Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 339 (2004) 137–140
COMPTES RENDUS MATHEMATIQUE

Probability Theory/Statistics

Non-parametric estimation of lifetime and repair time criteria for a semi-Markov process

Anne-Laure Afchain

Cemagref, unité GPAN, parc de Tourvoie, BP 44, 92163 Antony cedex, France

Received 20 August 2002; accepted after revision 13 May 2004

Available online 17 June 2004

Presented by Paul Deheuvels

Abstract

In this Note, we model an industrial system by a semi-Markov process where failure and repair phenomena are in mutual competition. A non-parametric estimation method for system component lifetime and repair time distributions and for associated hazard rate functions is proposed. The lifetime and repair time empirical distributions are reduced to two Kaplan–Meier estimators. A numerical example from an industrial system with three components and one repair man modeled by a birth and death process is provided to illustrate the previous results. ***To cite this article: A.-L. Afchain, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

Résumé

Estimation non paramétrique des indicateurs de survie et de réparation pour un processus semi-markovien. Cette note modélise, par un processus semi-markovien, un système industriel où les phénomènes de panne et de réparation sont en compétition mutuelle. Une méthode d'estimation non paramétrique pour les distributions de durée de survie et de réparation d'un composant du système et pour les fonctions taux de hasard associées est proposée. Les distributions empiriques de durée de survie et de réparation sont réduites à deux estimateurs de Kaplan–Meier. Un exemple numérique tiré d'un système industriel à trois composants et un réparateur modélisé par un processus de naissance et de mort, est fourni à titre d'illustration. ***Pour citer cet article : A.-L. Afchain, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

Let assume a situation where two phenomena, the failure and the repair of an industrial system are in mutual competition, but independent. On a censored period, we observe this industrial system composed by m components and one repair man [1]. The components are independent and identical with respect to failure and complete repair

E-mail address: anne-laure.afchain@cemagref.fr (A.-L. Afchain).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2004.05.006

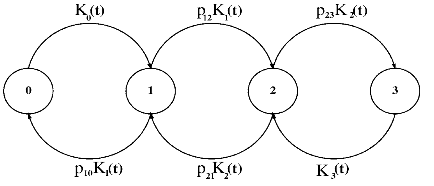

Fig. 1. Example of semi-Markov graph with $m = 3$ and one repair man

treatments. The system with m components is modelled by a semi-Markov process [7], whose states correspond to the number of breakdown state components (see Fig. 1).

In this report, we consider a random censure model as described in Fleming and Harrington [5] (p. 90) and a non-parametric estimation approach for the system component lifetime and repairtime distributions and the associated instantaneous hazard rate functions is proposed.

Definition 1. A random censure model is defined by the pairs $(Z_i, \delta_i)_{0\leqslant i\leqslant m}$ where

- $Z_i = \min(U_i, V_i)$, Z_i is a positive random variable corresponding to the sojourn time in state i with distribution K_i, U_i being the repair time in state i with distribution G_i and $V_i = \min(S_{i+1}, S_{i+2}, \dots, S_m)$, S_{i+1} being the onset of the first failure for the $m - i$ components which are not in breakdown order with distribution F_{i+1}. The positive random variables U_i and V_i are independent.
- δ_i defines the dummy variable, taking values in the set $\{0, 1\}$:

$$\delta_i = 1_{\{U_i \geqslant V_i\}} = \begin{cases} 1, & i \to i+1, \\ 0, & i \to i-1, \end{cases} \qquad i = 0, 1, \dots, m$$

  with $i \to i - 1$ ($i \to i + 1$, respectively) meaning that one out of m components is under repair (i.e. fails, respectively).

Definition 2. From the random censure model, the (Z_i, δ_i)-distribution is defined, for $i = 0, \dots, m$, by $K_{ij}(t) = \mathbb{P}(Z_i \leqslant t, \delta_i = j)$ with $j = 0, 1$.

