MAXIMUM LIKELIHOOD ESTIMATION IN ALTERNATING RENEWAL PROCESSES UNDER WINDOW CENSORING

Enrique E. Alvarez - University of Connecticut
215 Glenbrook Rd, Storrs CT 06269 - 4120
ealvarez@merlot.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 stationary. 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 incurred when ignoring the stationarity structure in the estimation.

Keywords: Alternating renewal process; Asymptotic efficiency; Markov chain; Window censoring

Running headline: Alternating Renewal Processes

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 [1], or the behavior of healthy-sick cycles in actuarial and insurance mathematics [2]. 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 [3].

The literature also contains inference for ARP processes. Ref. [4] studies the estimation of the "point availability", i.e., the probability that a system modeled by an ARP is "on" at a given time t. It considers data arising from one (not necessarily stationary) ARP and the estimator that results from plugging empirical distributions into the so-called "availability functional".

Under a parametric context, [5] considers estimation in Markov processes with state space $\{0, 1\}$ under different censoring mechanisms. A single realization of the process is available, which is not assumed to be stationary. The choice of estimators is based on simplicity of computation and the possibility of recursive calculation, instead of maximum likelihood, but nevertheless his estimators turn out to be strongly consistent and asymptotically normal.

The present study is concerned with how to estimate the distribution of the time spent in each of the states with maximum likelihood, when the data consist of "windows" from several stationary ARPs. Estimation in window censored ARPs has received some attention in the literature. When stationarity is not assumed, [6] proposes estimators of Kaplan-Meier type and derives their asymptotic properties. Under stationarity, estimation of the inter-arrival time distribution F for "ordinary" renewal processes (as opposed to "alternating") under window censoring was addressed for the first time by Vardi [7], who proposes an algorithm to find the Non-Parametric Maximum Likelihood Estimator of F when data is collected over windows from processes with an *integer-valued* inter-arrival distribution. Later, [8] and [9] extended Vardi's work to non-arithmetic inter-arrival distributions and proved the strong consistency of the maximum likelihood estimators.

The research in references [7-9] could be extended to *alternating* renewal processes, but the resulting $\hat{F}$ would be a step function and if the distribution is known to be continuous with respect to the Lebesgue measure, the maximum likelihood method would fails to provide an answer for estimation of the density unless extra assumptions are incorporated. In this paper we follow a parametric treatment, assuming either geometric or exponential distributions for the duration of the on and off times.

The outline of this paper is as follows. Section 2 provides the necessary definitions for ARPs. In Section 3, an example on discrete time Markov chains addresses the estimation problem when a pair of consecutive points are observed from several chains. We compare "classic" estimators that ignore stationarity with the m.l.e. Interestingly, for some function of the transition probabilities ignoring stationarity in the estimation process results in no loss of efficiency asymptotically.

Section 4 aims at a general formula for the likelihood ratio when data are gathered continuously over an interval and the distributions of the on and off times belong to some non-parametric family. It is shown that the likelihood is a Radon-Nykodym derivative restricted to the σ-field that corresponds to the window censoring mechanism.

Section 5 considers a detailed example of an ARP with exponential on and off times, which results in a continuous time Markov chain. It is shown that under some data configurations the m.l.e. may fail to exist. However, the chance of such a data configuration decreases as the number of observed windows increases. Further, the estimators are asymptotically normal. As with the discrete chain of Section 3, a comparison is done with the maximum likelihood estimator conditioned on the initial state and its duration. Also in similarity with the discrete chain, we obtain that the full data m.l.e. is better in terms of asymptotic relative efficiency, except for some functions of the parameters in which both methods are equivalent. One such function is the product of the hazard rates. For other parameters, a numerical example shows that gains in asymptotic efficiency by considering the stationarity can be substantial.

To facilitate the reading, most to the proofs in the paper are deferred to the appendix.

2 CONSTRUCTION AND STATIONARITY OF ARPs

A sequence of i.i.d. pairs of positive random variables $\{(Z_1, Y_1), (Z_2, Y_2), \ldots\}$ with joint distribution $(Z_i; Y_i) \sim Q$ constitutes an *alternating renewal sequence* with *inter-arrival times* $X_i := Z_i + Y_i$, and *renewal times* $S_0 := 0$ and $S_n := \sum_1^n X_i$ for $n > 0$.

Consider the counting process $N(t) := \sum_1^\infty I\{S_n \in [0;t]\}$ for the number of completed on-off cycles until time t. In order to further record the state of the process at each time, introduce $W(t) := I\{S_{N(t)} + Z_{N(t)+1} > t\}$, which is the *pure alternating renewal process* associated with the renewal sequence.

In this study we allow the first pair of sojourn times to be distributed differently though independently from the rest. That is, $(Z_0; Y_0) \sim Q_0$ independent of the sequence $\{(Z_i, Y_i), i \geq 1\}$. The importance of this *delayed* case is that with an appropriate choice of Q_0 the process $W(\cdot)$ is stationary, in a sense to be defined shortly.

Figure 1 shows a typical sample path observed over the "window" of time $[0, T]$.

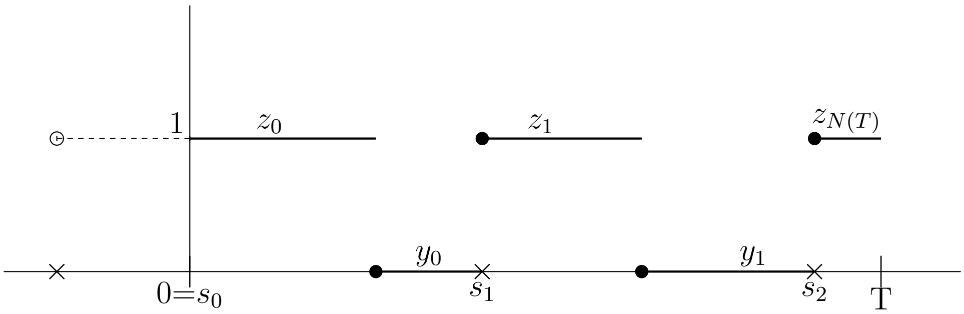

Figure 1: A Sample Path from a Delayed ARP over $[0, T]$

2.1 Stationarity

Choose any $t \in I\!R^+$ (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

$$\begin{aligned} Z_0^t &= (S_{N(t)-1} + Z_{N(t)} - t)^+, \\ Y_0^t &= Y_{N(t)} - (t - S_{N(t)-1} - Z_{N(t)-1})^+; \end{aligned}$$

and subsequently $Z_i^t = Z_{N(t)+i}$ and $Y_i^t = Y_{N(t)+i}$, for $i \geq 1$.

Definition 2.1. Call the ARP *stationary* iff the two sequences $\{(Z_i, Y_i), i \geq 0\}$ and $\{(Z_i^t, Y_i^t), i \geq 0\}$ are equal in distribution for every t.

