Nonparametric Estimation for Semi-Markov Processes Based on K-Sample Paths with Application to Reliability

Nikolaos Limnios[1] and Brahim Ouhbi[2]

[1] Laboratoire de Mathématiques Appliquées,
Université de Technologie de Compiègne,
B.P. 20529, 60205 Compiègne Cedex, France
[2] Ecole Nationale Supérieure d'Arts et Métiers
Marjane II, Meknès Ismailia,
Béni M'Hamed, Meknès, Maroc,

Abstract. The problem concerned here is the estimation of ergodic finite semi-Markov processes from data observed by considering K independent censored sample paths with application in the reliability.

1 Introduction

Semi-Markov modeling, as a generalization of Markov modeling, is a an active area in research. See, e.g., [Alvarez, 2005]-[Voelkel and Cronwley, 1984].

In our previous work [Ouhbi and Limnios, 1999], we have considered one trajectory in the time interval $[0, T]$, and given the estimators and their asymptotic properties, as $T \to \infty$. In the present work, we consider K trajectories in the time interval $[0, T]$, generated by K independent semi-Markov processes having the same semi-Markov kernel Q and initial distribution α. We obtain asymptotic properties of the estimators when $K \to \infty$. In this case the time T is finite and fixed. This type of observation can be viewed as a generalization of the fixed (or type I) censoring of a single failure time. Our method, as in our previous works [Ouhbi and Limnios, 1999, Ouhbi and Limnios, 2003, Ouhbi and Limnios, 2001], consists in obtaining estimators of the semi-Markov kernel, by using a maximum likelihood estimator (MLE) of the hazard rate function of transitions between states, and then considering estimators of other quantities, as the semi-Markov transition function, Markov renewal function, and reliability functions as statistical functionals of the semi-Markov kernel via analytic explicit formula.

2 Estimation of the hazard rate function of transitions

We will consider in this paper a semi-Markov process with a finite state space, $E = \{1, 2, ..., s\}$ say, with irreducible embedded Markov chain and finite sojourn time in all states [Limnios and Oprişan, 2001].

In this section, we will derive and study the maximum likelihood estimator of the hazard rate functions of piecewise constant type estimator (PEXE).

Let us suppose that the semi-Markov kernel Q is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}_+$ and denote by q its Radon-Nikodym derivative, that is, for any $i, j \in E$,

$$\frac{Q_{ij}(dt)}{dt} =: q_{ij}(t). \tag{1}$$

So, we can write also $q_{ij}(t) = P(i,j)f_{ij}(t)$, where f_{ij} is the density function of the distribution F_{ij}.

For any i and j in E, let us define the hazard rate function of transition distributions between states, $\lambda_{ij}(t)$, $t \geq 0$, of a semi-Markov kernel by

$$\lambda_{ij}(t) = \begin{cases} \frac{q_{ij}(t)}{1-H_i(t)} & \text{if} \quad P(i,j) > 0 \quad \text{and } H_i(t) < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Let us also define the cumulative hazard rate from state i to state j at time t by $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(u)du$ and the total cumulative hazard rate of state i at time t by $\Lambda_i(t) = \Sigma_{j\in E}\Lambda_{ij}(t)$. We have also

$$Q_{ij}(t) = \int_0^t \exp[-\Lambda_i(u)]\lambda_{ij}(u)du. \tag{3}$$

Let us consider now a family of K independent E-valued Markov renewal processes $(J_n^r, S_n^r, n \geq 0), 1 \leq r \leq K$, defined by the same semi-Markov kernel Q, and the initial distribution α, that is, for any r, $1 \leq r \leq K$,

$$Q_{ij}(t) := \mathbf{P}(J_{n+1}^r = j, S_{n+1}^r - S_n^r \leq t \mid J_n^r = i), \quad i, j \in E, t \in \mathbf{R}_+, n \in \mathbf{N},$$

$$\alpha(i) = \mathbf{P}(J_0^r = i), \quad i \in E.$$

For any r, let us denote by $N_i^r(t), N_{ij}^r(t), N^r(t), ...$ the corresponding quantities $N_i(t)$, $N_{ij}(t)$, $N(t), ...$, and define further

$$N_i(t,K) := \sum_{r=1}^{K} N_i^r(t), \quad N_{ij}(t,K) := \sum_{r=1}^{K} N_{ij}^r(t), \quad N(t) := \sum_{r=1}^{K} N^r(t). \tag{4}$$