From (Z_i, δ_i)-distribution (see Definition (2)), we construct the estimator of $(F_{i+1}(t), G_i(t))_{1\leqslant i\leqslant m-1, t\in\mathbb{R}^+}$ defined by $(\widehat{F}_{i+1}(t), \widehat{G}_i(t))_{1\leqslant i\leqslant m-1, t\in\mathbb{R}^+}$. From definition of $Z_i = \min(U_i, V_i)$, we can write the Efron mechanism [2], written as $1 - K_i(t) = [1 - F_{i+1}(t)]^{m-i}[1 - G_i(t)]$. Thus, a *m-differential equation system* is provided with respect to unknowns F_{i+1} and G_i written as:

$$\begin{cases} \mathrm{d}K_{i0}(t) = \big[1 - F_{i+1}(t)\big]^{m-i}\, \mathrm{d}G_i(t) = \left[\dfrac{1 - K_i(t)}{1 - G_i(t)}\right] \mathrm{d}G_i(t), \\ \mathrm{d}K_{i1}(t) = (m-i)\big[1 - F_{i+1}(t)\big]^{m-i-1}\big[1 - G_i(t)\big]\, \mathrm{d}F_{i+1}(t) = (m-i)\left[\dfrac{1 - K_i(t)}{1 - F_{i+1}(t)}\right] \mathrm{d}F_{i+1}(t). \end{cases} \tag{1}$$

Differential equation system (1) must undergo the *estimation procedure* on observation period $[0; T]$ of $\mathbb{R}^+$ [5] (p. 26, Eq. (3.1)) [4]. The lifetime and repair time empirical distributions can be reduced to two Kaplan–Meier estimators [6] from random sample $(Z_{i(k)}, \delta_{i(k)})_{0\leqslant i\leqslant m, 1\leqslant k\leqslant n}$ with size n defined by $n = \sum_{i=0}^m n_i = \sum_{i=0}^m (n_{i0} + n_{i1})$ where $i(k)$ means that, at the k-th jump, the semi-Markov process stays in state i. Besides, $Z_{i(k)}$ ($\delta_{i(k)}$, respectively) is defined as the k-th realization of random variable Z_i (indicator δ_i, respectively) when the process visits n_i times the state i (and when it makes a transition towards the previous in n_{i0} times or the following state in n_{i1} times, respectively). The *discretization* stage consists of splitting up $[0; T]$ into $m + 1$ irregular sub-intervals $[0; Z_{i(n_i)}[$ with i varying from 0 to m. Then, each sub-interval $[0; Z_{i(n_i)}[= \bigcup_{k=1}^{n_i} [Z_{i(k-1)}; Z_{i(k)}[$ is split into N constant steps with length $h_i = Z_{i(n_i)}/N$, N being arbitrarily fixed. In each interval $[Z_{i(k-1)}; Z_{i(k)}[$ for all $k = 1, \ldots, n_i$, beginning at the failure time and ending before the true one which follows, the lifetime distribution is piecewise constant.

Let us take for example the case of $\widehat{K}_{i1}$ (the case of $\widehat{K}_{i0}$ is treated in the same way to obtain $\widehat{G}_i$, the repair time empirical distribution). After the discretization stage and the estimation procedure with $\Delta\widehat{F}_{i+1}(k) = \widehat{F}_{i+1}(k+1) - \widehat{F}_{i+1}(k) = \widehat{\overline{F}}_{i+1}(k) - \widehat{\overline{F}}_{i+1}(k+1)$ and $\overline{F}_{i+1}(k) = 1 - F_{i+1}(k)$, we deduce that:

$$\widehat{\overline{F}}_{i+1}(k+1) = \left\{1 - \frac{\Delta\widehat{K}_{i1}(k)}{(m-i)[1-\widehat{K}_i(k)]}\right\}\widehat{\overline{F}}_{i+1}(k).$$

By simplification member to member of the previous equation for $l = 0, \ldots, k-1$ (with $\widehat{\overline{F}}_{i+1}(0) = 1$), we obtain the following expressions:

$$1 - \widehat{F}_{i+1}(k) = \prod_{l=0}^{k-1}\left\{1 - \frac{\Delta\widehat{K}_{i1}(l)}{(m-i)[1-\widehat{K}_i(l)]}\right\} \Rightarrow \widehat{F}_{i+1}(t,T) = 1 - \prod_{s\leqslant t}\left\{1 - \frac{\Delta\widehat{K}_{i1}(s,T)}{(m-i)[1-\widehat{K}_i(s,T)]}\right\} \tag{2}$$

with $\widehat{K}_{i1}(t,T) = \frac{1}{n_i}\sum_{k=1}^{n} 1_{\{Z_{i(k)}\leqslant t, \delta_{i(k)}=1\}}$, $\Delta\widehat{K}_{i1}(t,T) = \frac{1}{n_i}$ and $\widehat{K}_i(t,T) = \frac{1}{n_i}\sum_{k=1}^{n} 1_{\{Z_{i(k)}\leqslant t\}}$.