Theorem 2.2. *The process* $\{W(t), t \geq 0\}$ *is stationary iff* $(Z_0, Y_0) \sim Q_0(z, y)$ *given by* $Q_0(z, y) = \mu_X^{-1} E_Q \{(z \wedge Z)\, 1[Y \leq y] + (y \wedge Y)\}$.

In the special case when the on-time $Z \sim H$ is independent of the off-time $Y \sim G$ this gives

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

which is the sum of: (i) the probability that the system starts off multiplied by the excess life distribution of the off times, and (ii) the product of the probability that the system starts on multiplied by the excess life distribution of the on times and the regular distribution of the off times.

2.2 Example: System Availability.

In engineering reliability, it is of interest the *availability* of a system, defined as the probability that the system is on at a given time. Since $N(t) < \infty$ w.p.1,

$$P\{W(t) = 1\} = P\{Z_0 > t\} + \sum_{n=1}^{\infty} P\{Z_n > t - S_{n-1}; 0 \leq t - S_{n-1} < Z_n + Y_n\}.$$

Conditioning on S_{n-1} and using the renewal measure $\nu = \sum_{n=0}^{\infty} F_{S_n}$ this entails

$$P\{W(t) = 1\} = 1 - H_0(t) + \int_0^t 1 - H(t - s)\nu(ds),$$

where $H_0(\cdot) = Q_0(\cdot, \infty)$ and $H(\cdot) = Q(\cdot, \infty)$. Further, since under stationarity

$$\nu(ds) = \tfrac{ds}{\mu_X} \quad \text{and} \quad H_0(z) = \tfrac{\mu_Y}{\mu_X} + \tfrac{\mu_Z}{\mu_X} \int_0^z \tfrac{1-H(u)}{\mu_Z} du,$$

the availability becomes $P\{W(t) = 1\} = \mu_Z/\mu_X$, which coincides with the asymptotic availability as $t \to \infty$.

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

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

The joint density of a pair of consecutive observations is

$$P(W_t = x_i; W_{t+1} = y_i) = \frac{\pi_0(1-\pi_1)}{1-\pi_1+\pi_0}\left(\frac{\pi_1}{1-\pi_1}\right)^{x_i y_i}\left(\frac{1-\pi_0}{\pi_0}\right)^{(1-x_i)(1-y_i)} . \quad (2)$$

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

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

By standard results in exponential families theory [10], the maximum likelihood estimators are

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

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

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

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.

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

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

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

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 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 5.

4 LIKELIHOOD RATIOS

This section investigates how to obtain a likelihood for a sample path of an ARP observed on a window $[0, T]$. As noted in Van der Vaart [11], work in nonparametric statistics often uses "*likelihoods that work*", without clearly making explicit with respect to which measures and in which σ-fields, if these even exist. Here we rigorously produce a Radon-Nykodym derivative with respect to an appropriately chosen dominating measure and restricted to a filtration that corresponds to the censoring mechanism.

It is natural to view alternating renewal sequences as elements in

$$\Omega := \{(z_0, y_0, z_1, y_1, ...) : z_0 \in [0, \infty) \text{ and } y_0, z_i, y_i \in (0, \infty); i > 0\},$$

(where the first z_0 is allowed to take the value 0 so as to make it possible to start observing a process when it is in the off state). To introduce two competing measures, ν and φ, in Ω we consider the coordinate random variables $\{(Z_i, Y_i),\ i \geq 0\}$.

Assume that under one formulation $(Z_0, Y_0) \sim Q_{\nu_0}$ and $(Z_i, Y_i) \overset{\text{i.i.d}}{\sim} Q_\nu$, while under a competitive formulation, $(Z_0, Y_0) \sim Q_{\varphi_0}$ and $(Z_i, Y_i) \overset{\text{i.i.d}}{\sim} Q_\varphi$. With these, define

$$\nu = Q_{\nu_0} \times \textstyle\prod_{i=1}^{\infty} Q_\nu \quad \text{and} \quad \varphi = Q_{\varphi_0} \times \textstyle\prod_{i=1}^{\infty} Q_\varphi.$$

Further let $\mathcal{F}_n := \sigma\{(Z_0; Y_0; Z_1; Y_1; \ldots; Z_n; Y_n)\}$ be the natural filtration and introduce dominating measures χ_0 and χ chosen so that $Q_{\nu_0}, Q_{\varphi_0} \ll \chi_0$ and

$Q_\nu, Q_\varphi \ll \chi$. In practice, for χ_0 and χ we take Lebesgue or counting measure whenever possible or otherwise we let $\chi_0 = Q_{\nu_0} + Q_{\varphi_0}$ and $\chi = Q_\nu + Q_\varphi$.

Denote by $l_n(w)$ the $\mathcal{F}_n$-restricted likelihood ratio

$$l_n(w) := \frac{d\nu|_{\mathcal{F}_n}(w)}{d\varphi|_{\mathcal{F}_n}(w)} = \frac{q_{\nu_0}(z_0, y_0)}{q_{\varphi_0}(z_0, y_0)} \prod_{i=1}^{n} \frac{q_\nu(z_i, y_i)}{q_\varphi(z_i, y_i)}, \tag{3}$$

where the q's (i.e. q_ν, q_φ, q_{ν_0} and q_{φ_0}) are Radon-Nykodym derivatives of the general and initial distributions under ν or φ with respect to χ_0 or χ.

The likelihood ratio in (3) would suffice for inference if the data consisted of a left censored on or off time, followed by a predetermined number n of pairs of non-truncated on and off times. However, when sample paths are window censored, two other complications arise:

1. the duration of the last observed state is censored on the right, and
2. more importantly, a random number of pairs τ are observed.

To adapt to this observation process, we need to redefine the filtration and obtain the corresponding Radon-Nykodym derivatve. This treatment is fairly technical but relies on well known results in probability. The proof is deferred to the Appendix, but the main result is that the window-censored likelihood ratio $l^*_\tau(w)$ is a product of three types of factors:

1. In a typical sample path where at least one transition in observed, we multiply
   (a) the value of initial density
   $$\frac{q_{\nu_0}}{q_{\varphi_0}}(z_0, y_0),$$
   (b) the values of the densities at all non-censored on and off times
   $$\prod_{i=1}^{\tau-1} \frac{q_\nu}{q_\varphi}(z_i, y_i)$$
   (c) the survival function for the duration of the last state in the window
   $$\frac{1 - Q_\nu(T - s_{\tau-1}, \infty)}{1 - Q_\varphi(T - s_{\tau-1}, \infty)} \quad \text{or} \quad \frac{1 - Q_\nu(z_\tau, T - z_\tau + s_{\tau-1})}{1 - Q_\varphi(z_\tau, T - z_\tau + s_{\tau-1})},$$
   according as to whether the last state is on or off.
2. Secondly, if the window $[0, T]$ contains no jumps, the likelihood equals either
   $$\frac{1 - Q_{\nu_0}(T, \infty)}{1 - Q_{\varphi_0}(T, \infty)} \quad \text{or} \quad \frac{1 - Q_{\nu_0}(\infty, T)}{1 - Q_{\varphi_0}(\infty, T)}. \ .$$

