positive probability. If the mechanin leading to incomplete data had been deterministically fixed to be equal to \( r \), the observed \( \sigma \)-field would be \( \tilde{\Omega} = \sigma(r_j T_p, 0 \leq t \leq C; j = 1, \ldots , p \). The mechanism leading to incomplete data is ignorable if on the event \( \{ R = r \} \) using \( L^0_{\tilde{\Omega}} \) leads to the same inference as using \( L^0_\Omega \). Commenges & Gégout-Petit (2005) gave general conditions of ignorability.

Let us treat here the specific problem in which we assume that the stochastic nature of \( R \) comes only from the right censoring of the other processes by death. In that case the response processes can be written:

\[
R_{jt} = r_j(t)1_{\{N_{pt} = 0\}}, \quad 0 \leq t \leq C; \quad j = 1, \ldots, p - 1,
\]

and we assume that \( R_{jt} = 1, 0 \leq t \leq C; \) the \( r_j(t) \)s are deterministic functions as in definition 2. Define \( M_j = \max_t(l' : v'_l < T_j) \) and consider the event \( A = \cap_{j=1}^{p-1} \{ M_j = \tilde{m}_j \} \) for some choice of the \( \tilde{m}_j \)s compatible with the deterministic discontinuities \( v'_l \); \( A \) has then a positive probability if the intensity of \( N_p \) does not vanish on \([0, C]\). As the \( R_j \) are functions of \( N_p \), the observed \( \sigma \)-field can be written: \( \bar{O} = \sigma(R_j(t)1_{\{N_{pt} = 0\}}), \quad j = 1, \ldots, p - 1; \quad N_{pt}, 0 \leq t \leq C). \) For each \( A \) we can define deterministic functions as: \( \tilde{r}_j(t) = r_j(t), t \leq v_{\tilde{m}_j}; \tilde{r}_j(t) = r_j(t_+), t > v_{\tilde{m}_j}; \quad j = 1, \ldots, p - 1. \)

For the deterministic GCMP specified by the \( r_j(t) \)s the observed \( \sigma \)-field is \( \bar{O} = \sigma(\tilde{r}_j(t)N_{jt}), \quad 0 \leq t \leq C; \quad j = 1, \ldots, p \), where \( \tilde{r}_j(t) = 1, 0 \leq t \leq C. \) Because the stochastic right censoring depends only on \( N_p \), which is observed, it is clear that \( A \in \bar{O}. \) Moreover, because the intensities of the processes \( N_j, \quad j = 1, \ldots, p - 1 \) are zero after death we have \( \bar{O} = O \) on \( A \). Thus we can apply lemma 3 to find that on \( A \) we have \( L_{\tilde{O}}^O = L_O^O \). The consequence is that in that case we can use the formulae derived in the deterministic framework (while it can not be deterministic), interpreting \( v'_{\tilde{m}_j} \) as the last discontinuity of \( r_j(t) \) before death.

### 4.6. Extension to general multistate models

Any multistate process can be represented by an informationally equivalent counting process (obvious by the Borgan representation). The approach we have developed for deriving the likelihood in the GCMP context can be applied to general counting processes: that is, compute the likelihood ratio for continuous-time observation and take the conditional expectation given \( O \). However, it seems nearly impossible to obtain a general formula in that case, in particular because the number of transitions which may occur between two observation times is not bounded. In applications this is of course not realistic. For instance, consider a two-state reversible model for diarrhoea: state 0, no diarrhoea; state 1, diarrhoea. By definition a period of diarrhoea lasts a certain time, say 1 day. In that case the number of transitions between two observation times is bounded. It is then possible to cast the problem in the OJC framework because a counting process making at most \( k \) jumps can be represented by a \( k \)-dimensional OJC process.

### 5. Illustration on dementia–institution–death models

#### 5.1. Multistate and counting process models

A five-state non-homogeneous Markov model for dementia, institutionalization and death was proposed by Commenges & Joly (2004) (see Fig. 1); we present this model with somewhat different notations to be in agreement with our general notations of section 4. The first aim was to estimate the eight transition intensities. It was proposed to make no parametric assumption on \( x_{01}(t), x_{02}(t), x_{04}(t) \) but to relate parametrically the other transition intensities to these three basic intensities. Proportionality of the transition intensities towards dementia was assumed: \( x_{23}(t) = x_{01}(t)e^{\theta_1}, \) as well as proportionality of the transition intensities toward institution: \( x_{13}(t) = x_{02}(t)e^{\theta_2}. \) It was assumed that the transitions towards death are all proportional: \( x_{14}(t) = x_{04}(t)e^{\theta_5}; \) \( x_{24}(t) = x_{04}(t)e^{\theta_5}; \) \( x_{34}(t) = x_{04}(t)e^{\theta_5}. \) As shown in section 3 this model can be represented by a trivariate process \( N = (N_1, N_2, N_3) \) where \( N_1 \) counts dementia, \( N_2 \)
counts institutionalization and $N_3$ counts death; the value of the five-state process $X$ at $t$ can be represented in base 2 as: $X_t = \min(N_3,N_2,N_1,4)$. The processes $X$ and $N$ are equivalent in the sense that they generate the same filtration. Moreover if under a probability measure $P$, $X$ is Markov with transition intensities $(\xi_{ij}, h = 0,\ldots, 4; j = 0,\ldots, 4)$ the intensity of $N$ can be deduced by the general formula (4). The model which was proposed by Commenges & Joly (2004) is a Markov semiparametric multiplicative model; we write the intensities as functions of $T_j, j = 1, 2, 3$ and we can verify that the intensities at $t$ only depend on $T_j \land t, j = 1, 2, 3$ (because for instance, $\{T_j \geq t\} = (T_j \land t \geq t)$); they are:

\[
\begin{align*}
\lambda^0_1(t; T_1, T_2, T_3) &= \alpha_0(t) 1_{\{T_1 \geq t\}} 1_{\{T_3 \geq t\}} \exp(1_{T_2 < t}) \\
\lambda^0_2(t; T_1, T_2, T_3) &= \alpha_0(t) 1_{\{T_2 \geq t\}} 1_{\{T_3 \geq t\}} \exp(1_{T_1 < t}) \\
\lambda^0_3(t; T_1, T_2, T_3) &= \alpha_0(t) 1_{\{T_3 \geq t\}} \exp(1_{T_1 < t} + 1_{T_2 < t} + 1_{T_3 < t} + 1_{T_2 < t} + 1_{T_3 < t}),
\end{align*}
\]

(14)

where $\theta = (\eta; \alpha_0(\cdot), j = 1, 2, 4)$ and $\eta$ represents the vector of parameters named with this letter. It is attractive to consider some non-Markovian models and, including an explanatory variable $Z_i$. The model for subject $i$ may be:

\[
\begin{align*}
\lambda^0_1(t; T'_1, T'_2, T'_3) &= \alpha_0(t) 1_{\{T'_1 \geq t\}} 1_{\{T'_3 \geq t\}} \exp(1_{T'_2 < t} + 1_{T'_2 < t} + 1_{T'_3 < t}) \\
\lambda^0_2(t; T'_1, T'_2, T'_3) &= \alpha_0(t) 1_{\{T'_2 \geq t\}} 1_{\{T'_3 \geq t\}} \exp(1_{T'_1 < t} + 1_{T'_1 < t} + 1_{T'_3 < t}) \\
\lambda^0_3(t; T'_1, T'_2, T'_3) &= \alpha_0(t) 1_{\{T'_3 \geq t\}} \exp(1_{T'_1 < t} + 1_{T'_2 < t} + 1_{T'_3 < t} + 1_{T'_2 < t} + 1_{T'_3 < t}).
\end{align*}
\]

5.2. Likelihoods for coarsened observation from the dementia-institution-death model

Dementia is commonly observed in discrete time (because it is diagnosed at planned visits), while time of death and institutionalization can be known exactly. A heuristic likelihood for this MDCO scheme of the semiparametric Markov model was proposed in Commenges & Joly (2004). The results of the present paper allow us to rigorously derive the likelihood, to obtain a more concise formula and this can also be performed for non-Markov models. The model can be described as a trivariate counting process with $N_1$ observed at discrete times $v_0,\ldots, v_m$ and $N_2$ and $N_3$ are observed in continuous time. We will consider that $v_m$ is the last visit really carried out ($v_m < T_3$) and thanks to the result of section 4.5 we will be able to treat $v_m$ as deterministic. Thus we have $r_2(t) = r_3(t) = 1, 0 \leq t \leq C$ and we observe (for each subject) $(T_j, \delta_j), j = 2, 3$, using the standard notation of section 4.1; note that $\delta_j = 1_{C_j}, j = 2, 3$.
with the event $C_j$ (not to be confounded with time $C$) defined in section 4.4. As for $N_1$
we observe the events $D_{ij}, l=1, \ldots, m$ and $E_1$ (or equivalently the value of $1_{D_{ij}}, l=1, \ldots, m$
and $1_{E_1}$). Using the formula of theorem 2, and dropping the multiplicative constant, we have
for instance, on the pseudo-atom $D_1 \cap C_2 \cap C_3$:
\[
\mathcal{L}_o^0 = \int_{v_{i-1}}^{v_i} \lambda_0^0(s_1; s_1, T_2, T_3) \lambda_0^0(T_2; s_1, T_2, T_3) \lambda_0^0(T_3; s_1, T_2, T_3) e^{-\Lambda(t_3, s_1, T_2, T_3)} \, ds_1.
\]

