Asymptotic Efficiency in Censored Alternating Renewal Processes

Enrique E. Alvarez

University of Connecticut
215 Glenbrook Road CLAS 328
Storrs, CT 06269 - 4120, USA
(e-mail: ealvarez@stat.uconn.edu)

Abstract. Consider a process that jumps back and forth between two states, with random times spent in between. Suppose the durations of subsequent on and off states are i.i.d. and that the process has started far in the past, so it has achieved stationarity. We estimate the sojourn distributions through maximum likelihood when data consist of several realizations observed over windows of fixed length. For discrete and continuous time Markov chains, we also examine if there is any loss of efficiency when ignoring the stationarity structure in the estimation.
Keywords: Alternating renewal process, Asymptotic efficiency, Window censoring.

1 Introduction

Consider a machine which periodically fails, undergoes technical service, and is put to work again, so that the working and out-of-service times form an alternating renewal process (ARP). Suppose further that the machine was placed in service in the indefinite past, so that the process may be regarded as stationary. Our interest here is to estimate the distribution of the on and off times when several such processes are observed over a time interval, or when the same process is observed over several "well separated" windows.

Such alternating renewal processes have been taken as models for diverse phenomena such as system availability and reliability in engineering [Pham-Gia and Turkkan, 1999], or the behavior of healthy-sick cycles in actuarial and insurance mathematics [Ramsay, 1984]. They have also been of interest as building blocks for other processes where the cumulative count from many alternating renewal processes whose inter-arrival times have high or infinite variance can produce aggregate network traffic that exhibits long range dependence [Murad S. Taqqu and Sherman, 1997].

The present study is concerned with estimating the distribution of the time spent in each of the states with maximum likelihood methods, when the data consist of "windows" from several stationary ARPs.

2 Construction and stationarity of ARPs

Consider a set of pairs of positive random variables $\{(Z_*, Y_*), (Z_1, Y_1), \ldots\}$ with the property that the first pair $(Z_*, Y_*)\mathrm{sim}Q_0$ and it is independent from the remaining $(Z_i; Y_i) \overset{\mathrm{iid}}{\mathrm{sim}}\, Q$. That kind of arrangement constitutes an *alternating renewal sequence* with *inter-arrival times* $X_* = Z_* + Y_*$, $X_i := Z_i + Y_i$, and *renewal times* $S_0 := X_*$ and $S_n := S_0 + \sum_1^n X_i$ for $n > 0$.

Consider the counting process $N(t) := \sum_0^\infty I\{S_n \in [0; t]\}$ and in order to record the state of the process at each time, introduce $W(t) := I\{S_{N(t)-1} + Z_{N(t)} > t\}$, which is the *alternating renewal process* associated with the renewal sequence. Thus the distribution of $W := \{W(t), t \geq 0\}$ is determined by Q_0 and Q; call the process *pure* if $X_* \equiv 0$ or *delayed* otherwise. Think of the Z's and Y's denoting durations of *on* and *off* times respectively; and for identifiability assume throughout that $P(Z_i = 0) = P(Y_i = 0) = 0$ for all $i \in \mathbb{Z}^+$.

Note also that the initial random vector (Z_*, Y_*) can be thought of as resulting from an ordinary pair $(Z_0, Y_0)\mathrm{sim}Q$ through truncation, as

$$Z_* = (X_* - Y_0)^+ \text{ and } Y_* = X_* \wedge Y_0. \tag{1}$$

In particular, situations with $Z_* = 0$ correspond to paths beginning in the off-state.

In this study we are concerned not with pure but with delayed alternative renewal processes, the importance of which is that with an appropriate choice of Q_0 the process W is stationary, in a sense to be defined shortly. Figure 1 shows a typical sample path observed over the "window" of time $[0, T]$.