If $t = T$ fixed, then we will note simply $N_i, N_{ij}, ...$

The maximum likelihood estimator of the hazard rate functions will be based upon the observation of the above K independent MRP $\{(J^r, S^r) = [(J_n^r, S_n^r)_{n\geq 0}], 1 \leq r \leq K\}$.

We assume hereafter that we observe each MRPs over the period of time $[0, T]$ for some finite and fixed T. A sample or history for the r-th MRP is given by

$$\mathcal{H}^r(K) = (J_0^r, J_1^r, ..., J_{N^r(T)}^r, X_1^r, X_2^r, ..., X_{N^r(T)}^r, U_T^r), \tag{5}$$

where $U_T^r = T - S_{N^r(T)}^r$ is the backward recurrence time.

The log-likelihood function associated to $(\mathcal{H}^r(T), 1 \leq r \leq K)$ is:

$$l(K) = \log L(K) = \sum_{r=1}^{K} \Big\{ \sum_{l=1}^{N^r(T)} [\log \lambda_{J_{l-1}^r, J_l^r}(X_l^r) - \Lambda_{J_{l-1}^r}(X_l^r)] - \Lambda_{J_{N^r(T)}^r}(U_T^r) \Big\}. \tag{6}$$

In the sequel of this paper, we will approximate the hazard rate function $\lambda_{ij}(t)$ by the piecewise constant function $\lambda_{ij}^*(t)$ defined by $\lambda_{ij}^*(t) = \lambda_{ij}(v_k) = \lambda_{ijk} \in \mathbf{R}_+$

for $t \in (v_{k-1}, v_k] = I_k$, $1 \leq k \leq M$, where $(v_k)_{0 \leq k \leq M}$ is a regular subdivision of $[0, T]$, that is, $v_k = k\Delta$, $0 \leq k \leq M$, $M = M(K)$, with step $\Delta := T/M$, such that, as $K \to \infty$, $\Delta \to 0$, and $K\Delta \to \infty$.

Hence,

$$\lambda^*_{ij}(t) = \sum_{k=1}^{M} \lambda_{ijk} \mathbf{1}_{(v_{k-1}, v_k]}(t), \tag{7}$$

where $\mathbf{1}_{(v_{k-1}, v_k]}(t)$ is equal to 1 if $t \in (v_{k-1}, v_k]$, and 0 otherwise. We get

$$l(K) = \sum_{i,j \in E} \sum_{k=1}^{M} (d_{ijk} \log \lambda_{ijk} - \lambda_{ijk} \nu_{ik}), \tag{8}$$

where $d_{ijk} = \sum_{r=1}^{K} \sum_{l=1}^{N^r(T)} \mathbf{1}_{\{J^r_{l-1}=i, J^r_l=j, X^r_l \in I_k\}}$ is the number of transitions from state i to state j for which the observed sojourn time in state i belongs to I_k, and ν_{ik} is the trace of the sojourn time in state i on the interval time I_k, given for $N(T) \geq 1$. The r.v. ν_{ik} can be represented by the sum of two r.v. as follows

$$\nu_{ik} := \nu^1_{ik} + \nu^2_{ik},$$

where ν^1_{ik} is the trace of the sojourn time on the interval I_k, of the sojourn times in state i, and ν^2_{ik} is the trace of the cumulated censored time T greater than v_k, in state i.

