Evolution of Bladder Cancer Estimated by Using a State-Space Model with a Semi-Markov Process and Censored Data: A Case Study

Alicia Perez A. P. das Neves Yedig, Cazorla Delia Montoro, Limnios Nikolaos

Austin Journal of Infectious Diseases

Open Access

Austin Publishing Group

Research Article

Evolution of Bladder Cancer Estimated by Using a State-Space Model with a Semi-Markov Process and Censored Data: A Case Study

Alicia Pereira das Neves Yedig[1]; Nikolaos Limnios[2]; Delia Montoro Cazorla[3]

[1]Department of Statistics and Operational Research. University of Jaén, Spain

[2]Laboratoire de Mathématiques Apliquées. Université de Technologie de Compiègne, France

[3]Department of Statistics and Operational Research. University of Jaén, Spain

***Corresponding author:** **Alicia Pereira das Neves Yedig**
Department of Statistics and Operational Research. University of Jaén, Spain.
Email: apyedig@ujaen.es

Received: September 19, 2023
Accepted: December 04, 2023
Published: December 11, 2023

Abstract

We consider a Semi-Markov Process (SMP) to model the evolution of bladder cancer, which takes different states over time. A multi-state model has been constructed and applied to data collected from 847 patients during a period of fifteen years. Biomedicine databases usually contain censored data and this study shows that, despite this, a good fit of the main survival measures is achieved by using our specific model. This paper aims to present estimators for the semi-Markov kernel, the survival function and the mean time to disease progression. The strong consistency properties of the estimators are proved.

Keywords: Bladder cancer; State-space model; semi-Markov process; Censored data; Kaplan-Meier estimator; Survival function; Strong consistency

Introduction

There is a relatively high incidence of bladder cancer in the community. Fortunately, it does not usually progress to the more invasive stages and it as a low mortality rate, but frequently it requires lengthy and expensive treatment. Hence, in the study of the evolution of the disease, indentifying the time in which patients spend in the different states, and the analysis of survival rates are important for the development of more appropriate medical treatments. In the literature, surviving bladder cancer and other diseases has been addressed by using different approaches and methods. The Cox model [4] is often used in these types of studies and the Kaplan-Meyer estimator of the survival function is commonly applied when there are censured data [3]. For bladder cancer, survival from the first recurrence or progression has been studied by using a non-parametric analysis with the Kaplan-Meyer estimator and the Cox model, or an extension of this model, in order to compare the risks among different groups [7,8]. Other applications are [16,17]. Patient monitoring has provided information on the evolution of this disease. Data reveal that after the first surgery to remove the tumor, a patient may experience several recurrences, which, may lead to disease progression in some cases. Multi-states models are dynamic systems for the study of recurrent events, as can be seen in [1,2,9]. They are appropriate for the study of chronic diseases when monitoring a cohort of patients.

In general, dependence on recurrent events is not often addressed, independence is assumed and, consequently, a Markov model can be used to govern the evolution of the bladder carcinoma [20]. Nevertheless, in [18], several multistate models focusing on a cohort of patients derived from the Cox model were applied, considering time-dependence among the covariates.

It is well known that in Markov processes, the future evolution of the process depends only on the current state, while homogeneous semi-Markov processes relax this hypothesis by assuming that the trajectory of the process depends not only on the current state but also on the time spent in it. Besides this, the sojourn time in each state can have an arbitrary distribution. Therefore, the use of phase-type distributions in Markovian processes is a good approach, since they are a dense set in the set of distributions defined in R+, which also involve the versatile Matrix Analytic Methods (MAMs). [15] is a recent paper on bladder cancer and Phase-type distributions, which provides a general fitting procedure for many applications. Multi-state models using semi-Markov processes are applied in [19], and the authors emphasise their suitability for use and how well their method can be adjusted to the problem they are dealing with. The main theoretical reasons are set out in [12-14].

Citation: das Neves Yedig AP, Limnios N, Cazorla DM. Evolution of Bladder Cancer Estimated by Using a State-Space Model with a Semi-Markov Process and Censored Data: A Case Study. Austin J Infect Dis. 2023; 10(4): 1094.