Let take for example the case of λ_{i+1} called functional of component lifetime density [9] (p. 147). The case of the repairtime density, μ_i is treated in the same way. From system (1) whose solutions are $\widehat{F}_{i+1}$ and $\widehat{G}_i$, the failure rate function and its estimator verify, for all interval $[0; Z_{i(n_i)}]$ defined above, the following relation, such as, for all $i = 0, \ldots, m-1$, we have:

$$\lambda_{i+1}(t)\,\mathrm{d}t = \frac{\mathrm{d}F_{i+1}(t)}{1-F_{i+1}(t)} = \frac{\mathrm{d}K_{i1}(t)}{(m-i)[1-K_i(t)]} \quad\Rightarrow\quad \hat{\lambda}_{i+1}(t,T)h_i = \frac{\Delta\widehat{K}_{i1}(t,T)}{(m-i)[1-\widehat{K}_i(t,T)]} \tag{3}$$

where term $\mathrm{d}K_{ij}(t)$ is estimated by $\frac{1}{n_{ij}}\sum_{k=1}^{n} 1_{\{Z_{i(k-1)}\leqslant t\leqslant Z_{i(k)}, \delta_{i(k)}=j\}}$. The term dt is replaced by constant step h_i. Under Definition 1, the failure rate function wanted is estimated by the following expression:

$$\hat{\lambda}_{i+1}(t,T) = \frac{1_{\{Z_{i(k-1)}\leqslant t\leqslant Z_{i(k)}, \delta_{i(k)}=1\}}}{n_i(m-i)h_i[1-\widehat{K}_i(Z_{i(k-1)})]}. \tag{4}$$

Proposition 3. *The lifetime and repairtime empirical distributions defined by (2) are almost sure consistent estimators, when t tends towards infinity, in the sense that:*

$$\begin{cases} \widehat{F}_{i+1}(t,T) - F_{i+1}(t) \to 0 & \text{(a.s.)},\\ \widehat{G}_i(t,T) - G_i(t) \to 0 & \text{(a.s.)}. \end{cases}$$

Proof of Proposition 3. Let us take for example the case of $\widehat{F}_{i+1}$ (the case of $\widehat{G}_i$ is treated in the same way). By combining Eq. (2) with Eq. (3) and by using the limited development of $\log(1-x)$ for all x on the neighbourhood of zero, we obtain the following result:

$$1 - \widehat{F}_{i+1}(t,T) = \prod_{1\leqslant k\leqslant n_i : Z_{i(k)}\leqslant t}\left\{1 - h_i\hat{\lambda}_{i+1}(Z_{i(k)})\right\}^{\delta_{i(k)}} \approx \exp\left\{-\sum_{k=1}^{n_i}\delta_{i(k)}h_i\hat{\lambda}_{i+1}(Z_{i(k)})\right\}$$

with $1 - \widehat{K}_i(t,T) \neq 0$ and h_i being finite. Let consider an interval $[0;\tau]$, $\tau < T$ with $Z_{i(0)} = 0, \ldots, Z_{i(n_{i1})} = t$, $\ldots, Z_{i(n_i)} = \tau$, such as:

$$\sum_{k=1}^{n_{i1}} h_i\hat{\lambda}_{i+1}(Z_{i(k)}) \approx \int_0^t \hat{\lambda}_{i+1}(s,T)\,\mathrm{d}s \quad \text{with} \quad \hat{\Lambda}_{i+1}(t,T) = \int_0^t \hat{\lambda}_{i+1}(s,T)\,\mathrm{d}s \quad \forall t\in[0;\tau].$$