So, the maximum likelihood estimator $\widehat{\lambda}_{ijk}$ of λ_{ijk} is given by:

$$\widehat{\lambda}_{ijk} = \begin{cases} d_{ijk}/\nu_{ik} & \text{if} \quad \nu_{ik} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the estimator $\widehat{\lambda}_{ij}(t, K)$ of $\lambda_{ij}(t)$ is then given by

$$\widehat{\lambda}_{ij}(t, K) = \sum_{k=1}^{M} \widehat{\lambda}_{ijk} \mathbf{1}_{(v_{k-1}, v_k]}(t). \tag{9}$$

Let us also define

$$\widehat{\Lambda}_i(t, K) = \sum_{j \in E} \int_0^t \widehat{\lambda}_{ij}(u, K) du, \quad \widehat{\Lambda}_{ik} = \widehat{\Lambda}_i(v_k, K) = \Delta \sum_{j \in E} \sum_{l=1}^{k} \widehat{\lambda}_{ijl}, \tag{10}$$

and

$$\nu^l_i(t) = \sum_{k=1}^{M} \nu^l_{ik} \mathbf{1}_{(v_{k-1}, v_k]}(t), \quad l = 1, 2.$$

3 Maximum likelihood and empirical estimators of the semi-Markov kernel

Let us define estimators of the semi-Markov kernel by putting estimators (9) to (10), as follows

$$\widehat{Q}_{ij}(t, K) := \Delta \sum_{\{k: 0 \leq v_k \leq t\}} e^{-\widehat{\Lambda}_{ik}} \widehat{\lambda}_{ijk}.$$

Consider now the empirical kernel function defined by

$$\widetilde{Q}_{ij}(t,K) := \frac{1}{N_i}\sum_{r=1}^{K}\sum_{l=1}^{N^r}\mathbf{1}_{\{J^r_{l-1}=i,J^r_l=j,X^r_l\le t\}}, \tag{11}$$

and the empirical density kernel given by:

$$\widetilde{q}_{ij}(t,K) = \frac{\widetilde{Q}_{ij}(v_k,K)-\widetilde{Q}_{ij}(v_{k-1},K)}{\Delta}, \quad \text{if } t\in I_k.$$

Define also the estimator $\widetilde{H}_i(t,K)$ by

$$\widetilde{H}_i(t,K) := \sum_{j\in E}\widetilde{Q}_{ij}(t,K). \tag{12}$$

We define a function $G_i(\cdot,K)$, for $t\in I_k$, $1\le k\le M$, by

$$G_i(t,K) := \sum_{r=1}^{K}\Big\{\sum_{l=1}^{N^r}\frac{X^r_l-v_{k-1}}{N_i\Delta}\mathbf{1}_{\{J^r_{l-1}=i,X^r_l\in I_k,X^r_l\ge t\}}+\frac{U^r_T-v_{k-1}}{N_i\Delta}\mathbf{1}_{\{J^r_{N^r(T)}=i,U^r_T\in I_k\}}\Big\}.$$

Now, let us write estimator (9), as follows

$$\widehat{\lambda}_{ij}(t,K) = \frac{\widetilde{q}_{ij}(v_k,K)}{1-\{\widetilde{H}_i(v_k,K)-G_i(v_k,K)\}+h^r_i(t,K)}, \text{ if } t\in I_k,$$

where

$$h^r_i(t,K) := \frac{\nu^2_{ik}}{N_i\Delta} = \frac{1}{N_i}\sum_{r=1}^{K}\mathbf{1}_{\{J^r_{N^r(T)}=i,U^r_T>v_k\}}.$$

In order to obtain a consistent estimator, we will neglect the term $h^r_i(t,K)$ from the denominator of estimator $\widehat{\lambda}_{ij}(t,K)$, and obtain a new modified estimator denoted by $\widehat{\lambda}^0_{ij}(t,K)$. That is,

$$\widehat{\lambda}^0_{ij}(t,K) = \frac{\widetilde{q}_{ij}(v_k,K)}{1-\{\widetilde{H}_i(v_k,K)-G_i(v_k,K)\}}, \quad \text{if } t\in I_k. \tag{13}$$

Denote the corresponding cumulative hazard rates estimator by $\widehat{\Lambda}^0_i(t)$, and $\widehat{\Lambda}^0_{ij}(t)$.