In this work, specific conditions for the construction of our model based on a semi-Markov process are required. The main measures for survival can easily be calculated from estimators which have excellent behavioural properties. These measures are achieved by using versatile procedures and methods suitable for the introduction of censored data and, therefore, they can be applied to other studies in Biomedicine. For the analysis of the strong consistency of the estimators, we propose the definitions and properties in the corresponding section, as well as some essential results that have been studied previously in [5,6,10,11]. This paper is organised as follows: Section 2 presents the database and a statistical description of the main characteristics. In Section 3, we explain the methodology and give the estimators that will be used in Section 4 for calculating the corresponding estimations. In Section 5, we prove the properties of strong consistency for the estimators. Finally, the conclusions and some future extensions of this work are included in Section 6.

The Database

The database comes from the Department of Urology at La Fe University Hospital in Valencia (Spain). Data were collected from patients with bladder cancer from January 1995 to January 2010, a total of 847 patients whose carcinogenic cells were removed. These patients were submitted to periodical check-ups and treatment in accordance with the protocol for this disease. The model we describe in the Methodology was fitted to this set of patients and it was constructed by including the total number of patients in the database. From the sojourn times between recurrences in patients, the estimators of the kernel and the survival function were determined. We have used all the information available, including the censored data.

The total number of patients is submitted to surgery at time $t = 0$. When a patient has a recurrence, the carcinoma is removed and they arrive at state 1; from this point, forward transitions between successive transient states can occur. However, if a patient has disease progression, the bladder is removed and they occupy state P.

In subsequent medical check-ups, patients with recurrence entered new states up to the end of the observation period. A maximum of $i = 13$ recurrences within the observation period was reached by one patient. Thus, the set of transient states was reduced to $E_0 = \{0,1,2,3\}$ since from state 3 there were only two progressions, one patient with 4 recurrences and another with 8 recurrences. Then, the state space can be considered as $E = E_0 \cup D$, where 3 represents three or more than three recurrences and $D = \{4\}$ is the set with an absorbent state of disease progression P. State 4 means patients leaving the system with progression (P), and 5 represents patients leaving the system without progression (NP). The graph for the transitions between states of $E = E_0 \cup D$, is presented in *Figure 1*.

The patients without disease progression leave the system from state NP. This occurs when they are treated in other hospitals or they can not return to the original hospital for other reasons. Specifically, the time spent in the state i of E_0, given that the next state that is visited is *NP*, is considered as a censored sojourn time in state i.

In *Table* 1, we give a summary of the sojourn times, in days. The great difference between state 3 and the other states is, in part, due to the fact that the accumulation of patients in state 3.

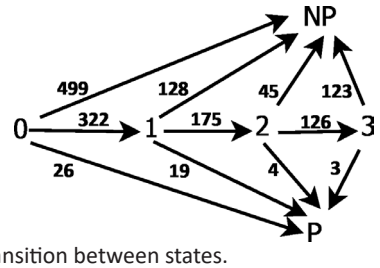

Figure 1: Transition between states.

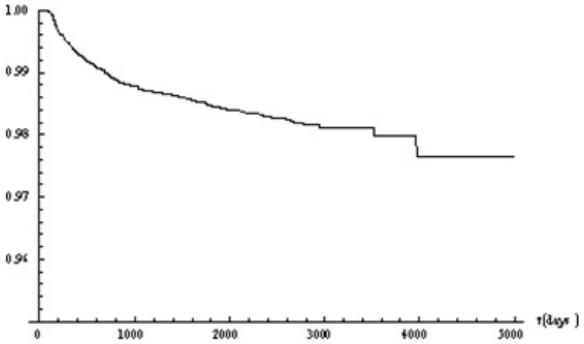

Figure 2: Estimation of the survival function, from state-to-state Kernel estimators.

Table 1: Empirical statistics of the stay times in the different states.

States	n	rang	min	max	med	mean	sd
0	847	4248	85	4333	656	971.02	880.66
1	322	3895	87	3982	399.5	679.55	671.48
2	175	4009	83	4092	372	628.73	660.28
3	126	3192	82	3274	891.5	1101.56	800.25

The minimum number of days that patients sojourn in hospital is similar in all the states, however, the maximum number varies and these numbers differ greatly. The mean times are not very informative since a great standard deviations in all the states. The empirical distributions of these times are biased to the left.

