Parametric and Non Homogeneous semi-Markov Process for HIV Control

Eve Mathieu[1], Yohann Foucher[1], Pierre Dellamonica[2], and Jean-Pierre Daures[1]

[1] Clinical Research University Institute. Biostatistics Laboratory.
641 av. D.G. Giraud.
34093 Montpellier , France (e-mail: `emathieu@iurc.montp.inserm.fr`)
[2] Infectious Disease Department
Archet Hospital
BP 3079. 06202 Nice , France

Abstract. In AIDS control, physicians have a growing need to use pragmatically useful and interpretable tools in their daily medical taking care of patients. In that sense, semi-Markov process seems to be well adapted to model the evolution of HIV-1 infected patients. In this study, we introduce and define a Non Homogeneous semi-Markov Model (NHSMM) in continuous time. Then the problem of finding the equations that describe the biological evolution of patient is studied and the interval transition probabilities are computed. A parametric approach is used and the maximum likelihood estimators of the process are given. As results, follow-up time has an impact on the evolution of patients and interval transition probabilities are computed.
Keywords: Semi-Markov process, Non homogeneity, Maximum likelihood estimation, Right censored data, interval transition probabilities.

1 Introduction

The CD4 count and the VL measurement are both fundamental markers of the state of an HIV-1 infected patient. The potential of these immunological and virological reservoirs determines the way the patients are handled. In the context of HIV, it seems reasonable to think that the probability of a patient's transition from one state to another depends on how long he has spent in this state. Therefore the semi-Markov Models (SMM) seem to be appropriated [Janssen and Limnios, 1999].

The SMM have been considered in the HIV modelling [Wilson and Solomon, 1994], [Satten and Sternberg, 1999], [Joly and Commenges, 1999]. These models were time Homogeneous semi-Markov Models (HSMM) and unidirectional. Nowadays it seems to be appropriated to take into account the impact of the follow-up time on the patients' evolution. The goal of this paper is to formulate a Non Homogeneous semi-Markov Model (NHSMM) of the HIV biological process and to compute its interval transition probabilities. The NHSMM have found many applications, in breast cancer [Davidov,

1999], [Davidov and Zelen, 2000] in manpower system [Papadopoulou and Vassiliou, 1999],

[Vassiliou and Papadopoulou, 1992],[Papadopoulou, 1998], [McClean *et al.*, 1998] and [Janssen and Manca, 2001].

This paper is organized as follows. In the next section, the model and associated notation are introduced. Section 3 defines the semi-Markovian interval transition probabilities and solves integral equations. In section 4, the emi-Markov process is parametrically modelled and the likelihood function is built. Section 5 illustrates an application to HIV control. Finally, section 6 is a summary and discussion.

2   Model description and Notation

The natural history of HIV infection can be considered as a series of stages through which a patient progresses. Based both on currently information and physicians' opinion, we have taken four immunological and virological states: state 1 ($VL \leq 400$ and $CD4 \leq 200$), state 2 ($VL \leq 400$ and $CD4 > 200$), state 3 ($VL > 400$ and $CD4 > 200$), state 4 ($VL > 400$ and $CD4 \leq 200$). Patients move thought these four states according ten transitions given in figure 1.
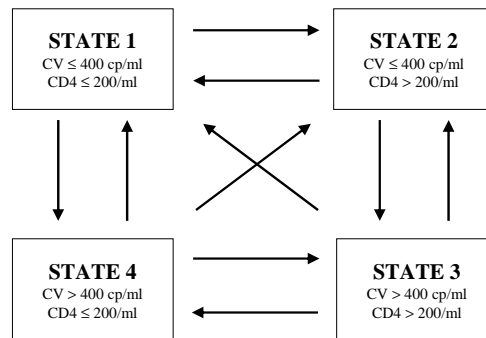
Fig. 1. An HIV Multi-state model, with 4 immunological and virological states and 10 transitions.

More formally, let $E = \{1, 2, 3, 4\}$ be the state space and (Ω, F, P) be a probability space. We define the following random variables [Janssen and Manca, 2001]:

$$J_n : \Omega \to E, \qquad S_n : \Omega \to [0, +\infty),$$

where J_n represents the state at the n-th transition and S_n represents the chronological time of the n-th transition. Let $N(t)$ be the counting process

$(N(t), t \geq 0)$ associated to the point process $(S_n)_{n \in \text{N}}$ defined for any time $t \geq 0$ by :

$$N(t) = \sup\{n : S_n \leq t\}.$$

The random variable $N(t)$ represents the number of transitions occured in the interval of time $[0, t]$. Let us define the $(X_n)_{n \in \text{N}}$ 'duration process' by :

$$X_0 = 0,$$
$$X_{n+1} = S_{n+1} - S_n, \quad n \in \text{N}^*$$

where X_{n+1} represents the duration time spent in state J_n.