Lemma 1 *The estimator $\widehat{\lambda}^0_{ij}(t,K)$, is a consistent estimator of $\lambda_{ij}(t)$, as $K\to\infty$.*

Since $h^r_i(t,K)$ converges to a positive quantity, it is clear the estimator $\widehat{\lambda}_{ij}(t,K)$ is not consistent.

In the sequel of this paper, we will consider only the estimator $\widehat{\lambda}^0_{ij}(t,K)$. So, the MLE $\widehat{Q}_{ij}(t,K)$ in (11) is obtained by using this estimator. In the remaining of this section we will study the asymptotic properties of the semi-Markov kernel estimator given by (11).

Theorem 1 *The empirical estimator of the semi-Markov kernel is uniformly strongly consistent, in the sense that, as $K\to\infty$,*

$$\max_{i,j}\sup_{t\in[0,T]}\left|\widetilde{Q}_{ij}(t,K)-Q_{ij}(t)\right|\xrightarrow{a.s.}0.$$

We will prove now that the semi-Markov kernel estimator, obtained from modified PEXE of the hazard rate function $\widehat{\lambda}^0_{ij}(t,K)$, is asymptotically uniformly equivalent to the empirical estimator $\widetilde{Q}_{ij}(t,K)$.

Lemma 2 *Let i and j be any two fixed states. Then we have, for any $t \in [0,T]$,*

$$\widehat{Q}_{ij}(t,K) - \widetilde{Q}_{ij}(t,K) = O(K^{-1}), \quad \text{as } K \to \infty.$$

¿From the previous lemma, we conclude that the estimator of the semi-Markov kernel is asymptotically uniformly a.s. equivalent to the empirical estimator of the semi-Markov kernel for which we will prove the uniform strong consistency and derive a central limit theorem.

Corollary 1 *The estimator of the semi-Markov kernel $\widehat{Q}_{ij}(t,K)$ is uniformly strongly consistent, that is, when K tends to infinity,*

$$\max_{i,j} \sup_{t\in[0,T]} \left| \widehat{Q}_{ij}(t,K) - Q_{ij}(t) \right| \xrightarrow{a.s.} 0.$$

Theorem 2 *For any $i,j \in E$ and $t \in [0,T]$ fixed, $K^{1/2}[\widehat{Q}_{ij}(t,K) - Q_{ij}(t)]$ converges in distribution, as $K \to \infty$, to a zero mean normal random variable with variance $Q_{ij}(t)(1-Q_{ij}(t))[(\alpha\psi)(T)\mathbf{1}]$.*

4 The estimator of the reliability function and its asymptotic properties

After having outlined the problem of estimating the semi-Markov transition matrix, it is appropriate to give some concrete applications of these processes as models of evolution of the reliability function of some system.

Let the state space, E, be partitioned into two sets, $U = \{1,...,r\}$ the patient is in good health and $D = \{r+1,...,s\}$ the patient is ill due to some causes or the component is failed and under repair. Reliability models whose state space is partitioned in the above manner will be considered here. As indicated above, it is of interest to estimate the distribution function of the waiting time to hit down states (failure).

We focus on the estimation of the reliability function for a semi-Markov process which describes the stochastic evolution of system. The general definition of the reliability function in the case of semi-Markov processes is

$$R(t) = \mathbf{P}(Z_u \in U, \quad \forall\, u \le t).$$

The reliability function $R(t)$ is given by:

$$R(t) = \sum_{i\in U} \alpha(i) R_i(t), \tag{14}$$

where $R_i(t)$ is the conditional reliability function, that the hitting time to D, starting from a state $i \in U$, is greater than the time t. It is easy to show, by a renewal argument, that $R_i(t)$ satisfies the following Markov renewal equation

$$R_i(t) - \sum_{i\in U} \int_0^t R_j(t-u) Q_{ij}(du) = 1 - H_i(t), \quad i \in U.$$

The solution of this MRE, see Section 2, together with (14), in matrix form, gives

$$R(t) = \alpha_0 (I - Q_0(t))^{(-1)} * (I - H_0(t))\mathbf{1}, \tag{15}$$

where $\mathbf{1} = (1, ..., 1)^\top$ is an r-dimensional column vector. Index 0 means restriction for matrices on $U \times U$, and for vectors on U.