5 A CONTINUOUS TIME MARKOV CHAIN (CTMC)

When the on and off times follow independent exponential distributions $Z_i \sim Q_z = \exp\ (\lambda_1)$ and $Y_i \sim Q_y = \exp(\lambda_2)$, 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 (1):

$$Q_0(z, y) = \frac{\lambda_2}{\lambda_1 + \lambda_2}(1 - e^{-z\lambda_1})\left(1 - e^{-\lambda_2 y}\right) + \frac{\lambda_1}{\lambda_1 + \lambda_2}(1 - e^{-y\lambda_2}), \tag{4}$$

with marginal distributions

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

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

After some algebra we obtain the likelihood over a window $[0, T]$ as

$$l(T) = \frac{{\lambda_1}^{\tau+1\{W(T)=0\}}{\lambda_2}^{\tau+1\{W(0)=1\}}}{\lambda_1 + \lambda_2} \exp\left[-\lambda_1 \text{on}(T) - \lambda_2 \text{off}(T)\right], \tag{5}$$

where $\text{on}(t) := \int_0^t W(t)dt =: 1 - \text{off}(t)$. 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

$$\begin{aligned}\ln l_m(T) = -m \ln(\lambda_1 + \lambda_2) + (\ln \lambda_1)\left[N(T, m) + d_0\right] \\ + (\ln \lambda_2)\left[N(T, m) + r_1\right] - \lambda_1 \text{on}(T, m) - \lambda_2 \text{off}(T, m),\end{aligned} \tag{6}$$

where $\text{on}(T, m) = \sum \text{on}_k(T)$ and $\text{off}(T, m) = \sum \text{off}_k(T)$ are the total on and off times overall m windows, $N(T, m)$ is the total number of jumps into state on, $d_0 = \sum 1\{W_k(T) = 0\}$ is the number of windows that end in state off, and $r_1 = \sum 1\{W_k(0) = 1\}$ is the number of windows that start in state on.

5.1 Data Configurations

Maximization of (6) depends on data configurations, for some of which the m.l.e. fails to exist.

Configuration I When only on-times are observed $d_0 = 0$, $r_1 = m$, $N(T, m) = 0$, $\text{on}(T, m) = Tm$, and $\text{off}(T, m) = 0$. So

$$\ln l_m(T) = -m \ln\left(1 + \frac{\lambda_1}{\lambda_2}\right) - \lambda_1 T m$$

increases monotonically for any sequence in which $\lambda_1 \to 0$, and $\lambda_1/\lambda_2 \to 0$.

Configuration II The case when only off-times are observed is similar taking any sequence in which $\lambda_2 \to 0$, and $\lambda_2/\lambda_1 \to 0$.

Configuration III Suppose we observed some on and some off-times, but no jump in any window. To analyze this case, reparameterize as

$$p := \frac{\lambda_2}{\lambda_1 + \lambda_2} =: 1 - q, \quad \mu := \lambda_1 + \lambda_2, \quad \text{and} \quad \hat{p} := \frac{r_1}{m} =: 1 - \hat{q}.$$

The log-likelihood

$$\ln l_m(T) = m \ln\left[(1-p)^{\widehat{q}}(p)^{\widehat{p}}\right] - mT\mu\left[(1-p)\widehat{p} + p\widehat{q}\right]$$

is maximized for $p = \hat{p}$ and it increases monotonically as $\mu \to 0$.

Configuration IV When in at least one window there are both on and off times, call the data configuration "regular". In this case, $[N(T, m) + d_0] \geq 1$, $[N(T, m) + r_1] \geq 1$, $\text{on}(T, m) > 0$, and $\text{off}(T, m) > 0$. The log-likelihood is bounded,

$$\begin{aligned}\ln l_m(T) \leq &-\mu\{\text{on}(T, m) \wedge \text{off}(T, m)\} \\ &+ \ln\left\{q^{[N(T,m)+d_0]} p^{[N(T,m)+r_1]}\right\} + \ln \mu[-m + 2N(T, m) + d_0 + r_1].\end{aligned}$$

Since $2N(T, m) + d_0 + r_1 < m$,

$$\limsup_{\mu\to\infty} l_m(T) = \limsup_{\mu\to 0} l_m(T) = -\infty.$$

Thus the existence of a maximum likelihood estimator in the interior of the parameter space follows thus by the continuity of $l_m(T)$, and it is found in closed form by calculus in the usual manner.

5.2 Asymptotic Normality

Following standard theorems in asymptotic statistics, it is established in the Appendix 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

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

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.

5.3 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 $\sigma\left\{Z_0 1(Z_0 > 0), Y_0 1(Z_0 = 0)\right\}$. 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

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

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

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

As in the discrete chain, Δ is negative semidefinite since $\operatorname{tr}(\Delta) < 0$ 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

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

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

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

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

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. $\hat{\eta_2}$

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.

Acknowledgement

This work forms part of the author's Ph.D. dissertation, presented to the Department of Statistics, University of Michigan (December, 2002). The author wishes to thank his advisor, Robert W. Keener, for superb guidance throughout the development of the thesis work.

APPENDIX

A Stationary Distribution

In order to prove Theorem 2.2 it is convenient to proceed by showing uniqueness first and then existence.

Consider the triplet (Z_i, Y_i, X_i) where Z_i and Y_i denote the actual duration of the ith on and off times, and X_i is the observed length of the ith cycle. This is, for the middle observations in the window, $X_i = Z_i + Y_i$, but for the first and last ones,

$X_i < Z_i + Y_i$ indicating the left or right censoring. In particular, what is observed at the initiation of the observation window is not the actual on-off times (Z_0, Y_0), but the *censored* (Z_0^*, Y_0^*) given by $Z_0^* := (X_0 - Y_0)^+$ and $Y_0^* := (Y_0 \wedge X_0)$ .We seek an answer to the question: *which distribution for the first triplet* (Z_0, Y_0, X_0)*, if any, makes the ARP stationary?* We show next that any such distribution should fulfill equation (7) and that gives uniqueness because that equation defines a *determining class* [12].