The $(J_n, S_n)_{n \in \text{N}}$ process is called 'non-homogeneous Markov renewal process' if :

$$P(J_{n+1} = j, S_{n+1} \leq t| \ J_n = i, S_n = s, J_{n-1}, S_{n-1}, ..., J_0, S_0) = P(J_{n+1} = j, S_{n+1} \leq t| J_n = i, S_n = s),$$

and for $j \neq i$

$$Q_{ij}(s, t) = P(J_{N(s)+1} = j, S_{N(s)+1} \leq t| J_{N(s)} = i, S_{N(s)} = s),$$

is the associated *non-homogeneous semi-Markov kernel* Q. The semi-Markov kernel is written again :

$$Q_{ij}(s, x) = P(J_{N(s)+1} = j, X_{N(s)+1} \leq x| J_{N(s)} = i, S_{N(s)} = s).$$

The second composant of Q, namely x, represents a duration time whereas s represents a chronological time.

As is well known [Wadjda, 1992],

$$p_{ij}(s) = \lim_{x \to \infty} Q_{ij}(s, x), \quad i, j \in E, j \neq i$$
$$= P(J_{N(s)+1} = j| J_{N(s)} = i, S_{N(s)} = s),$$

represents the probability of a patient making its next transition to state j, given that he entered state i at time s and $\mathbf{P}(s) = [p_{ij}(s)]_{i,j}$ is the (4×4) transition probability matrix of the *embedded non-homogeneous Markov chain* $(J_n)_{n \in \text{N}}$.

However, before the entrance into j, the patients 'holds' for a time x in state i. The conditional cumulative distribution function of the waiting time in each state, given the state subsequently occupied, is defined by :

$$F_{ij}(s, x) = P(X_{N(s)+1} \leq x| J_{N(s)+1} = j, J_{N(s)} = i, S_{N(s)} = s).$$

This probability function is obtained by :

$$F_{ij}(s, x) = \begin{cases} \frac{Q_{ij}(s,x)}{p_{ij}(s)} & \text{if } p_{ij}(s) \neq 0 \\ 1 & \text{if } p_{ij}(s) = 0 \end{cases}$$

and for more feasability, it is supposed free of the chronological time s, namely $F_{ij}(x)$. Without loss of generality, the waiting time also has a probability density function, namely $f_{ij}(x)$ and $\mathbf{D}(x) = [f_{ij}(x)]_{i,j}$ represents the (4×4) duration matrix.

Let introduce the probability that the process stays in state i for at least a duration time x, given state i is entered at chronological time s :

$$H_i(s,x) = P(X_{N(s)+1} \leq x | J_{N(s)} = i, S_{N(s)} = s).$$

Of course,

$$H_i(s,x) = \sum_{j \neq i}^{4} Q_{ij}(s,x) = \sum_{j \neq i}^{4} p_{ij}(s) F_{ij}(x).$$

Therefore, the marginal cumulative distribution functions of the waiting time in each state depend on both time. Let us define $S_i(s,x) = 1 - H_i(s,x)$.

Now it is possible to define the continuous time *non homogeneous semi Markov process* $Z(t)$, which represents, for each time t, the state occupied by the process [Cox and Isham, 1980], [Janssen, 1986], as :

$$Z(t) = J_{N(t)}, \quad t \in \mathrm{R}_+.$$

with :

$$P[Z(t) = j] = P[S_{N(t)} \leq t < S_{N(t)+1}, J_{N(t)} = j].$$

This SM process is both characterized by a set of Markov transition matrices $\{\mathbf{P}(t)\}_{t \geq 0}$, and a set of duration matrices $\{\mathbf{D}(x)\}_{x \geq 0}$. Note that two time scales arise, the chronological time and the internal time scales. The chronological time, namely t, is relative to an arbitrary origin. In our case, $t = 0$ represents the first immunological and virological measurement experimented by the patient in hospital. The internal time, namely x, is relative to the duration time in each state [Davidov and Zelen, 2000]. Our model is quite simple and completely defined by both the jump and duration processes. The advantage of semi-Markov model is their mathematical tractability and simple interpretation. The SMM presented in this section is non homogeneous with time since the jump process $(p_{ij}(t))_{i,j,t \geq 0}$ depends on the chronological time.