We will give an estimator of the reliability function of semi-Markov processes and prove its uniform strong consistency and weak convergence properties as $K \to \infty$.

Let $\widehat{Q}$ be the modified MLE of PEXE type of the transition probability of the semi-Markov kernel Q. Then we propose the following estimator for the reliability function

$$\widehat{R}(t,K) = \widehat{\alpha}_0 (I - \widehat{Q}_0(t,K))^{(-1)} * (I - \widehat{H}_0(t,K))\mathbf{1}, \tag{16}$$

and we will prove now its uniform strong consistency and central limit theorems.

Theorem 3 *The estimator of the the reliability function of the semi-Markov process is uniformly strongly consistent in the sense that,*

$$\sup_{t \in [0,T]} \left| \widehat{R}(t,K) - R(t) \right| \xrightarrow{a.s.} 0, \quad K \to \infty.$$

Set

$$B_{ij}(t) := \sum_{n \in U} \sum_{k \in U} \alpha(n) B_{nijk} * (1 - H_k)(t).$$

Theorem 4 *For any fixed $t \in [0,T]$, the r.v. $K^{1/2}[\widehat{R}(t,K) - R(t)]$ converges in distribution to a zero mean normal random variable with variance*

$$\sigma_S^2(t) := \sum_{i \in U} \sum_{j \in U} \mu_{ii} \{[B_{ij} - (\alpha\psi)_i]^2 * Q_{ij}(t) - [(B_{ij} - (\alpha\psi)_i) * Q_{ij}(t)]^2\}.$$

5 Numerical Application

In this section we present a numerical example for a three state semi-Markov process for which we will consider $K = 50$ censored trajectories. The time interval is $[0,T]$, with $T = 1000$.

The conditional transition functions $F_{ij}(t)$ are the following $F_{12}(t)$, and $F_{31}(t)$ are exponential with parameters respectively 0.1 and 0.2, and $F_{21}(t)$, $F_{23}(t)$ are Weibull with parameters respectively $(0.3, 2)$, and $(0.1, 2)$ (scale and shape parameter). The other functions are identically 0.

The transition probabilities $P(2,1)$ and $P(2,3)$ are:

$$P(2,1) = 1 - P(2,3) = \int_0^\infty [1 - F_{23}(t)] dF_{23}(t).$$

The results obtained here are illustrated in figure 1. These results concern the reliability function.