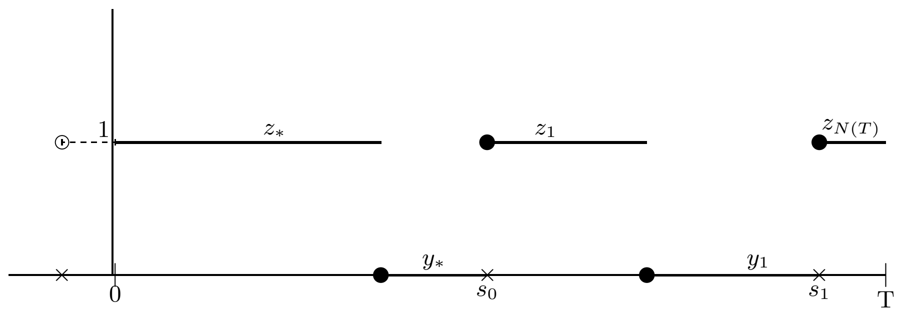

Fig. 1. A Sample Path from a Delayed ARP over $[0, T]$

2.0.0.1 Stationarity Choose any $t \in \mathbb{R}^+$ (deterministically or randomly but independent of the process) and construct a new alternating renewal sequence $\{(Z_i^t, Y_i^t), i \geq 0\}$ by censoring everything to the left of t. This is, the new sequence has an initial pair

$$Z_*^t = (S_{N(t)-1} + Z_{N(t)} - t)^+,$$
$$Y_*^t = Y_{N(t)} - (t - S_{N(t)-1} - Z_{N(t)})^+;$$

and subsequently $Z_i^t = Z_{N(t)+i}$ and $Y_i^t = Y_{N(t)+i}$, for $i \geq 1$. Notice that because this construction implies that $Z_*^t = 0$ on the event $C := \{S_{N(t)-1} + Z_{N(t)} \leq t\}$, the distribution of the random variable Z_*^t has a point mass at zero whenever C has positive probability.

Definition 1 *Call the ARP stationary if and only if the two sequences* $\{(Z_*, Y_*), (Z_i, Y_i), i \geq 1\}$ *and* $\{(Z_*^t, Y_*^t), (Z_i^t, Y_i^t), i \geq 1\}$ *are equal in distribution for every* $t \in [0, \infty)$.

Assume that $X := X_1$ has finite expectation μ_X and denote $Z := Z_1$, $Y := Y_1$.

Theorem 21 *If the distribution of the initial pair* (Z_*, Y_*) *is given by*

$$Q_0(z, y) = \frac{1}{\mu_X} E_Q \{(z \wedge Z)\, 1\,[Y \leq y] + (y \wedge Y)\}, \tag{2}$$

then process $\{W(t), t \geq 0\}$ *is stationary in the sense of definition 1.*

See [4]. In the special case when the on-time $Z\mathrm{sim}H$ is independent of the off-time $Y\mathrm{sim}G$ this gives

$$Q_0(z, y) = \frac{\mu_Y}{\mu_X} \int_0^y \frac{1 - G(u)}{\mu_Y} du + \frac{\mu_Z}{\mu_X} G(y) \int_0^z \frac{1 - H(u)}{\mu_Z} du. \tag{3}$$

3 A two-states Markov chain

The simplest example of a window censored alternating renewal process is a pair of consecutive observations from a Markov chain on {0,1}. When the transition probabilities are $\pi_0 := P(W_{t+1} = 1 | W_t = 0)$ and $\pi_1 := P(W_{t+1} = 1 | W_t = 1)$, the stationary distribution is given by

$$q := P\{W_t = 0\} = \tfrac{1-\pi_1}{1-\pi_1+\pi_0},\ p := P\{W_t = 1\} = \tfrac{\pi_0}{1-\pi_1+\pi_0}.$$

The joint density of a pair of consecutive observations is

$$P(W_t = x_i; W_{t+1} = y_i) = \frac{\pi_0(1-\pi_1)}{1-\pi_1+\pi_0} \left(\frac{\pi_1}{1-\pi_1}\right)^{x_i y_i} \left(\frac{1-\pi_0}{\pi_0}\right)^{(1-x_i)(1-y_i)}. \tag{4}$$

This is of exponential family form with complete sufficient statistic T, and canonical parameter η given respectively by

$$T = \begin{pmatrix} X_i Y_i \\ (1 - X_i)(1 - Y_i) \end{pmatrix} \text{ and } \eta = \begin{pmatrix} \ln \pi_1 - \ln(1 - \pi_1) \\ \ln(1 - \pi_0) - \ln \pi_0 \end{pmatrix}.$$