Methodology

Let E be a finite set of states and a stochastic process with values in E, $Z = (Z_t,\ t \in R^+)$, in such a way that:

- The jump times of Z are $S_0 = 0 \leq S_1 \leq \cdots \leq S_n \leq S_{n+1} \cdots$
- The successive states that are visited are: $J_0, J_1, \ldots, J_n, J_{n+1} \cdots$
- The sojourn times in each state, between jumps are: $X_0 = 0$, and

$X_n = S_n - S_{n-1}$, with $n \geq 1$.

The stochastic process (J_n, S_n) defines a Markov renewal process, or equivalently, a semi-Markov process in the state space E if

$$P(J_{n+1} = j,\ S_{n+1} - S_n \leq t \mid J_0, \ldots, J_n,\ S_1, \ldots, S_n) = P(J_{n+1} = j,\ S_{n+1} - S_n \leq t \mid J_n) \quad (a.s.)$$

is verified, where $j \in E$ and $t \in R^+$. The process $Z = (Z_t,\ t \in R^+)$ is said to be a semi-Markov process (SMP).

The process (S_n, J_n) is a Markov chain with a state space $E \times R^+$ and a transition kernel between states $Q_{ij}(t)$, where

$$Q_{ij}(t) = P(J_{n+1} = j,\ S_{n+1} - S_n \leq t \mid J_n = i), \quad i, j \in E$$

and $t \geq 0$

The semi-Markov kernel is a square matrix $Q(t) = \left(Q_{ij}(t)\right)_{(i,j) \in E \times E}$, which takes the following form in our model

$$Q(t) = \begin{bmatrix} 0 & Q_{01}(t) & 0 & 0 & | & Q_{04}(t) \\ 0 & 0 & Q_{12}(t) & 0 & | & Q_{14}(t) \\ 0 & 0 & 0 & Q_{23}(t) & | & Q_{24}(t) \\ 0 & 0 & 0 & 0 & | & Q_{34}(t) \\ - & - & - & - & | & - \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

The matrix block that corresponds to the transient states is

$$Q_0(t) = \begin{bmatrix} 0 & Q_{01}(t) & 0 & 0 \\ 0 & 0 & Q_{12}(t) & 0 \\ 0 & 0 & 0 & Q_{23}(t) \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We suppose that (J_n) is homogeneous in the time Markov chain with the state space E and transition probabilities

$$p_{ij} = Q_{ij}(+\infty) = P(J_{n+1} = j \mid J_n = i)$$

So, (J_n) is called the embedded Markov chain (EMC) in Z. The semi-Markov process associated with the Markov renewal process (J_n, S_n) is $Z = (Z_t)_{t \in R^+}$, where

$$Z_t = J_n, \text{if } S_n \leq t < S_{n+1}$$

Let $F_{ij}(t) = P(X_{n+1} \leq t \mid J_{n+1} = j, J_n = i)$ be the conditional distribution function for the times between jumps.

Then, we can write

$F_{ij}(t) = \frac{P(X_{n+1} \leq t, J_{n+1} = j \mid J_n = i)}{P(J_{n+1} = j \mid J_n = i)} = \frac{Q_{ij}(t)}{p_{ij}}$ for $p_{ij} \neq 0$, and

$$Q_{ij}(t) = p_{ij} F_{ij}(t), \ i, j \in E, \ t \in R^+$$

Moreover, the distribution function of the sojourn time spent in state i by Z is

$$H_i(t) = \sum_{j \in E} Q_{ij}(t), \ i \in E, t \geq 0$$

We also write $H(t) = diag((H_i(t))$.

An interesting property that uses the concept of convolution is

$$P(J_n = j, S_n \leq t \mid J_0 = i) = Q_{ij}^{*(n)}(t)$$

Where $Q_{ij}^{*(n)}$ is the *n-th Lebesgue-Stieljes convolution* of Q_{ij}, then

$$Q_{ij}^{*(n)}(t) = \sum_{h \in E} \int_0^t Q_{ih}(ds) Q_{hj}^{*(n-1)}(t-s), n \geq 2$$

The above equation assumes that $Q_{ij}^{*(1)}(t) = Q_{ij}(t)$ and

$$Q_{ij}^{*(0)}(t) = \mathbf{1}(t) = \begin{cases} 1, if \ t \geq 0 \\ 0, if \ t < 0 \end{cases}$$