Theorem A.1. *When the process is stationary, then* $\mu_X = EX_i < \infty$ *and the initial distribution* (Z_0, Y_0, X_0) *is such that for all bounded continuous* $g : I\!R^{+2} \mapsto I\!R$*,*

$$E[g(Z_0, Y_0)1(X_0 > c)] = \frac{1}{\mu_X} E\left\{g(Z, Y)(X - c)^+\right\}. \tag{7}$$

Proof. Choose a (large) positive M, and introduce a random time at which we start to observe the process by $T_M \sim U(0, M)$, independent of the field $\mathcal{F} = \sigma\{Z_0, Y_0, Z_1, Y_1, \ldots\}$. Let $\eta_M := N(M)$ and $\tau_M := N(T_M)$. A quantity of interest is the "residual life" defined by the random variable $R_M := S_{\tau_M} - T_M$.

Consider the conditional expectation

$$E\{g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)|\mathcal{F}\} = \sum_{i=0}^{\eta} E\{g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)1(T_M \in I_i)|\mathcal{F}\} \tag{8}$$

where $I_i := (S_{i-1}, S_i]$ for $0 \le i \le \eta - 1$, $S_{-1} := 0$, and $I_\eta := (S_{\eta-1}, M]$. Note that $\tau_M = i$ iff $T_M \in I_i$. Since Z_i and Y_i are $\mathcal{F}$ measurable and T_M is independent of the process, (8) is in turn equal to

$$\begin{aligned}
&= \sum_{i=0}^{\eta-1} g(Z_i, Y_i) \frac{X_i}{M} \frac{(X_i - c)}{X_i} 1(X_i > c) + g(Z_\eta, Y_\eta) \frac{M - S_{\eta-1}}{M} 1(M - S_{\eta-1} > c) \\
&= \sum_{i=0}^{\eta} g(Z_i, Y_i) \frac{(X_i - c)^+}{M} + \frac{1}{M} g(Z_\eta, Y_\eta) \left[(M - S_{\eta-1})\, 1(M - S_{\eta-1} > c)\right] \\
&= \frac{\eta}{M} \frac{1}{\eta} \sum_{i=0}^{\eta} g(Z_i, Y_i)(X_i - c)^+ + \frac{1}{M} g(Z_\eta, Y_\eta) \left[(M - S_{\eta-1})\, 1(M - S_{\eta-1} > c)\right].
\end{aligned}$$

The second term converges to zero as $M \to \infty$. To analyze the first term, notice that as $M \to \infty$, $N(M) \to \infty$ a.s. Then by the strong law of large numbers,

$$\frac{1}{\eta} \sum_{i=0}^{\eta} g(Z_i, Y_i)(X_i - c)^+ \to Eg(Z, Y)(X - c)^+ \quad \text{a.s.}.$$

Also, since by the elementary renewal theorem $N(M)/M \to \mu_X^{-1}$ a.s.,

$$E\{g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)|\mathcal{F}\} \to \frac{1}{\mu_X} Eg(Z,Y)(X-c)^+ \quad \text{a.s.}$$

Now by smoothing,

$$E[g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)] = E\{E[g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)|\mathcal{F}]\},$$

and by the Dominated (Bounded) Convergence Theorem,

$$E[g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)] \to \frac{1}{\mu_X} Eg(Z,Y)(X-c)^+.$$

For the process to be stationary it should hold that

$$E[g(Z_{\tau_M}, Y_{\tau_M})1(R_M > c)] = E[g(Z_0, Y_0)1(X_0 > c)],$$

and so, the process is stationary only if

$$E[g(Z_0, Y_0)1(X_0 > c)] = \frac{1}{\mu_X} Eg(Z,Y)(X-c)^+ \tag{9}$$

for all bounded $g \in C(\mathbb{R}^{2+})$. $\diamondsuit$

Having characterized the stationary distribution above, we now exhibit one particular choice for which (9) holds.

Lemma A.2. *Let $U \sim U(0,1)$ be independent of $\sigma\{Z_1, Y_1, Z_2, Y_2, \ldots\}$, and suppose that Υ_0 satisfies*

$$E\ f(Z_0, Y_0, X_0) = \frac{1}{\mu_X}\ E\ Xf(Z, Y, XU) \tag{10}$$

for all measurable $f : \mathbb{R}^{+3} \to \mathbb{R}$. Then, Υ_0 satisfies (9).

Proof. Take $f(Z_0, Y_0, X_0) = g(Z_0, Y_0)1(X_0 > c)$, for $g \in C(\mathbb{R}^{+2})$. Then

$$\begin{aligned}
Ef(Z_0, Y_0, X_0) &= Eg(Z_0, Y_0)1(X_0 > c) \\
&= \frac{1}{\mu_X} EXg(Z,Y)1(XU > c) \\
&= \frac{1}{\mu_X} E\left\{E\left[Xg(Z,Y)\ 1\left(U > \frac{c}{X}\right)\Big|\ X\right]\right\} \\
&= \frac{1}{\mu_X} E\{g(Z,Y)(X-c)^+\},
\end{aligned}$$

which coincides with (9). $\diamondsuit$

It remains to show that the distribution defined in (10) does in fact give rise to a stationary process, in the sense of Definition 2.1.

Theorem A.3. *For* $(Z_0, Y_0, X_0) \sim \Upsilon_0$ *as defined in (10),* $(Z_i, Y_i) \sim Q$ *and* $X_i = Z_i + Y_i$ *for* $i > 0$*, the alternating renewal process* $\{W(t), t \geq 0\}$ *is stationary.*

Proof. With any $t \in I\!R$, let $\tau := N(t)$ and $R_t := S_\tau - t$. Take an arbitrary measurable function $f : I\!R^{+3} \to I\!R$ and decompose the expectation

$$E\ f(Z_\tau, Y_\tau, R_t) = E\ \{f(Z_\tau, Y_\tau, R_t); N(t) = 0\} + \sum_{i=1}^{\infty} E\{f(Z_\tau, Y_\tau, R_t); N(t) = i\}, \quad (11)$$

since N$(t) < \infty$ w.p.1; and observe that the second term in the display above is

$$\begin{aligned}
&= \sum_{i=1}^{\infty} E\ \{f(Z_i, Y_i, S_{i-1} + X_i - t); N(t) = i\} \\
&= \sum_{i=1}^{\infty} E\ \{f(Z_i, Y_i, S_{i-1} + X_i - t); S_{i-1} \leq t; S_i > t\} \\
&= \sum_{i=1}^{\infty} \int_{s<t} E\,\{f(Z_i, Y_i, S_{i-1} + X_i - t);\ X_i > a - S_{i-1} | S_{i-1} = s\}\, dF_{S_{i-1}}(s).
\end{aligned}$$