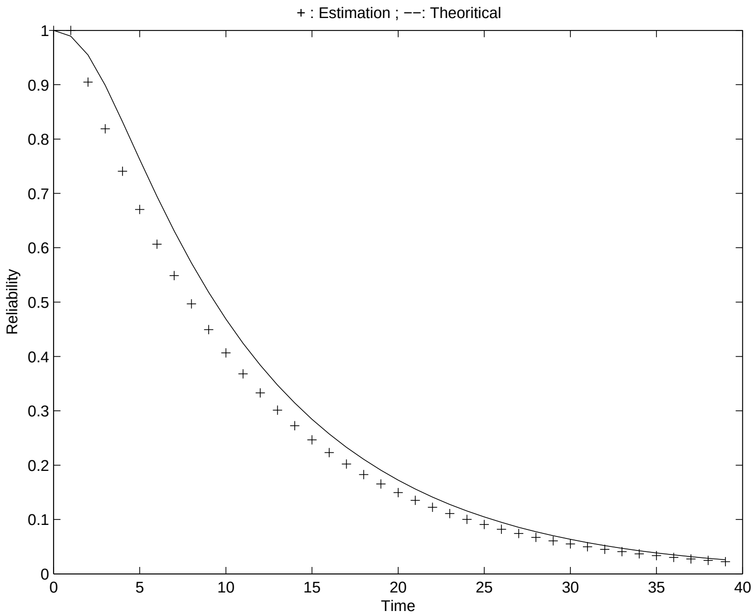

Fig. 1. Reliability estimation

References

[Alvarez, 2005]E.E.E. Alvarez (2005). Smothed nonparametric estimation in window censored semi-Markov processes, *J. Statist. Plann. Infer.*, 131, 209–229.

[Andersen *et al.*, 1993]P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding (1993). *Statistical Models Based on Counting Processes*, Springer, N.Y.

[Dabrowska *et al.*, 1994]D. Dabrowska, G. Horowitz, M. Sun (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *J. Amer. Statist. Ass.*, 89, 867-877.

[Gill, 1980]R.D. Gill (1980). Nonparametric estimation based on censored observations of Markov renewal process. *Z. Wahrsch. verw. Gebiete.* 53, 97–116.

[Greenwood and Wefelmeyer, 1996]P. E. Greenwood, W. Wefelmeyer (1996). Empirical estimators for semi-Markov processes. *Math. Methods Statist.* 5(3), 299–315.

[Lagakos *et al.*, 1978]S.W. Lagakos, C.J. Sommer, M. Zelen (1978). Semi-Markov models for partially censored data, *Biometrika*, 65(2), 311–317.

[Limnios, 2004]N. Limnios (2004). A functional central limit theorem for the empirical estimator of a semi-Markov kernel, *J. Nonparametric Statist.*, 16(1-2), pp 13-18.

[Limnios and Oprişan, 2001]N. Limnios, G. Oprişan (2001). *Semi-Markov Processes and Reliability*, Birkhäuser, Boston.

[Limnios and Ouhbi, 2003]N. Limnios, B. Ouhbi, "Empirical estimators of reliability and related functions for semi-Markov systems", In *Mathematical and*

Statistical Methods in Reliability, B. Lindqvist, K. Doksum (Eds.), World Scientific, 2003.

[Ouhbi and Limnios, 1996]B. Ouhbi, N. Limnios (1996). Non-parametric estimation for semi-Markov kernels with application to reliability analysis, *Appl. Stoch. Models Data Anal.*, 12, 209–220.

[Ouhbi and Limnios, 1999]B. Ouhbi, N. Limnios (1999). Non-parametric estimation for semi-Markov processes based on their hazard rate. *Statist. Infer. Stoch. Processes*, 2(2), 151–173.

[Ouhbi and Limnios, 2001]B. Ouhbi, N. Limnios (2001). The rate of occurrence of failures for semi-Markov processes and estimation. *Statist. Probab. Lett.*, 59(3), 245–255.

[Ouhbi and Limnios, 2003]B. Ouhbi, N. Limnios (2003). Nonparametric reliability estimation of semi-Markov processes. *J. Statist. Plann. Infer.*, 109(1/2), 155–165.

[Phelan, 1990]M.J.Phelan (1990). Estimating the transition probabilities from censored Markov renewal processes. *Statist. Probab. Letter.*, 10, pp 43–47.

[Pyke, 1961]R. Pyke (1961). Markov renewal processes: definitions and preliminary properties, *Ann. Math. Statist.*, 32, 1231–1241.

[Ruiz-Castro and Pérez-Ocon, 2004]J. E. Ruiz-Castro, R. Pérez-Ocon (2004). A semi-Markov model in biomedical studies, *Commun. Stat. - Theor. Methods*, 33(3).

[Voelkel and Cronwley, 1984]J.G. Voelkel, J. Cronwley (1984). Nonparametric inference for a class of semi-Markov processes with censored observations, *Ann. Statist.*, 12(1), pp 142–160.