The transition function of Z is defined by

$$P_t(i,j) = P(Z_t = j \mid Z_0 = i), i, j \in E, t \in R^+$$

Then, we write in matrix form

$$P(t) = (P_t(i,j))_{(i,j) \in E \times E}$$

A property that allows us to define the survival function [12], is

$$P(t) = [I - Q(t)]^{*(-1)} * [I - H(t)]$$

Then,

$$S(t) = \alpha [I - Q_0(t)]^{*(-1)} * [I - H_0(t)] \mathbf{1}$$

$$= \alpha [I - Q_0(t)]^{*(-1)} * \overline{H_0}(t)$$

where $\overline{H_0}(t)$, a matrix with one column, has its *i-th* element

In this application the initial law vector is $\alpha = (1,0,0,0)$, and

$$\overline{H_0}(t) = \begin{bmatrix} 1 - (Q_{01}(t) + Q_{04}(t)) \\ 1 - (Q_{12}(t) + Q_{14}(t)) \\ 1 - (Q_{23}(t) + Q_{24}(t)) \\ 1 - Q_{34}(t) \end{bmatrix}$$

and

$$[I - Q_0(t)]^{*(-1)} = \begin{bmatrix} 1 & Q_{01}(t) & Q_{01}(t) * Q_{12}(t) & Q_{01}(t) * Q_{12}(t) * Q_{23}(t) \\ 0 & 1 & Q_{12}(t) & Q_{12}(t) * Q_{23}(t) \\ 0 & 0 & 1 & Q_{23}(t) \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then, we obtain the survival function

$$S(t) = 1 - [Q_{04}(t) + Q_{01}(t) * Q_{14}(t) + Q_{01}(t) * Q_{12}(t) * Q_{24}(t) + Q_{01}(t) * Q_{12}(t) * Q_{23}(t) * Q_{34}(t)]$$

In the sequel, we first estimate the semi-Markov kernel $Q(t)$ and then, we introduce it into above equation, in order to obtain an estimator of the survival function $S(t)$.

Empirical Estimator of The Semi-Markov Kernel

We are interested in the estimation of $Q_{ij}(t)$ from a sample of K trajectories, each with an independent semi-Markov process. In fact, for each individual trajectory, we consider an SMP $Z^{(k)}$, $k = 1, \dots, K$. These processes are i.i.d. with a common semi-Markov kernel Q and an initial law α.

If there is no reason to assume that the sojourn time in state i depends on the following state, it is possible to write

$$Q_{ij}(t) = p_{ij} F_i(t), \ i, j \epsilon E, t \in R^+ \text{ and } F_i(t) = P(X_{n+1} \leq t \mid J_n = i)$$

T denotes the length of the observation period in the study, as described in Section (2). The independent processes considered are

$$\left\{ Z_t^{(k)}, 0 \leq t \leq T \right\}, k = 1, \dots, K$$

For each individual process k, with $k = 1, \dots, K$, we define:

- The total number of jumps before reaching T, for the individual process k:

$$N_T^k := sup\{n \in \mathbb{N} : S_n^k \leq T\}$$

- The total number of visits to state i before reaching T, for the individual process k:

$$N_T^k(i) := \sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i\}}, \ i \epsilon E$$

- The total number of jumps from i to j before reaching T, for the individual process k:

$$N_T^k(i,j) := \sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i, J_h^k = j\}}, \ i, j \epsilon E$$

- Regarding the K processes, the total number of visits to i and the total number of jumps from i to j before reaching T are, respectively:

$$N_T(i,K) := \sum_{k=1}^{K} N_T^k(i), \qquad i \in E$$

$$N_T(i,j,K) := \sum_{k=1}^{K} N_T^k(i,j), \qquad i,j \epsilon E$$

From this point, it is possible to apply a filter to our database to select patients who visit state i, so that, the estimation of each factor and finally the SM kernel Q can be calculated.