Next, introduce the renewal measure

$$\begin{aligned}
\nu([0, s]) &= E\{N(s)\} = E \sum_{n=1}^{\infty} 1\,\{S_{n-1} \in [0, s]\} \\
&= \sum_{n=1}^{\infty} P\,\{S_{n-1} \in [0, s]\} = \sum_{n=1}^{\infty} F_{X_0} * F_{X_1}^{*(n-1)}([0, s]),
\end{aligned}$$

so $\nu = \sum_{n=1}^{\infty} F_{X_0} * F_{X_1}^{*n-1}$, where $*$ denotes convolution. It is known that for a stationary process $\nu(ds) = ds/\mu_X$ [13]. Using this,

$$E\ \{f(Z_\tau, Y_\tau, R_t); N(t)\ \geq\ 1\}\ =\ \frac{1}{\mu_X} \iint\limits_A f(z, y, s + x - t)\Upsilon(dz, dy, dx)ds, \quad (12)$$

where $A = \{0 < s \leq t, x > t - s\}$. Also, making the change of variables $r = x + s - t$, $A = \{r > 0, r \leq x, r > x - t\}$ the expectation becomes

$$E\ \{f(Z_\tau, Y_\tau, R_t); N(t) \geq 1\} = \frac{1}{\mu_X} \iint\limits_A f(z, y, r)\Upsilon(dz, dy, dx)dr.$$

In a similar manner we transform the first term in the r.h.s. of (11),

$$E\{f(Z_\tau, Y_\tau, R_t); N(t) = 0\} = \frac{1}{\mu_X} \iint_B f(z, y, s-t)\Upsilon(dz, dy, dx)ds.$$

where $B = \{0 < s \leq x; s > t\}$. Replacing $s - t = r$ this set becomes $B = \{r > 0, r \leq x, r \leq x - t\}$ and the expectation is

$$E\ \{f(Z_\tau, Y_\tau, R_t); N(t) = 0\} = \frac{1}{\mu_X} \iint_B f(z, y, r)\Upsilon(dz, dy, dx)dr.$$

Finally, adding the two pieces we get

$$Ef(Z_\tau, Y_\tau, R_t) = \frac{1}{\mu_X} \iint_{0<r\leq x} f(z, y, r)\Upsilon(dz, dy, dx)dr = Ef(Z_0, Y_0, X_0).$$

Now define $\rho(z, y, r) := Eh(z, y, r, Z_1, Y_1, X_1, Z_2, Y_2, X_2, \ldots)$, for some integrable h, and let κ be a stopping time. Then with $\mathcal{F}_\kappa = \sigma\{Z_0, Y_0, X_0, , \ldots, Z_\kappa, Y_\kappa, X_\kappa\}$, the expectation $E[h(Z_\kappa, Y_\kappa, R_\kappa, Z_{\kappa+1}, Y_{\kappa+1}, X_{\kappa+1}, \ldots)|\mathcal{F}_\kappa] = \rho(Z_\kappa, Y_\kappa, R_\kappa)$, and by smoothing,

$$\begin{aligned} E[h(Z_\kappa, Y_\kappa, R_\kappa, Z_{\kappa+1}, Y_{\kappa+1}, X_{\kappa+1}, \ldots)] &= E\rho(Z_\kappa, Y_\kappa, R_\kappa) \\ &= E\rho(Z_0, Y_0, X_0) = E[h(Z_0, Y_0, X_0, Z_1, Y_1, X_1, \ldots)]. \end{aligned}$$

Hence the process is stationary. $\diamondsuit$

Since (Z_0, Y_0, X_0) gets observed only partially, we use now equation (10) to obtain the distribution of the *censored* initial on and off times (Z_0^*, Y_0^*) given by $Z_0^* := (X_0 - Y_0)^+$ and $Y_0^* := (Y_0 \wedge X_0)$. This is,

$$\begin{aligned} Q_0(z, y) &= E_Q\left[1\left(Z_0^* \leq z\right)1\left(Y_0^* \leq y\right)\right] \\ &= E_Q\left\{1\left[(X_0 - Y_0)^+ \leq z\right]1\left[(Y_0 \wedge X_0) \leq y\right]\right\} \\ &= \frac{1}{\mu_X}E_Q X\left\{1\left[(XU - Y)^+ \leq z\right]1\left[Y \wedge XU \leq y\right]\right\} \\ &= \frac{1}{\mu_X}E_Q X\left\{1\left[U \leq \frac{z+Y}{X}\right]1\left[Y \wedge XU \leq y\right]\right\}. \end{aligned}$$

Conditioning now on Z and Y we observe that

$$Q_0(z, y) = \frac{1}{\mu_X}E_Q X\, E_Q\left\{1\left[U \leq \frac{z+Y}{X}\right]1\left[Y \wedge XU \leq y\right]\Big| Z, Y\right\}.$$

Since further $1[Y \wedge XU \leq y] = 1 - 1[U > y/X]\, 1[Y > y]$, the argument inside the conditional expectation is

$$1\left[U \leq \frac{z+Y}{X}\right] 1[Y \wedge XU \leq y] = 1\left[U \leq \frac{z+Y}{X}\right] - 1\left[\frac{y}{X} < U \leq \frac{z+Y}{X}\right] 1[Y > y].$$

Then $Q_0(z, y)$ equals

$$\begin{aligned}
&= \frac{1}{\mu_X} E_Q X\, E_Q \left\{ 1\left[U \leq \frac{z+Y}{X}\right] 1[Y \wedge XU \leq y] \Big| Z, Y \right\} \\
&= \frac{1}{\mu_X} E_Q X \left\{ \left(\frac{z+Y}{X} \wedge 1\right) - \left[\left(\frac{z+Y}{X} \wedge 1\right) - \frac{y}{X}\right] 1[Y > y] \right\} \\
&= \frac{1}{\mu_X} E_Q \left\{ (z \wedge Z)\, 1[Y \leq y] + y \wedge Y \right\}.
\end{aligned}$$

B Likelihood Ratios

In what follows, we arrive at the relevant likelihood ratio by building up from a few steps.

Censored paths but with a deterministic number of transitions n

- In order to incorporate the right censoring introduce a new filtration $\{\mathcal{F}_n^*, n \geq 0\}$ defined separately for $n = 0$ and $n \geq 1$ by

$$\begin{aligned}
\mathcal{F}_0^* : &= \sigma\{Z_0 \wedge T, (Z_0 + Y_0) \wedge T\}, \\
\mathcal{F}_n^* : &= \sigma\{Z_0, Y_0, Z_1, Y_1, \ldots, , Z_{n-1}, Y_{n-1}, \\
&\qquad Z_n 1[Z_n \leq T - S_{n-1}], 1[Y_n + Z_n + S_{n-1} \leq T]\}.
\end{aligned}$$

- In the next few steps, we derive an $\mathcal{F}_n^*$ restricted likelihood ratio. First we consider the case when $n \geq 1$. The case $n = 0$ represents windows that contain only censored observations and will be treated later.