By standard results in exponential families theory [11], the maximum likelihood estimators are

$$\widehat{\pi_0} = \frac{\sum_{i=1}^n (X_i - Y_i)^2}{2n - \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i} \text{ and } \widehat{\pi_1} = \frac{2\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i};$$

and $\sqrt{n}(\hat{\pi} - \pi) \Rightarrow N(0; \hat{\Sigma})$, where

$$\hat{\Sigma} = \frac{1}{2}(1 - \pi_1 + \pi_0) \begin{pmatrix} \pi_0 (1 - \pi_0) \frac{1+\pi_0}{1-\pi_1} & -\pi_1 (1 - \pi_0) \\ -\pi_1 (1 - \pi_0) & \pi_1 \frac{2-\pi_1}{\pi_0} (1 - \pi_1) \end{pmatrix}.$$

Alternatively, we could ignore stationarity in order to estimate π_0 and π_1 by the sample proportion of transitions into each state, i.e.

$$\widetilde{\pi_0} = \frac{\sum_{i=1}^n (1 - X_i) Y_i}{\sum_{i=1}^n (1 - X_i)} \text{ and } \widetilde{\pi_1} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i}.$$

By the multivariate central limit theorem and the delta method, $\sqrt{n}(\tilde{\pi} - \pi) \Rightarrow N(0; \tilde{\Sigma})$, with

$$\tilde{\Sigma} = (1 - \pi_1 + \pi_0) \begin{pmatrix} \frac{\pi_0(1-\pi_0)}{1-\pi_1} & 0 \\ 0 & \frac{\pi_1(1-\pi_1)}{\pi_0} \end{pmatrix}.$$

At this point, it is natural to ask what is lost in terms of efficiency by ignoring stationarity in the estimation. To address this question, consider the difference matrix $\hat{\Sigma} - \tilde{\Sigma} =: (1 - \pi_1 + \pi_0)\ \Delta$. It is easy to check that the diagonal entries of Δ are strictly negative and that the cross-products are equal. Therefore, the matrix difference $(\hat{\Sigma} - \tilde{\Sigma})$ has one eigenvalue which is negative and the other is zero. This result is surprising, because it implies that there exist functions of the transition probabilities for which ignoring stationarity is of no consequence asymptotically. Essentially, any function of (π_0, π_1) with gradient proportional to the eigenvector corresponding to the null eigenvalue of Δ will have that property. This will be explored further for continuous time Markov chains in section 4.

4 A continuous time Markov chain

When the on and off times follow independent exponential distributions Z_i sim $Q_z = \exp(\lambda_1)$ and $Y_i \text{sim} Q_y = \exp(\lambda_2)$, the process $\{W(t), t \geq 0\}$ is a continuous time Markov chain. At any given time, the excess life is independent of the history of the process.

The stationary distribution is, according to equation (3):

$$Q_0(z, y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - e^{-z\lambda_1})\left(1 - e^{-\lambda_2 y}\right) + \frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - e^{-y\lambda_2}), \qquad (5)$$

with marginal distributions

$$Q_0(\infty, y) = (1 - e^{-y\lambda_2}) \text{ and } Q_0(z, \infty) = \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - e^{-z\lambda_1}) + \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Notice that Z_0 is independent of Y_0, since $Q_0(z,y) = Q_0(\infty,y)Q_0(z,\infty)$.

Reference [Alvarez, 2003] investigates how to obtain a likelihood for a sample path of an ARP observed on a window $[0,T]$, as a Radon-Nykodym derivative with respect to an appropriately chosen dominating measure and restricted to a filtration that corresponds to the censoring mechanism. The main result is that the window-censored likelihood ratio is a product of three types of factors:

i) In a typical sample path where at least one transition in observed, we multiply
  (a) the value of initial density
  (b) the values of the densities at all non-censored on and off times
  (c) the survival function for the duration of the last state in the window

ii) Secondly, if the window $[0,T]$ contains no jumps, the likelihood equals the survival function of the excess life in either state.

Using the above recipe, after some algebra we obtain the likelihood over a window $[0,T]$ as

$$l(T) = \frac{{\lambda_1}^{\tau+1\{W(T)=0\}}{\lambda_2}^{\tau+1\{W(0)=1\}}}{\lambda_1+\lambda_2}\exp\left[-\lambda_1 \mathrm{on}(T) - \lambda_2 \mathrm{off}(T)\right], \qquad (6)$$

where $\mathrm{on}(t) := \int_0^t W(t)dt =: t - \mathrm{off}(t)$. This additive property is characteristic to the Markov chain and it is fairly intuitive. Because of the memoryless property of the exponential distribution, the break up of the total on or off times into subperiods does not provide any additional information on their distribution. When we observe m windows independently up to a same time T, the log-likelihood over the sample is the sum of the corresponding path likelihoods.