The first factor of the kernel, which is the transition probability of the EMC, is estimated as:

$$\hat{p}(i,j,K) = \frac{N_T(i,j,K)}{N_T(i,K)}$$

However, for each transient state, it is possible to estimate

$$\widehat{\bar{F}}_i(t) = 1 - \widehat{\bar{F}}_i(t), \; i \in E_0$$

Also, $\widehat{\bar{F}}_i(t)$ can be estimated by using the Kaplan-Meier estimator:

$$\widehat{\bar{F}}_i(t) = \prod_{j:\, t_j \le t} \left(1 - \frac{d_j}{r_j}\right), \quad i \in E_0$$

with $0 < t_1 < \cdots < t_m$ the survival times or arrival times at state 4 from state i, with r_j patients at risk just before t_j and d_j progressions at each t_j, where $j = 1, \ldots, m$.

From the survival times and their corresponding censorship codes, all the necessary values of the *Kaplan-Meier* estimator can be calculated by automatic algorithms, as well as the survival frequency in each survival time and the number of patients at risk.

Other algorithms allow us to calculate the values of the continuous and stepped function $\widehat{\bar{F}}_i(t)$. The calculation of these functions is the step taken before, obtaining the kernels and finally, the survival function, which is obtained by using approximation procedures for the corresponding convolutions.

From the above estimations, we are able to estimate the semi-Markov kernels:

$$\hat{Q}_{ij}(t,K) = \hat{p}_T(i,j,K)\hat{F}_i(t,K), \qquad i,j \epsilon E, t \in R^+$$

Survival Function Estimation

As a result of these calculations, we can define the survival function estimator for our model as:

$$\hat{S}_{K,T}(t) = 1 - [\hat{Q}_{04}(t,K) + \hat{Q}_{01}(t,K) * \hat{Q}_{14}(t,K) + \hat{Q}_{01}(t,K) * \hat{Q}_{12}(t,K) * \hat{Q}_{24}(t,K) + \hat{Q}_{01}(t,K) * \hat{Q}_{12}(t,K) * \hat{Q}_{23}(t,K) * \hat{Q}_{34}(t,K)]$$

Empirical Estimator of the Mean time to Progression

Let us consider the mean time to progression (MTTP), that is, the mean time for reaching the absorbing state P:

$$MTTP = \int_0^{+\infty} S(t)dt$$

It could be useful to estimate, by using right-censored survival data, the expected time to progression in a given interval $[0, T]$ [3]:

$$MTTP_T = \int_0^T S(t)dt$$

This can be estimated as following

$$\widehat{MTTP}_{K,T} = \int_0^T \hat{S}_{K,T}(t)dt$$

Estimation from our Data

In this section, we present the estimations of the semi-Markov kernel and the survival function from our database described in Section 2.

The conditional probabilities of transitions between the states we have been estimated and organised in the following matrix

$$\hat{P}(T,K) = \begin{bmatrix} 0 & \hat{p}_{01} & 0 & 0 & \hat{p}_{04} & \hat{p}_{05} \\ 0 & 0 & \hat{p}_{12} & 0 & \hat{p}_{14} & \hat{p}_{15} \\ 0 & 0 & 0 & \hat{p}_{23} & \hat{p}_{24} & \hat{p}_{25} \\ 0 & 0 & 0 & 0 & \hat{p}_{34} & \hat{p}_{35} \end{bmatrix}$$

The estimation of the survival function taken from the semi-Markov kernel estimator has been represented in Figure 2.

In addition, note how to calculate the estimation of the MTTP in $[0, T]$,

$$\widehat{MTTP}_{K,T} = \int_0^T \hat{S}_{K,T}(t)dt$$

where $T = 5000$ days and $\widehat{MTTP}_{K,T} = 4918.17$ (days) has been calculated by using a procedure with $i = 1549$ steps for successive approximations to the corresponding integral and an error close to 2.31×10^{-7}, between the last two terms, as areas under the stepped functions.

Strong Consistency of the Estimators

We analyse the strong consistency of the estimators defined in Section (3), as $K \to +\infty$ for a fixed $T > 0$.