  Assume $n \geq 1$ and let

$$\begin{aligned}
U_n &:= (Z_0; Y_0; Z_1; Y_1; \ldots; Z_{n-1}; Y_{n-1})' \\
W_n &= \begin{pmatrix} W_{1,n} \\ W_{2,n} \end{pmatrix} := \begin{pmatrix} Z_n 1[Z_n \leq T - S_{n-1}] \\ 1[Y_n + Z_n + S_{n-1} \leq T] \end{pmatrix},
\end{aligned}$$

  Notice that $W_{1,n} = 0$ implies that $W_{2,n} = 0$. For simplicity of notation we write U and W, omitting the subindex n in the sequel.

- For an arbitrary bounded measurable function h, consider

$$\begin{aligned} E_\nu h(U,W) &= E_\nu E_\nu \left[h(U,W)|U\right] \\ &= E_\nu E_\nu \left[h(U,W)1(W_1 = W_2 = 0)|U\right] \\ &\quad + E_\nu E_\nu \left[h(U,W)1(W_1 > 0; W_2 = 0)|U\right] \\ &\quad + E_\nu E_\nu \left[h(U,W)1(W_1 > 0; W_2 = 1)|U\right] \\ &= E_\nu\{(I) + (II) + (III)\}. \end{aligned} \tag{13}$$

In typical ARPs, except possibly for the first pair, the on and off times are independent. In our treatment, this is means $Q_\nu = Q_{\nu_z} \times Q_{\nu_y}$. This further enables a choice of $\chi = \chi_z \times \chi_y$ with $Q_{\nu_z} \ll \chi_z$ and $Q_{\nu_y} \ll \chi_z$. Similar decompositions hold under φ.

We now look at each of the conditional expectations on the rhs of equation (13) separately in order to express them as integrals. Let δ_x denote a point mass at x, and whenever its clear from the context we use the same notation for probability measures and cumulative distribution functions.

i) First,

$$\begin{aligned} (I) &= E_\nu \left[h(U,W)1(W_1 = W_2 = 0)|U = u\right] \\ &= h\left[u, \begin{pmatrix} 0 \\ 0 \end{pmatrix}\right] Q_{\nu z}[Z_n > T - s_{n-1}] \\ &= \int h\left[u, w\right]\left[1 - Q_{\nu z}(T - s_{n-1})\right]\delta_{(0,0)}(dw). \end{aligned}$$

ii) Since, S_{n-1} is a function of U, and $U \perp Z_n \perp Y_n$,

$$\begin{aligned} (II) &= E_\nu \left[h(U,W)1(W_1 > 0; W_2 = 0)|U = u\right] \\ &= E_\nu \left\{h\left[u, \begin{pmatrix} Z_n \\ 0 \end{pmatrix}\right] 1[Z_n \le T - s_{n-1}]1[Y_n + Z_n + s_{n-1} > T]\right\} \\ &= \int h\left[u, \begin{pmatrix} w_1 \\ 0 \end{pmatrix}\right] 1[w_1 \le T - s_{n-1}]\left[1 - Q_{\nu y}(T - w_1 - s_{n-1})\right] \\ &\qquad q_{\nu z}(w_1)\chi_z(dw_1)\delta_0(dw_2). \end{aligned}$$

iii) Finally,

$$\begin{aligned}
(III) &= E_\nu \left[h(U, W) 1(W_1 > 0; W_2 = 1) | U = u \right] \\
&= E_\nu \left\{ h \left[u, \begin{pmatrix} Z_n \\ 1 \end{pmatrix} \right] 1[Z_n \leq T - s_{n-1}] 1[Y_n + Z_n + s_{n-1} \leq T] \right\} \\
&= \int h \left[u, \begin{pmatrix} w_1 \\ 1 \end{pmatrix} \right] 1[w_1 \leq T - s_{n-1}] Q_{\nu y}(T - w_1 + s_{n-1}) \\
&\qquad\qquad q_{\nu z}(w_1) \chi_z(dw_1) \delta_1(dw_2).
\end{aligned}$$

Now, the expectation (13) can be expressed as a double integral with respect to the sum of the measures just obtained. So $E_\nu h(U, W)$ is equal to

$$\begin{aligned}
&\int_U \int_W h[u, w] \frac{d\nu_{n-1}}{d\chi^{n-1}}(u) \left\{ [1 - Q_{\nu z}(T - s_{n-1})] \, 1[z_n > T - s_{n-1}] \right. \\
&+ 1[z_n \leq T - s_{n-1}] 1[y_n + y_n + s_{n-1} > T] \left[1 - Q_{\nu y}(T - z_n + s_{n-1}) \right] q_{\nu z}(z_n) \\
&\quad \left. +1[z_n \leq T - s_{n-1}] 1[y_n + z_n + s_{n-1} \leq T] Q_{\nu y}(T - z_n + s_{n-1}) q_{\nu z}(z_n) \right\} \\
&\qquad\qquad [\chi^{n-1} \times (\delta_{(0,0)} + \chi_z \times \delta_0 + \chi_z \times \delta_1)](du, dw).
\end{aligned}$$

A similarly decomposition holds under φ. Thus, when $n \geq 1$, the $\mathcal{F}_n^*$-restricted likelihood ratio l_n^* is the product:

$$\begin{aligned}
l_n^* = l_{n-1} &\left[\frac{1 - Q_{\nu z}(T - s_{n-1})}{1 - Q_{\varphi z}(T - s_{n-1})} \right]^{1[z_n > T - s_{n-1}]} \\
&\left[\frac{[1 - Q_{\nu y}(T - z_n + s_{n-1})] \, q_{\nu z}(z_n)}{[1 - Q_{\varphi y}(T - z_n + s_{n-1})] \, q_{\varphi z}(z_n)} \right]^{1[z_n \leq T - s_{n-1}] 1[y_n + z_n + s_{n-1} > T]} \\
&\qquad \left[\frac{Q_{\nu y}(T - z_n + s_{n-1}) q_{\nu z}(z_n)}{Q_{\varphi y}(T - z_n + s_{n-1}) q_{\varphi z}(z_n)} \right]^{1[z_n \leq T - s_{n-1}] 1[y_n + z_n + s_{n-1} \leq T]} .
\end{aligned}$$

- For $n = 0$, similar calculations show that

$$\begin{aligned}
l_0^* = &\left[\frac{1 - Q_{\nu_{y_0}}(T)}{1 - Q_{\varphi_{y_0}}(T)} \right]^{1[z_0 = 0]} \left[\frac{1 - Q_{\nu_{y_0}}(T)}{1 - Q_{\varphi_{y_0}}(T)} \right]^{1[z_0 = 0] 1[y_0 > T]} \\
&\qquad \left[\frac{[1 - Q_{\nu_{y_0}}(T - z_0)] \, q_{\nu_{z_0}}(z_0)}{[1 - Q_{\varphi_{y_0}}(T - z_0)] \, q_{\varphi_{z_0}}(z_o)} \right]^{1[0 < z_0 < T] 1[y_0 + z_0 > T]} . \qquad (14)
\end{aligned}$$

Censored paths with a random number of transitions τ The field $\mathcal{F}^*_n$ is however still inadequate, for when we observe an alternating renewal process continually over a "window" $[0, T]$ the number of renewals in the interval is not predetermined but random. In the next steps we take this into account.