3   Interval transition probabilities

In the perspective of a more and more effective taking care of patients, physicians need tools of prediction and reference points. Let us define, $\forall i, j = 1, ..., 4$, $\phi_{ij}(t,x)$ as the following probability [Papadopoulou and Vassiliou, 1999] :

$$\phi_{ij}(t,x) = P\left[\text{a patient is in state } j \text{ at time } t+x \mid \text{he entered state } i \text{ at time } t\right].$$
$$= P\left[Z(t+x) = j \mid J_{N(t)} = i \ ; \ S_{N(t)} = t\right]$$

These probabilities are real quantities of interest in the medical practice. Let us precise that $\phi_{ij}(t, x) \neq \phi_{ij}(t + h, x + h), \forall h > 0$. We now turn on the question of developing a functional relationship between the probabilities $\phi_{ij}(t, x)$, which from now on we call the interval transition probabilities of the SM process, and the probabilities $p_{ij}(t)$ and $d_{ij}(x)$. This could be done by taking all the possible mutually exclusive ways in which it is possible for the event of interest to take place. With careful reasoning we could prove that $\forall t, x \geq 0$:

$$\phi_{ij}(t, x) = \delta_{ij} \times S_{i.}(t, x) + \sum_{\substack{l=1 \\ l \neq i}}^{4} \int_0^x p_{il}(t) d_{il}(u) \phi_{lj}(t + u, x - u) du.$$

This equation represents the evolution equation of a continuous NHSMM. Let $c_{il}(t, x)$ be the product $p_{il}(t) d_{il}(x)$. Then the previous equation is written:

$$\phi_{ij}(t, x) = \delta_{ij} \times S_{i.}(t, x) + \sum_{\substack{l=1 \\ l \neq i}}^{4} \int_0^x c_{il}(t, u) \phi_{lj}(t + u, x - u) du. \qquad (1)$$

Obviously $\phi_{ij}(t, 0) = 0$ for $j \neq i$, 1 otherwise. Using probabilistic arguments, we could find probabilities $\phi_{ij}(t, x)$ in closed analytic form. Let k be the index of the number of transitions in the interval of time $]t, t + x[$, and let $t + x_1, t + x_1 + x_2, ..., t + x_1 + x_2 + ... + x_k$ be the chronological times where they successively occur. Then the equation (1) is written as follows

$$\phi_{ij}(t, x) = \delta_{ij} \times S_{i.}(t, x)$$
$$+ \int_0^x c_{ij}(t, x_1) S_{j.}(t + x_1, x - x_1) dx_1$$
$$+ \sum_{\substack{l=1 \\ l \neq i, l \neq j}}^{4} \int_0^x \int_0^{x-x_1} c_{il}(t, x_1) c_{lj}(t + x_1, x_2) S_{j.}(t + x_1 + x_2, x - x_1 - x_2) dx_2 dx_1$$
$$+ \sum_{k=3}^{\infty} \sum_{\substack{l=1 \\ l \neq i}}^{4} \sum_{\substack{m=1 \\ m \neq l}}^{4} \cdots \sum_{\substack{w=1 \\ w \neq v}}^{4} \int_0^x \int_0^{x-x_1} \cdots \int_0^{x-x_1-x_2-...-x_{k-1}}$$
$$c_{il}(t, x_1) c_{lm}(t + x_1, x_2)...c_{wj}(t + x_1 + x_2 + ... + x_{k-1}, x_k)$$
$$\times S_{j.}(t + x_1 + x_2 + ... + x_{k-1} + x_k, x - x_1 - x_2 - ... - x_{k-1} - x_k) dx_k...dx_2 dx_1.$$
$$(2)$$

This previous expression formalizes the fact that the event of interest {a patient of the NHSMM is in in state j at time $t + x$, given he entered state i at time t} may be derived from no transition ($k = 0$) or from exactly one transition ($k = 1$) or from exactly two transitions ($k = 2$) or more ($k \geq 3$). Let us define $\phi_{ij}^k(t, x)$ by the following probability:

$$\phi_{ij}^k(t, x) = P[\text{patient in state } j \text{ at } t + x; k \text{ transitions during }]t, t + x[$$
$$| \text{ he entered state } i \text{ at time } t].$$