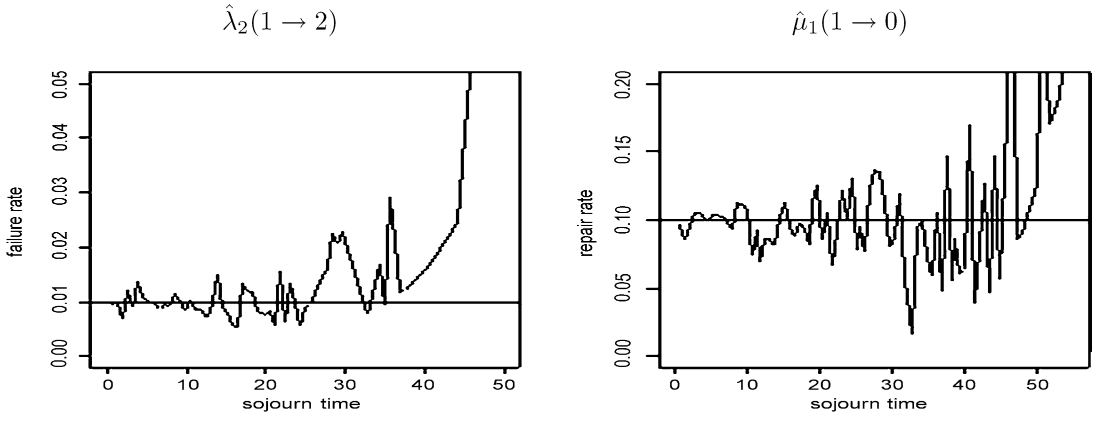

Fig. 2. Curves of failure rate function $\hat{\lambda}_2$ and of repair rate function $\hat{\mu}_1$ versus sojourn time sample $(Z_{1(k)})_{1\leqslant k\leqslant n_1}$ with $n = 10\,000$ and $(\lambda, \mu) = (0.1, 0.01)$ for an industrial system with $m = 3$ components and one repairman.

Thus, we obtain the relation $\widehat{F}_{i+1}(t, T) \approx 1 - \mathrm{e}^{-\hat{\Lambda}_{i+1}(t,T)}$. Given that series $(\hat{\Lambda}_{i+1})_{0\leqslant i\leqslant m-1}$ is in probability convergent [5] (pp. 92–94), then series $(\widehat{F}_{i+1})_{0\leqslant i\leqslant m-1}$ is consistent by exponential continuity for large t (for more details, see [8]). By analogy, we deduce that series $(G_i)_{1\leqslant i\leqslant m}$ admits a consistent estimator for large t. □

In order to validate the non-parametric estimation approach with a numerical viewpoint, we *simulate* a birth and death process (Markov process) which models an industrial system with three components and one repair man. The component lifetime and repairtime distributions are computed from a sample constituted by n successive sojourn times in the different states of the Markov process. They are exponential with constant parameters λ and μ, respectively and are well-numerically compared to the ones from formulae (2) (see [1], p. 171, Fig. 2.3). The failure and repair rate functions λ_2 and μ_1 deduced from formulae (3) are represented on Fig. 2 (see [3] for the simulation method).

Acknowledgements

The author would like to thank Professor Paul Deheuvels, Mr Gerard Derzko and the anonymous referees, for suggestions which helped in improving this report.

References

[1] A.L. Afchain, Robustesse d'évaluation appliquée à des problèmes de fiabilité markovienne en Sûreté de Fonctionnement, Université de Nantes, Thèse, 22 Octobre 2001.
[2] E. Bradley, The two-sample problem with cendored data, in: Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, 1967, pp. 831–853.
[3] G. Derzko, Une approche intrinsèque de l'estimation non paramétrique de la densité, C. R. Acad. Sci. Paris, Ser. I 327 (12) (1998) 985–988.
[4] G. Derzko, E. Leconte, Estimation non paramétrique d'incidence d'événements en compétition avec censure à droite, J. Soc. Française Stat. 1 (2004), in press.
[5] T.R. Fleming, D.P. Harrington, Counting Processes and Survival Analysis, Willey, New York, 1991.
[6] E.L. Kaplan, P. Meier, Non-parametric estimation from incomplete observations, J. Am. Statist. Assoc. 53 (1958) 457–481.
[7] N. Limnios, G. Oprisan, Semi-Markov Processes and Reliability, Birkhäuser, Boston, 2001.
[8] R.S. Liptser, A.N. Shiryaev, Statistics of random process. II: Applications, Springer, 1978.
[9] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, 1986.