- First, define a random time $\tau = \inf\{n : S_n > T\}$. Notice that since by the SLLN $S_n \to +\infty$ a.s. under either ν or φ, it follows that $\tau < \infty$ a.s. (ν, φ). In fact τ is a stopping time with respect to $\mathcal{F}^*_n$. To see this observe that $\{\tau = 0\} = \{Z_0 + Y_0 > T\} \subset \mathcal{F}^*_0$; and when alternatively $n \geq 1$, the set $\{1 \leq \tau \leq n\}$ equals the union

  $$\bigcup_{k=1}^{n} \Big(\{Z_k 1[Z_k \leq T - S_{k-1}] = 0\} \cup \{1[Y_k + Z_k + S_{n-1} \leq T] = 0\} \Big),$$

  and this is a subset of $\mathcal{F}^*_n$, for all $n \geq 1$.

  Now we are ready to define a filtration that corresponds to window censored paths. Under such, the events that we observe are all in

  $$\mathcal{F}^*_\tau := \sigma\{A \in \mathcal{F} : \forall n \; A \cap \{\tau = n\} \in \mathcal{F}^*_n\}. \tag{15}$$

- The final task is to obtain an $\mathcal{F}^*_\tau$-restricted likelihood ratio. We proceed as in Wald's fundamental likelihood ratio identity. For $A \in \mathcal{F}^*_\tau$,

  $$\nu(A) = \nu(A; \tau < \infty) = \sum_{n=0}^{\infty} \nu(A; \tau = n).$$

  But by the definition of $\mathcal{F}^*_\tau$, we see that for all n the intersection $A \cap \{\tau = n\}$ lies in $\mathcal{F}^*_n$. Hence,

  $$\nu(A; \tau = n) = \int_\Omega 1_A 1\{\tau = n\} \; l^*_n \; d\varphi.$$

  So that,

  $$\nu(A) = \sum_{n=0}^{\infty} E_\varphi[l^*_n 1_A 1(\tau = n)] = \sum_{n=0}^{\infty} E_\varphi[l^*_\tau 1_A 1(\tau = n)].$$

By Fubini,

$$\nu(A) = E_\varphi \left[l_\tau^* 1_A \sum_{n=0}^{\infty} 1(\tau = n) \right] = E_\varphi \left[l_\tau^* 1_A 1(\tau < \infty) \right] = E_\varphi [l_\tau^* 1_A],$$

and hence finally:

$$\begin{aligned} l_\tau^*(w) &:= \left. \frac{d\upsilon}{d\varphi} \right|_{\mathcal{F}_\tau^*} (w) \\ &= l_0(w)\ 1(\tau = 0) + 1(\tau \geq 1)\ l_{\tau-1}(w) \\ &\quad \left[\frac{1 - Q_{\nu z}(T - s_{\tau-1})}{1 - Q_{\varphi z}(T - s_{\tau-1})} \right]^{1[z_\tau > T - s_{\tau-1}]} \\ &\quad \left[\frac{[1 - Q_{\nu y}(T - z_\tau + s_{\tau-1})]\, q_{\nu z}(z_\tau)}{[1 - Q_{\varphi y}(T - z_\tau + s_{\tau-1})]\, q_{\varphi z}(z_\tau)} \right]^{1[z_\tau \leq T - s_{\tau-1}]} . \end{aligned} \tag{16}$$

C Asymptotic Normality

We appeal to the following results in Van der Vaart [11, chapter 5].

Theorem C.1. *For each θ in an open subset of Euclidean space, let $\theta \mapsto \psi_\theta(w)$ be twice continuously differentiable for every w. Suppose that $E_{\theta_0}\psi_{\theta_0}(w) = 0$, that $E_{\theta_0} \left\| \psi_{\theta_0}(w) \right\|^2 < \infty$ and that the matrix $E_{\theta_0} \nabla \psi_{\theta_0}(w)$ exists and is nonsingular. Assume that the second order partial derivatives are dominated by a fixed integrable function $\ddot{\psi}(w)$ for every θ in a neighborhood of θ_0.*

Then, every consistent estimator sequence $\widehat{\theta}_n$ such that $\Psi_n(\widehat{\theta}_n) := \frac{1}{n} \sum_{i=1}^n \psi_{\widehat{\theta}_n}(w_i) = 0$ for every n satisfies:

$$\sqrt{n} \left(\widehat{\theta}_n - \theta_0 \right) = - \left[E_{\theta_0} \nabla \psi_{\theta_0}(w) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta_0}(w_i) + o_p(1),$$

and in particular, the sequence $\sqrt{n} \left(\widehat{\theta}_n - \theta_0 \right)$ is asymptotically normal with mean 0 and variance

$$\left[E_{\theta_0} \nabla \psi_{\theta_0}(w) \right]^{-1} E_{\theta_0} \left[\psi_{\theta_0}(w) \psi'_{\theta_0}(w) \right] \left[E_{\theta_0} \nabla \psi_{\theta_0}(w) \right]^{-1}.$$

Theorem C.2. *Under the conditions of the preceding theorem, the probability that the equation $\frac{1}{n} \sum_{i=1}^n \psi_\theta(w_i) = 0$ has at least one root tends to 1, as $n \to \infty$ and there exists a sequence $\widehat{\theta}_n$ such that $\widehat{\theta}_n \to \theta_0$ in probability. If $\psi_\theta = \nabla m_\theta$ is the gradient of some function m_θ then and θ_0 is a point of local maximum of $\theta \mapsto E_{\theta_0} m_\theta$, then the sequence $\widehat{\theta}_n$ can be chosen to be local maxima of the maps $\theta \mapsto \frac{1}{n} \sum_{i=1}^n m_\theta(w_i)$.*

We now check that the assumptions of these theorems are satisfied in our context. We have m i.i.d. "windows" w_i on $[0,T]$. These random elements $w_i = \{W_i(s) : 0 \le s \le T\}$ represent the observed behavior of an exponential alternating renewal process. The parameter $\theta = (\lambda_1, \lambda_2)$ lies in the space $\theta \in \Theta = (0,\infty)^2$ and each observation has density

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

as in (5). Let $m_\theta = \ln f_\theta(w)$, with the corresponding $\psi_\theta(w) = \nabla \ln f_\theta(w)$ given by

$$\psi_\theta(w) = \begin{pmatrix} -\mathrm{on}(T) - \dfrac{1}{\lambda_1+\lambda_2} + \dfrac{N(T)+d_0}{\lambda_1} \\ \\ -\mathrm{off}(T) - \dfrac{1}{\lambda_1+\lambda_2} + \dfrac{N(T)+r_1}{\lambda_2} \end{pmatrix}.$$