Estimators of the Transition Probabilities

Proposition 5.1: The estimator of the transition probabilities of the S-M process is strongly consistent, i.e.,

$$\lim_{K\to+\infty} \frac{\sum_{k=1}^K N_T^k(i,j)}{\sum_{k=1}^K N_T^k(i)} = \lim_{K\to+\infty} \frac{N_T(i,j,K)}{N_T(i,K)} = p_{ij} \quad [a.s.]$$

Proof: In the In the context of semi-Markov processes, we have

$$p_{ij} = P(J_n = j \mid J_{n-1} = i), \qquad i,j \in E$$

Let $X^k = N_T^k(i,j) - p_{ij}N_T^k(i)$, $k = 1, \ldots, K$, with $E[X^k] = a$

We are interested in this value because we know that, according to the Strong Law of Large Numbers, when $K \to +\infty$, the next convergence is almost sure:

$$\lim_{K\to+\infty} \overline{X}_K = \lim_{K\to+\infty} \frac{1}{K} \sum_{k=1}^K \left(N_T^k(i,j) - p_{ij}N_T^k(i)\right) = a \;\; [a.s.]$$

We will demostrate that $a = 0$, from which we obtain:

$$\lim_{K\to+\infty} \frac{\sum_{k=1}^K N_T^k(i,j)}{\sum_{k=1}^K N_T^k(i)} = \lim_{K\to+\infty} \frac{N_T(i,j,K)}{N_T(i,K)} = p_{ij} \quad [a.s.]$$

From state $l \in E$, where $l = J_0$:

$$E_l[N_T^k(i,j)] = E_l\left[\sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i,\, J_h^k = j\}}\right] =$$

$$= E_l\left[E_l\left[\sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i,\, J_h^k = j\}} \mid N_T^k\right]\right] =$$

$$= \sum_{n\ge1}\left[E_l\left[\sum_{h=1}^{N_T^k} \mathbf{1}_{\{J_{h-1}^k = i,\, J_h^k = j\}} \mid N_T^k = n\right] P_l(N_T^k = n)\right] =$$

$$= \sum_{n\ge1}\left[E_l\left[\sum_{h=1}^{n} \mathbf{1}_{\{J_{h-1}^k = i,\, J_h^k = j\}}\right] P_l(N_T^k = n)\right] =$$

$$= \sum_{n\ge1}\sum_{h=1}^{n} P_l(J_{h-1}^k = i,\; J_h^k = j)\, P_l(N_T^k = n) =$$

$$= \sum_{n\ge1}\sum_{h=1}^{n} P_l(J_{h-1}^k = i)\, P_l(J_h^k = j \mid J_{h-1}^k = i)\, P_l(N_T^k = n)$$

$$= p_{ij} \sum_{n\ge1}\sum_{h=1}^{n} P^{h-1}(l,i)\, P_l(N_T^k = n)$$

With a similar procedure, we can get:

$$E_i[N_T^k(i)] = E_i\left[\sum_{n=1}^{N_T^k} \mathbf{1}_{\{J_{n-1}^k = i\}}\right] =$$

$$= E_i\left[E_i\left[\sum_{n=1}^{N_T^k} \mathbf{1}_{\{J_{n-1}^k = i\}} \mid N_T^k\right]\right] =$$

$$= \sum_{n\geq 1}\left[E_i\left[\sum_{n=1}^{N_T^k} \mathbf{1}_{\{J_{n-1}^k = i\}} \mid N_T^k = n\right] P_i(N_T^k = n)\right] =$$

$$= \sum_{n\geq 1}\left[E_i\left[\sum_{n=1}^{n} \mathbf{1}_{\{J_{n-1}^k = i\}}\right] P_i(N_T^k = n)\right] =$$

$$= \sum_{n\geq 1}\sum_{n=1}^{n} P_i(J_{n-1}^k = i)\, P_i(N_T^k = n) =$$

$$= \sum_{n\geq 1}\sum_{n=1}^{n} P^{n-1}(l,i)\, P_i(N_T^k = n)$$

Finally,

$$E\left[N_T^k(i,j) - p_{ij}\, N_T^k(i)\right] = 0$$

The Kaplan-Meier Product Limit Estimator

Let $X_1, \ldots, X_n$ be independent positive random variables with the common continuous distribution function F. Independent to those variables, let $U_1, \ldots, U_n$, also be independent positive random variables, which possibly have the non-continuous and defective common distribution function G.

