KYBERNETIKA — VOLUME *52* (2016), NUMBER 3, PAGES 348–358

A VERSATILE SCHEME FOR PREDICTING RENEWAL TIMES

Gusztáv Morvai and Benjamin Weiss

There are two kinds of universal schemes for estimating residual waiting times, those where the error tends to zero almost surely and those where the error tends to zero in some integral norm. Usually these schemes are different because different methods are used to prove their consistency. In this note we will give a single scheme where the average error is eventually small for all time instants, while the error itself tends to zero along a sequence of stopping times of density one.

Keywords: nonparametric estimation, stationary processes

Classification: 62G05, 60G25,60G10

1. INTRODUCTION

In problems of universal estimation it is usually easier to prove that a certain scheme converges in probability as opposed to a pointwise result. On the other hand for particular estimation schemes it is not always so easy to go from a pointwise result to convergence in expectation since one needs to control the size of the rare errors that may occur. In this note we are interested in the problem of estimating the conditional expectation of the residual waiting time (given the first n outputs) to the next occurrence of the renewal state in a binary renewal process. Since the possible values of this residual waiting time are not bounded we don't expect to be able to give a reasonable estimate at all time instants and to obtain a positive results. One of the things that we will do here is to show by an explicit construction that indeed this is the case.

In order to obtain positive results the notion of intermittent estimation was introduced (cf. [8]). Here one defines a sequence of stopping times and only ventures an estimate at those times. In the favorable cases these times have density one so that effectively we are not giving up too much. We have already given such a scheme in [8]. Our purpose in this note is to show how to modify that scheme so that we also get convergence to zero of the expected value of the error that we will be making. For this we will venture a guess at all time instants, and what we propose in fact is a single scheme which will converge in expectation for all time instants and converge almost surely along a sequence of stopping times of density one.

DOI: 10.14736/kyb-2016-3-0348

We turn now to a more formal description of the problem. It is easiest to formally define a renewal process in terms of an underlying Markov chain. Consider a Markov chain on the state space $\{0, 1, 2, \dots\}$ with transition probabilities $p_{i,i-1} = 1$ for all $i \geq 1$ and $p_{0,i} = p_i$ a probability distribution π on $\{0, 1, 2 \dots\}$, cf. [5] Ex. 12.13. This chain is positive recurrent exactly when $\sum_{i=0}^{\infty} i p_{0,i} = \mu < \infty$ and the unique stationary probability assigns mass $\frac{1}{1+\mu}$ to the state 0, cf. [4] Ch. XIII and [13] Sec. I.2.c. Collapsing all states $i \geq 1$ to 1 gives rise to the classical binary renewal process. Even though our primary interest is in one sided processes, stationarity implies that there exists a two sided process with the same statistics and we will use the two sided version whenever it is convenient to do so.

For conciseness sake, we will denote $X_i^j = (X_i, \dots, X_j)$ and also use this notation for $i = -\infty$ and $j = \infty$. Our interest is in the waiting time to renewal (the state 0) given some previous observations, in particular given X_0^n. Recall that if the data segment X_0^n doesn't contain a zero the expected time to the first occurrence of a zero may be infinite; this depends on the finiteness of the second moment of π. If a zero occurs then the expected time depends on the location of the zero and so we introduce the notation:

$$\tau(X_{-\infty}^n) = \text{the } t \geq 0 \text{ such that } X_{n-t} = 0, \text{ and } X_i = 1 \text{ for } n - t < i \leq n.$$

Note that this is well defined with probability one. If a zero occurs in X_0^n then $\tau(X_{-\infty}^n)$ depends only on X_0^n and so we will also write for $\tau(X_{-\infty}^n)$, $\tau(X_0^n)$ with the understanding that this is defined only if a zero occurs in X_0^n.

Define σ_i as the length of runs of 1's starting at position i. Formally put

$$\sigma_i = \max\{0 \leq l : X_j = 1 \ \text{ for } \ i < j \leq i + l\}.$$

Now for the classical binary renewal process $\{X_n\}$ define θ_n as

$$\theta_n = E(\sigma_n | X_0^n).$$

(Note that

$$\theta_n = \frac{\sum_{k=0}^{\infty} k p_{k+\tau(X_0,\dots,X_n)}}{\sum_{k=\tau(X_0,\dots,X_n)}^{\infty} p_k}$$

as soon as there is at least one zero in X_0^n. As we have already mentioned if no zero occurs then it might happen that $\theta_n = \infty$.) Our goal is to estimate θ_n without prior knowledge of the distribution function of the process. We will measure how well we do either by a direct comparison or by taking the expectation of the difference between θ_n and our estimate for it.