4.1 Asymptotic normality

Following standard theorems in asymptotic statistics it is established that the likelihood equation has a unique root with probability tending to 1 as $m \to \infty$ and that $\sqrt{n}\left(\widehat{\lambda}_n - \lambda_0\right) \Rightarrow N(0,\hat{\Sigma})$ with

$$\hat{\Sigma} = \frac{(\lambda_1+\lambda_2)}{(\lambda_1 T + \lambda_2 T + 2)}\begin{pmatrix} \lambda_1\frac{\lambda_1 T+\lambda_2 T+1}{\lambda_2 T} & 1/T \\ 1/T & \lambda_2\frac{\lambda_1 T+\lambda_2 T+1}{\lambda_1 T}\end{pmatrix}.$$

Notice that while the main diagonal entries are $O(1/T)$, the off-diagonal entries are $O(1/T^2)$ as $T \to \infty$. This is intuitive, since the only reason why the estimators of λ_1 and λ_2 are dependent is the presence in the data of the initial (left censored) observations. As the observation window enlarges, the information provided by the first two observations becomes negligible and the estimators closer to being independent.

4.2 Comparison with classic estimators

As in the discrete Markov chain example of Section 3, it is natural to ask if there is any loss in efficiency by ignoring stationarity in the estimation.

Suppose that we "condition away" the initial states. That is, we seek a log-likelihood function conditioned on $\sigma\{Z_0 1(Z_0 > 0), Y_0 1(Z_0 = 0)\}$. This is given over a single window by

$$\ln l^c(T) = [\tau + r_1 + d_0 - 1](\ln \lambda_1) + \tau(\ln \lambda_2) - \lambda_1 \mathrm{on}(T) - \lambda_2 \mathrm{off}(T),$$

and its gradient is

$$\nabla \ln l^c(T) = \begin{pmatrix} (\tau + r_1 + d_0 - 1)/\lambda_1 - \mathrm{on}(T) \\ \tau/\lambda_2 - \mathrm{off}(T) \end{pmatrix}.$$

The conditional maximum likelihood estimators can be easily found over m windows to be

$$\widetilde{\lambda_1} = \tfrac{\tau + r_1 + d_0 - m}{\mathrm{on}(T)} \text{ and } \widetilde{\lambda_2} = \tfrac{\tau}{\mathrm{off}(T)}.$$

It is easy to check that

$$E\left[-\nabla^2 \ln l^c(T)\right]^{-1} = \frac{\lambda_1 + \lambda_2}{T} \begin{pmatrix} \frac{\lambda_1}{\lambda_2} & 0 \\ 0 & \frac{\lambda_2}{\lambda_1} \end{pmatrix}.$$

Therefore, $\sqrt{m}(\tilde{\lambda} - \lambda) \Rightarrow N(0; \tilde{\Sigma})$ with

$$\tilde{\Sigma} = \frac{\lambda_1 + \lambda_2}{T} \begin{pmatrix} \frac{\lambda_1}{\lambda_2} & 0 \\ 0 & \frac{\lambda_2}{\lambda_1} \end{pmatrix},$$

which coincides with the approximation for the unconditional m.l.e's for large T's. To compare the two methods asymptotically let

$$\hat{\Sigma} - \tilde{\Sigma} =: \tfrac{\lambda_1 + \lambda_2}{T} \tfrac{1}{\lambda_1 T + \lambda_2 T + 2} \Delta \;\text{ with }\; \Delta = \begin{pmatrix} -\frac{\lambda_1}{\lambda_2} & 1 \\ 1 & -\frac{\lambda_2}{\lambda_1} \end{pmatrix}.$$

As in the discrete chain, Δ is negative semidefinite since $\mathrm{tr}(\Delta) < 0$ and $|\Delta| = 0$. The m.l.e. is then better than its conditional version, with a gain in efficiency that depends inversely on the truncation time and which is also affected by the relative means of the on and off times.