For a more compact expression of the likelihood we may group the formulae for the pseudo-atoms included in $D_1$
by making use of the $\delta_j$s and the $T_j$; the likelihood on $D_1 = \{ T_1 \in (v_{i-1}, v_i] \}$ is:
\[
\mathcal{L}_o^0 = \int_{v_{i-1}}^{v_i} \lambda_0^0(s_1; s_1, \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_2; s_1, \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_3; s_1, \tilde{T}_2, \tilde{T}_3) e^{-\Lambda(t_3, s_1, \tilde{T}_2, \tilde{T}_3)} \, ds_1.
\]

Similarly we obtain on $E_1 = \{ T_1 > v_m \}$:
\[
\mathcal{L}_o^0 = \int_{v_m}^{\tilde{T}_3} \lambda_0^0(s_1; s_1, \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_2; s_1, \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_3; s_1, \tilde{T}_2, \tilde{T}_3) e^{-\Lambda(t_3, s_1, \tilde{T}_2, \tilde{T}_3)} \, ds_1 + \lambda_0^0(\tilde{T}_2; \tilde{T}_3, \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_3; \tilde{T}_2, \tilde{T}_3) \lambda_0^0(\tilde{T}_3; s_1, \tilde{T}_2, \tilde{T}_3) e^{-\Lambda(t_3, \tilde{T}_3, s_1, \tilde{T}_2, \tilde{T}_3)},
\]

where $\Lambda'((t; s, v, w)) = \sum_{j=1}^3 \Lambda_j'((t; s, v, w))$ and $\Lambda_j'((t; s, v, w)) = \int_0^t \lambda_j^0((u; s, v, w)) \, du$, $j = 1, 2, 3$. In this
formula we have replaced the upper bound $C$ of the integral by $\tilde{T}_3$ because $\lambda_0^0(s_1; \tilde{T}_2, \tilde{T}_3) =
0, \tilde{T}_3 < s_1 \leq C$. For the same reason we have replaced $C$ in $\Lambda_0^0(\ldots, \ldots, \ldots)$ by $\tilde{T}_3$.

For the Markov model (14) for instance, the intensities appearing inside the integrals of
(15) and (16) can be computed to be:
\[
\lambda_0^0(s_1; s_1, \tilde{T}_2, \tilde{T}_3) = \xi_{01}(s_1) \exp(\eta_1^0 1_{(\tilde{T}_2 < s_1)})
\]
\[
\lambda_0^0(\tilde{T}_2; s_1, \tilde{T}_2, \tilde{T}_3) = \xi_{02}(\tilde{T}_2) \exp(\eta_2^0 1_{(s_1 < \tilde{T}_2)})
\]
\[
\lambda_0^0(\tilde{T}_3; s_1, \tilde{T}_2, \tilde{T}_3) = \xi_{03}(\tilde{T}_3) \exp(\eta_3^0 1_{(\tilde{T}_2 < \tilde{T}_3)}) + \eta_{12}^0 1_{(\tilde{T}_2 < \tilde{T}_3)}.
\]

Note that the last equality is available only on the set $\{ s_1 < \tilde{T}_3 \}$ (which is the case inside the
integrals) and note moreover that on this event $\lambda_0^0(\tilde{T}_3; s_1, \tilde{T}_2, \tilde{T}_3)$ does not depend on $s_1$; this
comes from the Markov property of this model. The other quantities appearing in (15) and
(16) can be computed by the same mechanical manipulations. The general formula can be
applied even if the model is non-Markovian.

A closer examination of the data available in the PAQUID study (Letenneur et al., 1999)
has revealed that the observations were more incomplete that what was assumed in Comm-
enges & Joly (2004). When a subject is visited at time $v_i$ it is observed whether he lives in
institution or not and the time of entry in institution is retrospectively recorded; information
about institutionalization between the last visit and $\tilde{T}_3$ is often unknown for subjects
who were not institutionalized at the last visit. The response processes are: $R_{ij} = r_i(t) \{ N_{ij} = 0 \}$
where $r_i(t) = 1$ for $t = v_i, j = 1, \ldots, m_i$ and zero elsewhere. We define $M$ as $M = \max_i(t; v_i < \tilde{T}_3)$,
so that $v_{M}$ is the last visit time before $C$ or death; $R_{2i} = 1_{(t < v_{M})}$ and $R_{3i} = 1$ for $0 \leq t \leq C$. It
is seen that $R_{2i}$ depends on the process $N_3$, and moreover, because of the retrospective rec-
collection of the time of entry in institution, it depends on values of $N_3$ for times larger than $t$
(future values !): so it is not obvious that the mechanism leading to incomplete data is
ignorable. The result of section 4.5 does not apply directly to this case but we can apply the
same approach to prove that ignorability holds in this case. So we can still apply the general
formula of theorem 2.
The observed variables for dementia and death are as before, but for institution we observe $C_2, I_C, T_2, E_2$. On $C_2 \cap D_{II}$ (i.e. we observe that the subject was institutionalized at $T_2$ and demented between $v_{i-1}$ and $v_i$) the likelihood is given by (15); on $C_2 \cap E_1$ (the subject was institutionalized at $T_2$ but not demented at the last visit) it is given by (16). Two different formulae are necessary to describe the likelihood on $E_2$. As before the formula of theorem 2 gives us the likelihood on pseudo-atoms and we may group the formulae for the pseudo-atoms included in $E_2 \cap D_{II}$ to obtain:

$$
L^0_{Oj} = \int_{t_{i-1}}^{t_i} \left[ \int_{\tau_M}^{\tau_3} \lambda_1^0(s_1; s_1, s_2, \bar{T}_3) \lambda_2^0(s_2; s_1, s_2, \bar{T}_3) \lambda_3^0(\bar{T}_3; s_1, s_2, \bar{T}_1) \beta_1 e^{-\lambda_1^0(\bar{T}_1; s_1, s_2, \bar{T}_3)} ds_2
\right. \\
+ \lambda_1^0(s_1; s_1, \bar{T}_3, \bar{T}_3) \lambda_2^0(\bar{T}_3; s_1, \bar{T}_3, \bar{T}_3) \beta_1 e^{-\lambda_1^0(\bar{T}_1; s_1, \bar{T}_3, \bar{T}_3)} ds_1,
$$

(17)

and those included in $E_2 \cap E_1$ to obtain:

$$
L^0_{Oj} = \int_{\tau_M}^{\tau_3} \int_{\tau_M}^{\tau_3} \lambda_1^0(s_1; s_1, s_2, \bar{T}_3) \lambda_2^0(s_2; s_1, s_2, \bar{T}_3) \lambda_3^0(\bar{T}_3; s_1, s_2, \bar{T}_3) \beta_1 e^{-\lambda_1^0(\bar{T}_1; s_1, s_2, \bar{T}_3)} ds_2 ds_1
\right. \\
+ \int_{\tau_M}^{\tau_3} \lambda_1^0(s_1; s_1, \bar{T}_3, \bar{T}_3) \lambda_2^0(\bar{T}_3; s_1, \bar{T}_3, \bar{T}_3) \beta_1 e^{-\lambda_1^0(\bar{T}_1; s_1, \bar{T}_3, \bar{T}_3)} ds_1
\right. \\
+ \int_{\tau_M}^{\tau_3} \lambda_2^0(s_2; \bar{T}_3, s_2, \bar{T}_3) \lambda_3^0(\bar{T}_3; \bar{T}_3, s_2, \bar{T}_3) \beta_1 e^{-\lambda_1^0(\bar{T}_1; \bar{T}_3, s_2, \bar{T}_3)} ds_2
\right. \\
+ \lambda_3^0(\bar{T}_3; \bar{T}_3, \bar{T}_3, \bar{T}_3) \beta_1 e^{-\lambda_1^0(\bar{T}_1; \bar{T}_3, \bar{T}_3, \bar{T}_3)}.
$$

(18)

Once the likelihood is computed different approaches for inference are possible; in particular, it was proposed in Commenges & Joly (2004) to use penalized likelihood, with smoothing coefficients chosen by cross-validation and this method gave very satisfactory results in a simulation study.

6. Conclusion

Many multistate models can be considered as generated by simple events so that a direct representation in terms of counting processes may be more economical; this is particularly the case when the events are not repeated and can be modelled by 0–1 counting processes. The multistate point of view, however, retains its interest in many applications, in particular for reversible models: for instance, epileptic crises, repeated diarrhoea periods or repeated hospitalization stays might be modelled by a reversible two-state model if we wish to take into account the duration of the crises or of the hospital stays. So the multistate and the counting process points of view are rather complementary.

Representing multistate models as counting processes models allows us to rigorously derive the likelihood by the use of Jacod’s formula. This was already known for continuous time observation schemes but we were able to apply this approach to the quite general GCMP scheme. This theory will be useful for developing complex models in life history events analysis. In addition to that, having a compact general formula for the likelihood could be exploited, for instance, for designing a software able to automatically treat any IM model: the user could specify his model by a routine giving the the values of the intensities of the OJC model which would be used to compute the likelihood corresponding to the observations described in the data set. This would be in particular feasible for parametric or penalized likelihood approaches; see Commenges et al. (2006) for a penalized likelihood approach of Markov and semi-Markov models.

Acknowledgements

We thank Pierre Joly and Alioum Ahmadou for their comments about the manuscript. We also thank a referee for his careful reading of the paper.

References


Received March 2005, in final form March 2006

Daniel Commenges, INSERM E0338, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, Bordeaux 33076, France.

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