Gill [11] and Borgan [3], analyse the problem that concerns how to perform a non parametric inference of F based on the censored observations $(C_i, \delta_i),\ i = 1, \ldots, n$, with

$$C_i = min\{X_i, U_i\}, \delta_i = \mathbf{1}_{\{X_i \leq U_i\}}$$

Let H be the distribution of the variables C_i, given by

$$1 - H = (1 - F)(1 - G)$$

Traditionally, F is estimated by the 1958 Kaplan-Meier product-limit estimator, defining processes N and Y on $[0, +\infty)$ by using

$$N(t) = \#\{i: C_i \leq t, \delta_i = 1\} \text{ and } Y(t) = \#\{i: C_i \geq t\}$$

Therefore,

$$\hat{F}(t) = 1 - \widehat{\bar{F}}(t) = 1 - \prod_0^t \left(1 - \frac{\Delta N(s)}{Y(s)}\right),$$

In the above equality, $\Delta N(s) = N(s) - N(s-)$, where $N(s-)$ denotes the left-hand limit of the N(t) at s, i.e. the limit of N(t) when $t \to s-$. Then, $\Delta N(s)$ defines the number of deaths at time s.

Proposition 5.2: Lemma 2.8 from Gill [11], says that for any τ such that $H(\tau^-) < 1$:

$$sup_{t\leq\tau}\left|\hat{F}(t) - F(t)\right| \xrightarrow{n\to+\infty} 0\ [a.s.]$$

Estimators of Kernels and Survival Function

Proposition 5.3: From the two previous results, we conclude that The S-M kernel estimator $\hat{Q}_{ij}(t,K)$ is strongly consistent, i.e.,

$$\hat{Q}_{ij}(t,K) \xrightarrow{K\to+\infty} Q_{ij}(t)\ [a.s.]$$

Proposition 5.4: The convolution of kernel estimators is also strongly consistent, i.e.,

$$(\hat{Q}_{ij} * \hat{Q}_{jh})(t,K) \xrightarrow{K\to+\infty} (Q_{ij} * Q_{jh})(t)\ [a.s.],\ i,j,h \in E$$

Proof: We know that $|\hat{Q}_{ij}(t,K)| \leq 1$ and $|\hat{Q}_{ij}(t,K)\ \hat{Q}_{jh}(t,K)| \leq 1$. From the above proposition, we also know that

$$\hat{Q}_{ij}(t,K) \xrightarrow{K\to+\infty} Q_{ij}(t)\ [a.s.]$$

So, we have

$$\hat{Q}_{ij}(s,K)\ \hat{Q}_{jh}(t-s,K) \xrightarrow{K\to+\infty} Q_{ij}(s)Q_{jh}(t-s)\ [a.s.]$$

Finally, the Lebesgue dominated convergence theorem gives us the conclusion.

From the definition of the survival function for this application and all the previous properties, we obtain:

Proposition 5.5: The estimator $\hat{S}_{K,T}(t)$ is strongly consistent, i.e.,

$$\hat{S}_{K,T}(t) \xrightarrow{K\to+\infty} S(t)\ [a.s.]$$

Estimator of the mean time to Progression

The MTTP in a given interval $[0, T]$ is

$$MTTP_T = \int_0^T S(t)dt$$

This is estimated by using

$$\widehat{MTTP}_{K,T} = \int_0^T \hat{S}_{K,T}(t)dt$$

Proposition 5.6: The estimator $\widehat{MTTP}_{K,T}$ is strongly consistent, i.e.,

$$\widehat{MTTP}_{K,T} \xrightarrow{K\to+\infty} MTTP_T\ [a.s.]$$

Proof: We know that: $|\hat{S}_{K,T}(t)| \leq 1$, from its definition.

$\hat{S}_{K,T}(t) \xrightarrow{K\to+\infty} S(t)\ [a.s.]$, according to the previous proposition.

Then, the Lebesgue dominated convergence theorem gives us the conclusion,

$$\int_0^T \hat{S}_{K,T}(t)dt \xrightarrow{K\to+\infty} \int_0^T S(t)\,dt\ [a.s.]$$

Conclusions

The analysis of the evolution of diseases from experimental data is a difficult task as it is carried out by working on databases in which the amount of information available is not as large as required, and much of the information comes from censored data. Therefore, the model we have developed shows a remarkable goodness of fit to the corresponding real process. This is a consequence of its strong properties even though the model has relaxed assumptions. These assumptions are very important for the applicability of the process to other medical contexts. The SMP model we have built for this disease enables its kernels to be estimated, that is, the conditional distribution function in each transient state until the process jumps to another one. The estimation of the survival function and the mean time to the progression of the disease are important achievements that have been made with the model. These achievements provide useful information to medical experts for the creation of new treatments and preventive actions, especially when the survival time is long, as observed in this case. Moreover, each estimation has involved procedures that have been generalised and they can be used in other applications.