$\psi_\theta(w)$ is continuous and differentiable in Θ. Note that

$$E_\theta \mathrm{on}(T) = \frac{\lambda_2}{\lambda_1+\lambda_2}T, \qquad E_\theta \mathrm{off}(T) = \frac{\lambda_1}{\lambda_1+\lambda_2}T,$$

$$E_\theta N(T) = \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}T,$$

$$E_\theta d_0 = \frac{\lambda_1}{\lambda_1+\lambda_2}, \qquad E_\theta r_1 = \frac{\lambda_2}{\lambda_1+\lambda_2}.$$

which imply $E_{\theta_0}\psi_{\theta_0}(w) = 0$. The densities $f_\theta(w)$ form a curved exponential family and the canonical statistics have a finite moment generating function. This justifies the reversal of the order of the operations of derivation and expectation, which gives that $E_\theta \ln f_\theta(w) = 0$, and $E_\theta\left[\nabla \ln f_\theta(w) \nabla' \ln f_\theta(w)\right] = -E_\theta\left[\nabla^2 \ln f_\theta(w)\right]$;see Brown (1986) for details. Explicitly,

$$\nabla^2 \ln f_\theta(w) = \begin{pmatrix} \frac{1}{(\lambda_1+\lambda_2)^2} - \frac{N(T)+d_0}{\lambda_1^2} & \frac{1}{(\lambda_1+\lambda_2)^2} \\ \frac{1}{(\lambda_1+\lambda_2)^2} & \frac{1}{(\lambda_1+\lambda_2)^2} - \frac{N(T)+r_1}{\lambda_2^2} \end{pmatrix}.$$

Negating and taking an expectation yields the Fisher Information matrix

$$I(\theta) = \begin{pmatrix} \frac{1}{\lambda_1+\lambda_2}\frac{1}{\lambda_1}(\lambda_2 T+1) - \frac{1}{(\lambda_1+\lambda_2)^2} & -\frac{1}{(\lambda_1+\lambda_2)^2} \\ -\frac{1}{(\lambda_1+\lambda_2)^2} & \frac{1}{\lambda_1+\lambda_2}\frac{1}{\lambda_2}(\lambda_1 T+1) - \frac{1}{(\lambda_1+\lambda_2)^2} \end{pmatrix}.$$

Also, $E_{\theta_0}\left\|\psi_{\theta_0}(w)\right\|^2 = E_{\theta_0}\left[\psi'_{\theta_0}(w)\psi_{\theta_0}(w)\right] = \mathrm{tr}\{-E_\theta\left[\nabla^2 \ln f_\theta(w)\right]\} < \infty$, and

$$E_{\theta_0}\left[\nabla\psi_{\theta_0}(w)\right] = \begin{pmatrix} \frac{1}{\lambda_1+\lambda_2}\frac{1}{\lambda_1}(\lambda_2 T+1) - \frac{1}{(\lambda_1+\lambda_2)^2} & -\frac{1}{(\lambda_1+\lambda_2)^2} \\ -\frac{1}{(\lambda_1+\lambda_2)^2} & \frac{1}{\lambda_1+\lambda_2}\frac{1}{\lambda_2}(\lambda_1 T+1) - \frac{1}{(\lambda_1+\lambda_2)^2} \end{pmatrix}$$

exists and is positive definite. To see this, check that $\operatorname{tr}\{E_{\theta_0}[\nabla\psi_{\theta_0}(w)]\} > 0$ and $\det\{E_{\theta_0}[\nabla\psi_{\theta_0}(w)]\} > 0$.

The second derivatives of ψ_θ are, using the notation $\ddot{\psi}_{ijk} := \dfrac{\partial^2\psi_\theta(w)_i}{\partial\theta_j\partial\theta_k}$,

$$\begin{aligned}
\ddot{\psi}_{111} &= 2\frac{N(T)+d_0}{\lambda_1^3} - 2\frac{1}{(\lambda_1+\lambda_2)^3} \\
\ddot{\psi}_{112} &= \ddot{\psi}_{121} = \ddot{\psi}_{122} = \ddot{\psi}_{211} = \ddot{\psi}_{212} = \ddot{\psi}_{221} = -2\frac{1}{(\lambda_1+\lambda_2)^3} \\
\ddot{\psi}_{222} &= 2\frac{N(T)+r_1}{\lambda_2^3} - 2\frac{1}{(\lambda_1+\lambda_2)^3},
\end{aligned}$$

and we can see that

$$\left|\ddot{\psi}_{ijk}\right| \leq 2\frac{N(T)+d_0}{(\lambda_{0_1}-\varepsilon)^3} + 2\frac{N(T)+r_1}{(\lambda_{0_2}-\varepsilon)^3} + 2\frac{1}{(\lambda_{0_1}+\lambda_{0_2}-2\varepsilon)^3}$$

in every (sufficiently small) ε-nbhd of $\theta_0 \in \Theta$. The bound is integrable since so are $N(T), d_0$, and r_1. The conditions of the theorems are thus satisfied.

References

1. Pham-Gia, T.; Turkkan, N. System availability in a gamma alternating renewal process. Naval Res. Logist. **1999**, *46*, 822-844.

2. Ramsay, C. The asymptotic ruin problem when the healthy and sick periods form an alternating renewal process. Insurance Mathematics and Economics **1984**, *3*, 139-143.

3. Willinger, W.; Taqqu, M. S.; Sherman, R. Proof of a fundamental result in self-similar traffic modelling. Computer Communication Review **1997**, *2*, 5-23.

4. Baxter, L. A.; Li, L. Non-parametric confidence intervals for the renewal function and the point availability. Scand. J. Statist. **1997**, *21*, 277-287.

5. Karr, A. F. Estimation and reconstruction for zero-one Markov processes. Stochastic Process. Appl. **1994**, *16*, 219-255.

6. Ouhbi, B.; Limnios, N. Non-parametric Estimation for semi-Markov kernels with application to reliability analysis. Applied Stochastic Models and Data Analysis **1996**, *12*, 209-220.

7. Vardi, Y. Nonparametric estimation in renewal processes. Ann. Statist. **1982**, *10*, 772-785.

8. Soon, G. *Large sample theory of empirical distributions in a window censoring model for renewal processes*; Ph.D. Thesis: The University of Michigan, Ann Arbor, 1995.

9. Soon, G.; Woodroofe, M. Nonparametric estimation and consistency for renewal processes. J. Statist. Plann. Inference **1996**, *53*, 171-195.

10. Brown, L. *Fundamentals of statistical exponential families*; Lecture notes monograph series, Institute of Mathematical Statistics, 1986.

11. Van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press, 1998.

12. Billingsley, P. *Convergence of probability measures*; Wiley Interscience, 1968.

13. Resnick, S. *Adventures in Stochastic Processes*; Birkhauser: Boston, 1992.