Finally the equation (2) can be written $\forall i, j \in \{1, 2, 3, 4\}$

$$\phi_{ij}(t, x) = \sum_{k=0}^{\infty} \phi_{ij}^k(t, x) \qquad (3)$$

and in matrix form, with $\mathbf{\Phi}(t,x) = (\phi_{ij}(t,x))_{i,j}$ and $\mathbf{\Phi}^k(t,x) = (\phi_{ij}^k(t,x))_{i,j}$

$$\mathbf{\Phi}(t,x) = \sum_{k=0}^{\infty} \mathbf{\Phi}^k(t,x). \tag{4}$$

4   The likelihood function

Over a period of time, M patients are observed $(p = 1, ..., M)$. Each patient begins his immunological and virological trajectory in any state, which is revealed by the first measurement at time $s = 0$. Let us assume that the pth subject changes state $(n_p - 1)$ times in the instants $s_{p,1} < s_{p,2} < ... < s_{p,n_p-1}$ and successively occupies states $J_{p,1}, J_{p,2}, ..., J_{p,n_p-1}$ with $J_{p,n} \neq J_{p,n+1}, \forall n \geq 1$. At the last observed time of the follow-up, namely s_{p,n_p}, the patient either may enter a new state J_{p,n_p} or stay in the state J_{p,n_p-1}. In the last case, the last duration time in state J_{p,n_p-1} is right censored. More generally, the contribution for an observed transition $i \rightarrow j$, after a duration time x spent in state i, equals $p_{ij}(t)f_{ij}(x)$, namely the probability $P[$duration time $= x; next = j|$ state i is entered at time $t]$. If the transition from state i is right censored, after a staying time x, then the contribution is the function $S_i(t,x)$. The likelihood function for all times and transition times observed, is written as follows

$$L = \prod_{p=1}^{M} \prod_{n=1}^{n_p} [p_{J_{p,n-1},J_{p,n}}(s_{p,n-1}) f_{J_{p,n-1},J_{p,n}}(s_{p,n}-s_{p,n-1})]^{\xi_{p,n}} [S_{J_{p,n-1}}(s_{p,n-1},\ s_{p,n}-s_{p,n-1})]^{1-\xi_{p,n}}$$

where $\xi_{p,n} = 1$, if the nth transition is observed for the individual p, and $\xi_{p,n} = 0$ if censored. Our parametric approach for both jump and duration processes consists respectively in a linear and a Weibull modelings

$$p_{ij}(t|\theta_{ij}) = a_{ij}t + b_{ij} \qquad \forall j \neq i \tag{5}$$

$$p_{ii}(t) = 0 \quad \forall i = 1, ..., 4$$

$$f_{ij}(x|\gamma_{ij}) = \nu_{ij}\sigma_{ij}^{\nu_{ij}} x^{\nu_{ij}-1} Exp[-(\sigma_{ij}x)^{\nu_{ij}}] \quad \forall j \neq i \tag{6}$$

5   Application to HIV control

In this section, we apply the previous parametric NHSMM to an HIV-1 infected patients database. The database NADIS is made of patients followed in the Nice Hospital, France. The study sample is made of 1313 patients and 17888 virologic and immunologic measurements. The chronological time is

measured from the first biological measurement. From the modelings (5) and (6), we test several restrictions in order to select the parametric model which offers the best adequacy (Likelihood Ratio Test). The selected parametric NHSMM is based on both exponential and Weibull duration times, but also on time linear and constant probabilities.The estimations of the NHSMM parameters are given in Table 1.

Transition $i \rightarrow j$	Estimators of the duration process $d_{ij}(x)$	Estimators of the jump process $p_{ij}(t)$
$1 \rightarrow 2$	$Weibull$ (1.1069 , 1.6795)	$(0.0450 \times t)+ 0.4748$
$1 \rightarrow 3$	$Weibull$ (1.4460 , 1.8283)	0.1111
$1 \rightarrow 4$	$Weibull$ (1.0971, 1.7254)	$(-0.0450 \times t) - 0.4141$
$2 \rightarrow 1$	$Weibull$ (0.5878 , 0.0940)	$(-0.0213 \times t)+ 0.3148$
$2 \rightarrow 3$	$Weibull$ (1.0500, 0.8844)	$(0.0213 \times t) + 0.6852$
$3 \rightarrow 2$	$Expo$ (1.0841)	0.8496
$3 \rightarrow 4$	$Weibull$ (0.7842, 0.7597)	0.1504
$4 \rightarrow 1$	$Weibull$ (0.9095, 1.0556)	$(-0.0276 \times t)+ 0.4779$
$4 \rightarrow 2$	$Weibull$ (1.1866, 1.5765)	0.1605
$4 \rightarrow 3$	$Expo$ (1.8410)	$(0.0276 \times t)+ 0.3616$

Table 1. Estimations of parameters in the NHSMM defined by the linear jump process $\{p_{ij}(t)\}_{i,j}$ and the duration process $\{d_{ij}(x)\}_{i,j}$