A way to continue this first stage of research, is the study of other properties in the estimators we have used, such as asymptotic normality. This desirable property allows a parametric

inference to be made, for example, the calculus of confidence intervals is of interest. Another possible extension of this research is the introduction of explanatory covariates into this study, pursuing a more detailed and differential analysis among possible different risk groups, which will surely be applicable to other diseases.

References

1. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. Int J Epidemiol. 2015; 44: 324-33.
2. Andersen PK, Keiding N. Multi-state models for event history analysis. Stat Methods Med Res. 2002; 11: 91-115.
3. Borgan O. Kaplan-Meier estimator, encyclopedia of biostatistics. John Wiley & Sons. 2005.
4. Cox DR. Regression models and life tables [with discussion]. J R Stat Soc B. 1972; 34: 187-202.
5. Dumitrescu M, Gámiz ML, Limnios N. Minimum divergence estimators for the Radon-Nikodym derivatives of the Semi-Markov kernel, A Journal of Theoretical Applied Statistics. Statistics. 2016; 50: 486-504.
6. G'amiz ML, Kulasekera KB, Limnios N, Lindqvist BH. Applied Nonparametric Statistics in Reliability. London: Springer. 2011.
7. Garćıa B, Rubio G, Santamaŕıa C, Pontones JL, Vera CD, Jimenez JF. A predictive mathematical model in the recurrence of bladder cancer. Math Comput Modell. 2005; 42: 621-34.
8. Garćıa-Mora B, Santamaŕıa C, Rubio G, Luis Pontones JL. Modeling the recurrence–progression process in bladder carcinoma. Comput Math Appl. 2008; 56: 619-30.
9. Geskus RB. Data analysis with competing risks and intermediate States. Chapman & Hall, CRC Press; 2016.
10. Gill RD. Testing with replacement and the product limit estimator. Ann Statist. 1981; 9: 853-60.
11. Gill RD. Large sample behaviour of the product limit estimator on the whole line. Ann Statist. 1983; 11: 49-58.
12. Limnios N, Oprisan G. Semi-Markov Processes and Reability, Birkha¨user. Boston; 2001.
13. Limnios N, Ouhbi B. Empirical estimators of Reliability and related functions for Semi-Markov Systems, Mathematical and Statistical Methods in Reliability, B. Lindqvist, K. Doksum. World Scientific, 2004.
14. Limnios N. Reliability measures of semi-markov systems with General State Space. Methodol Comput Appl Probab. 2012; 14: 895-917.
15. Montoro-Cazorla D, Pérez-Ocón R, Pereira das Neves-Yedig AM. A Longitudinal Study of the Bladder Cancer Applying a State-Space Model with non-exponential Staying Time in States. Mathematics. 2021; 9: 363.
16. Pérez-Ocón R, Ruiz-Castro JE, Gámiz-Pérez ML. A piecewise Markov process for analyzing survival from breast cancer in different risk groups. Stat Med. 2001; 20: 109-22.
17. Pérez-Ocón R, Ruiz-Castro JE, Gámiz-Pérez ML. Nonhomogeneous Markov Models in the Analysis of Survival after Breast Cancer. J R Stat Soc C. 2001; 50: 111-24.
18. Porta N, Calle ML, Malats N, Gómez G. A dynamic model for the risk of bladder cancer progression. Stat Med. 2012; 31: 287-300.
19. Ruiz-Castro JE, Pérez-Ocón R. A Semi-Markov model in biomedical studies. Commun Stat Theor Methods. 2004; 33: 437-55.
20. Santamaŕıa C, Garćıa-Mora B, Rubio G, Navarro E. A Markov model for analyzing the evolution of bladder carcinoma. Math Comput Modell. 2009; 50: 726-32.