On the other hand, Δ has eigenpairs

$$\left[0, (\lambda_2, \lambda_1)'\right] \text{ and } \left[\left(-\tfrac{\lambda_1}{\lambda_2} - \tfrac{\lambda_2}{\lambda_1}\right), (-\lambda_1, \lambda_2)'\right],$$

which can be used to decompose $\Delta = PDP'$, with

$$P = \frac{1}{\sqrt{\lambda_1^2 + \lambda_2^2}} \begin{pmatrix} \lambda_2 & -\lambda_1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \text{ and } D = \begin{pmatrix} 0 & 0 \\ 0 & -\frac{\lambda_1}{\lambda_2} - \frac{\lambda_2}{\lambda_1} \end{pmatrix}.$$

This suggests the definition of a new parameter $\eta = \eta(\lambda)$ by

$$\begin{pmatrix} \eta_1(\lambda_1, \lambda_2) \\ \eta_2(\lambda_1, \lambda_2) \end{pmatrix} := \begin{pmatrix} \lambda_1 \lambda_2 \\ \frac{1}{2}\lambda_2^2 - \frac{1}{2}\lambda_1^2 \end{pmatrix}.$$

This map is continuous and has the Jacobian matrix

$$D_\eta = \begin{pmatrix} \frac{\partial}{\partial \lambda_1}\eta_1(\lambda_1, \lambda_2) & \frac{\partial}{\partial \lambda_2}\eta_1(\lambda_1, \lambda_2) \\ \frac{\partial}{\partial \lambda_1}\eta_2(\lambda_1, \lambda_2) & \frac{\partial}{\partial \lambda_2}\eta_2(\lambda_1, \lambda_2) \end{pmatrix} = \begin{pmatrix} \lambda_2 & \lambda_1 \\ -\lambda_1 & \lambda_2 \end{pmatrix}.$$

By the delta method, the estimators $\widehat{\eta} = \eta\left(\widehat{\lambda}\right)$ and $\widetilde{\eta} = \eta\left(\widetilde{\lambda}\right)$ are asymptotically normal and the difference in covariance matrices is

$$D_\eta\left(\hat{\Sigma} - \tilde{\Sigma}\right)D_\eta' = \frac{1}{T}\frac{\lambda_1 + \lambda_2}{\lambda_2\lambda_1}\frac{1}{\lambda_1 T + \lambda_2 T + 2}\begin{pmatrix} 0 & 0 \\ 0 & -\left(\lambda_1^2 + \lambda_2^2\right)^2 \end{pmatrix}.$$

The product of the hazard rates is estimated equally efficiently by the two methods, asymptotically, but for estimation of the difference in the square of the hazard rates the unconditional m.l.e. is better. As before, the gain in efficiency depends inversely on the truncation time.

For the parameter $\eta_2(\lambda_1, \lambda_2) = \frac{1}{2}\lambda_2^2 - \frac{1}{2}\lambda_1^2$ the asymptotic relative efficiency (ARE) of $\tilde{\eta_2}$ w.r.t. $\hat{\eta_2}$ is given by

$$\text{A.R.E.}(\tilde{\eta_2}, \hat{\eta_2}) = 1 - \frac{(\lambda_1^2 + \lambda_2^2)^2}{2(\lambda_1^4 + \lambda_2^4)} \Big/ \left[1 + \frac{1}{2}(\lambda_1 + \lambda_2)T\right].$$

The fraction in the numerator varies between 0 when $\lambda_1 \to 0$ and 1 when $\lambda_1 = \lambda_2$. When T is small the gains in efficiency could be substantial. As an example, Table 1 quantifies these gains for a few combination of parameters values.

Case:	**i**	**ii**	**iii**	**iv**	**v**
λ_1	0.5	0.5	0.5	0.5	0.5
λ_2	1	1	0.5	0.5	0.5
T	4	20	2	1	0.5
A.R.E.$(\tilde{\eta_2}, \hat{\eta_2})$	**0.82**	**0.95**	**0.50**	**0.33**	**0.20**

Table 1. A.R.E. of $\tilde{\eta_2}$ w.r.t. $\hat{\eta_2}$

References

[Alvarez, 2003]Enrique E. Alvarez. *Likelihood Based Estimation of Stationary Semi-Markov Processes Under Window Censoring*. PhD thesis, Universtity of Michigan, 2003.

[Murad S. Taqqu and Sherman, 1997]Walter Willinger Murad S. Taqqu and Robert Sherman. Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review*, 2:5–23, 1997.

[Pham-Gia and Turkkan, 1999]T. Pham-Gia and N. Turkkan. System availability in a gamma alternating renewal process. *Naval Res. Logist.*, 46:822–844, 1999.

[Ramsay, 1984]Colin M. Ramsay. The asymptotic ruin problem when the healthy and sick periods form an alternating renewal process. *Insurance Mathematics and Economics*, 3:139–143, 1984.