Mathematical computing was preformed on R software version 1.9.1. The standard error deviations are not presented for more lisibility. The real quantities of interest are the semi-Markovian interval transition probabilities defined in Section 3. Indeed in medical practice, physicians are often interested in predictions. In this view, the 4×4 matrix of the interval transition probabilities for fixed chronological time t and duration time x,given by equation (4) in section 3, are useful. For exemple, the estimations of $\Phi(0, 1)$ are given in Table 2.

state i \ state j	1	2	3	4
1	0.2059	0.3212	0.2182	0.1935
2	0.0336	0.6530	0.2623	0.0210
3	0.0219	0.4311	0.4673	0.0094
4	0.1464	0.2557	0.2266	0.2649

Table 2. The 4×4 interval transition matrix $(\phi_{ij}(0, 1))_{i,j}$

Given a patient enters state 2 at $t = 0$, he has a 0.652 probability to be 1-year later in state 2; Given a patient enters state 3 at $t = 0$, there is a quasi

equiprobability to be 1-year later in state 2 or 3. Lastly, given a patient enters state 4 at $t = 0$, there is a quasi equiprobability to be 1-year later in state 2, 3 or 4.

6   Discussion

The HIV model considered in this study clearly relates to a 'macroscopic' view of the disease process and it is based both on the CD4 count and VL measurement. This multi-state model is made of 4 immunological and virological states and 10 transitions. The non homogeneous semi-Markov model captures the main features of the disease process and therefore provides a reasonable approximation of a very complicated process. The homogeneity hypothesis reveals to be too restrictive in the HIV context which nowadays becomes a chronic disease. The follow-up time has a significant impact on the disease process. We use a parametric approach and compute the maximum likelihood estimators of the NHSMM. The integral evolution equations of the continuous NHSMM are solved and the interval transition probabilities are computed. Therefore physicians have interesting reference points and some predictions can be made as regards the biological evolution of patients. Here are the three characteristics of a good model which should be mathematically tractable, pragmatically useful and interpretable.

References

[Cox and Isham, 1980]D.R. Cox and V. Isham. *Point Processes.* Chapman and hall, 1980.

[Davidov and Zelen, 2000]O. Davidov and M. Zelen. Designing cancer prevention trials: a stochastic model approach. *Statistics in Medicine*, pages 1983–1995, 2000.

[Davidov, 1999]O. Davidov. The steady-state probabilities for regenerative semi-markov processes with application to prevention and screening. *Applied Stochastic Models and Data Analysis*, pages 55–63, 1999.

[Janssen and Limnios, 1999]J. Janssen and N. Limnios. *Semi-Markov Models and Applications.* Kluwer Academic publishers, 1999.

[Janssen and Manca, 2001]J. Janssen and R. Manca. Numerical solution of non-homogeneous semi-markov processes in transient case. *Methodology and Computing in Applied Probability*, pages 271–293, 2001.

[Janssen, 1986]J. Janssen. *Semi Markov Models. Theory and Applications.* Plenum press, 1986.

[Joly and Commenges, 1999]P. Joly and D. Commenges. A penalized likelihood approach for a progressive three-state model with censored and truncated data: application to aids. *Biometrics*, pages 887–890, 1999.

[McClean *et al.*, 1998]S. McClean, E. Montgomery, and F. Ugwuowo. Non-homogeneous continuous-time markov and semi-markov manpower models. *Applied Stochastic Models and Data Analysis*, pages 191–198, 1998.

[Papadopoulou and Vassiliou, 1999]A.A. Papadopoulou and P.C.G. Vassiliou. Continuous time non homogeneous semi-markov systems. In J. Janssen and N. Limnios, editors, *Semi-Markov Models and Applications*, pages 241–251, 1999.

[Papadopoulou, 1998]A.A. Papadopoulou. Counting transitions-entrance probabilities in non-homogeneous semi-markov systems. *Applied Stochastic Models and Data Analysis*, pages 199–206, 1998.

[Satten and Sternberg, 1999]G.A. Satten and M.R. Sternberg. Fitting semi-markov models to interval-censored data with unknown initiation times. *Biometrics*, pages 507–513, 1999.

[Vassiliou and Papadopoulou, 1992]P.C.G. Vassiliou and A.A. Papadopoulou. Non-homogeneous semi-markov systems and maintainability of the state sizes. *J. Appl. Prob.*, pages 519–534, 1992.

[Wadjda, 1992]W. Wadjda. Uniformly strong ergodicity for non-homogeneous semi-markov processes. *Demonstration Mathematica*, pages 755–764, 1992.

[Wilson and Solomon, 1994]S.R. Wilson and P.J. Solomon. Estimates for different stages of hiv/aids disease. *Comput Appl. Biosci*, pages 681–683, 1994.