An intermittent scheme can be converted into a scheme defined for all n by simply setting the estimate to be zero at those time instants which are not a stopping time. In the next section we will give a general result showing under what hypotheses this will yield a two way universal scheme, while in §3 we will give a specific scheme which works. The last section has the construction that shows why we must use intermittent estimation to get almost sure results.

For further reading on these topics we refer the interested reader to [9, 12] and [11].

2. CONVERTING INTERMITTENT SCHEMES

Define ψ as the position of the first zero, that is,

$$\psi = \min\{t \geq 0 : X_t = 0\}.$$

Let

$$\mu_L = \sum_{i=L}^{\infty}(i-L)p_i / \sum_{i=L}^{\infty} p_i.$$

For a sequence of stopping times λ_n define χ_n as

$$\chi_n = \min\{k \geq 0 : \lambda_k = n\}$$

and ∞ if there is no such finite k.

Let

$$h_n^*(X_0^n) = \begin{cases} h_{\chi_n}(X_0^{\lambda_{\chi_n}}) & \text{if } \chi_n < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.1. Assume $\sum_{k=0}^{\infty} k^2 p_k < \infty$. Let λ_n be a sequence of stopping times and $h_n(X_0^{\lambda_n})$ an estimation scheme such that

$$\text{almost surely, } \tau(X_{-\infty}^{\lambda_n}) \leq \lambda_n, \tag{1}$$

$$\text{for any fixed } L, \quad \lim_{n\to\infty} E\left(I_{\{\tau(X_{-\infty}^n)\leq L\}} I_{\{\forall k\geq 0:\lambda_k\neq n\}} \mu_{\tau(X_{-\infty}^n)}\right) = 0, \tag{2}$$

$$\text{almost surely, } \lim_{n\to\infty} |h_n(X_0^{\lambda_n}) - \mu_{\tau(X_{-\infty}^{\lambda_n})}| = 0, \tag{3}$$

$$\lim_{n\to\infty} E\left(I_{\{\exists k\geq 0:\lambda_k=n\}} \left|h_{\chi_n}(X_0^{\lambda_{\chi_n}}) - \mu_{\tau(X_{-\infty}^n)}\right|\right) = 0. \tag{4}$$

Then

- almost surely, $\lim_{n\to\infty} I_{\{\chi_n<\infty\}} |h_n^*(X_0^n) - \mu_{\tau(X_{-\infty}^n)}| = 0$
- $\lim_{n\to\infty} E\left(|h_n^*(X_0^n) - \theta_n|\right) = 0.$

P r o o f. For any L we can estimate:

$$\begin{aligned}
&\lim_{n\to\infty} E\left(|h_n^*(X_0^n) - \theta_n|\right) \\
&\quad\leq \lim_{n\to\infty} E\left(I_{\{\exists k\geq 0:\lambda_k=n\}} \left|h_{\chi_n}(X_0^{\lambda_{\chi_n}}) - \mu_{\tau(X_{-\infty}^n)}\right|\right) \\
&\quad+ \lim_{n\to\infty} E\left(I_{\{\tau(X_{-\infty}^n)\leq L\}} I_{\{\forall k\geq 0:\lambda_k\neq n\}} \mu_{\tau(X_{-\infty}^n)}\right) \\
&\quad+ \lim_{n\to\infty} E\left(I_{\{L<\tau(X_{-\infty}^n)\leq n\}} \mu_{\tau(X_{-\infty}^n)}\right) \\
&\quad+ \lim_{n\to\infty} E\left(I_{\{\tau(X_{-\infty}^n)>n\}} \theta_n\right).
\end{aligned}$$

The first two terms are zero by (4) and (2). We deal with the third term. Now

$$P(\tau(X_{-\infty}^{k}) = l) = \frac{1}{1 + \sum_{h=0}^{\infty} h p_h} \sum_{h=l}^{\infty} p_h$$

(note that by Kac's theorem

$$P(X_{k-l} = 0) = \frac{1}{1 + \sum_{h=0}^{\infty} h p_h}$$

cf. [4] Ch. XIII and [13] Sec. I.2.c). It is easy to see that

$$\begin{aligned} \sum_{l=0}^{\infty} \mu_l P(\tau(X_{-\infty}^{0}) = l) &= \sum_{l=0}^{\infty} \frac{\sum_{h=0}^{\infty} h p_{h+l}}{\sum_{h=l}^{\infty} p_h} \frac{\sum_{h=l}^{\infty} p_h}{1 + \sum_{h=0}^{\infty} h p_h} \\ &\leq \frac{\sum_{h=0}^{\infty} h^2 p_h}{1 + \sum_{h=0}^{\infty} h p_h} \\ &< \infty. \end{aligned}$$

For any $\epsilon > 0$ we can choose L so that

$$\begin{aligned} \lim_{n\to\infty} E\left(I_{\{L<\tau(X_{-\infty}^{n})\leq n\}} \mu_{\tau(X_{-\infty}^{n})}\right) &= \lim_{n\to\infty} E\left(I_{\{L<\tau(X_{-\infty}^{0})\leq n\}} \mu_{\tau(X_{-\infty}^{0})}\right) \\ &= E\left(I_{\{L<\tau(X_{-\infty}^{0})\}} \mu_{\tau(X_{-\infty}^{0})}\right) \\ &= \sum_{l=L+1}^{\infty} \mu_l P(\tau(X_{-\infty}^{0}) = l) < \epsilon \end{aligned}$$

and then the third term will be bounded by ϵ.

Now we deal with the last term. Since $\sum_{k=0}^{\infty} k^2 p_k < \infty$,

$$\begin{aligned} E\left(I_{\{\tau(X_{-\infty}^{n})>n\}} \theta_n\right) &= E(I_{\{\tau(X_{-\infty}^{n})>n\}} E(\sigma_n | X_0^n)) \\ &= \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} j p_{i+n+j} P(X_{-i} = 0) \\ &\leq \frac{\sum_{k=n}^{\infty} k^2 p_k}{1 + \sum_{h=0}^{\infty} h p_h} \to 0. \end{aligned}$$

Combining all these we get that for arbitrary $\epsilon > 0$,

$$\lim_{n\to\infty} E\left(|h_n^*(X_0^n) - \theta_n|\right) < \epsilon.$$

Since $\epsilon > 0$ was arbitrary this completes the proof of Theorem 2.1. □

3. THE TWO WAY UNIVERSAL SCHEME

Let $0 < \gamma < 1$ be arbitrary. First define the stopping times λ_n as $\lambda_0 = \psi$ and for $n \geq 1$,

$$\lambda_n = \min\{\ t > \lambda_{n-1} : \exists \psi < i \leq \lfloor \log t \rfloor \text{ such that } \tau(X_0^i) = \tau(X_0^t) \text{ and }$$
$$\left|\left\{ \lfloor \log t \rfloor < j < 2^{\lfloor \log t \rfloor} :\ \tau(X_0^j) = \tau(X_0^t) \right\}\right| \geq 2^{\lfloor \log t \rfloor (1-\gamma)}\ \}.$$

(Note that all logarithms are to the base 2.) Put

$$\kappa_n = \min\{K : \left|\left\{ \lfloor \log \lambda_n \rfloor < j \leq K : \tau(X_0^j) = \tau(X_0^{\lambda_n}) \right\}\right| = \lceil 2^{\lfloor \log \lambda_n \rfloor (1-\gamma)} \rceil \}.$$

Note that $\kappa_n < 2^{\lfloor \log \lambda_n \rfloor}$. For $n > 0$ define our estimator $h_n(X_0, \dots, X_{\lambda_n})$ at time λ_n as

$$h_n(X_0, \dots, X_{\lambda_n}) = \frac{1}{\lceil 2^{\lfloor \log \lambda_n \rfloor (1-\gamma)} \rceil} \sum_{i=\lfloor \log \lambda_n \rfloor + 1}^{\kappa_n} I_{\{\tau(X_0^i) = \tau(X_0^{\lambda_n})\}} \sigma_i.$$

(Notice that κ_n ensures that we take into consideration exactly $\lceil 2^{\lfloor \log \lambda_n \rfloor (1-\gamma)} \rceil$ occurrences.) The above formula is simply the average of the residual waiting times that we have already observed in the data segment $X_{\lfloor \log \lambda_n \rfloor + 1}^{\kappa_n}$ when we were at the same value of τ as we see at time λ_n. Note that as long as $2^m \leq \lambda_n < 2^{m+1}$ the estimator $h_n(X_0, \dots, X_{\lambda_n})$ is not refreshed. Keeping the same estimate for many values of n enables us to use weaker moment assumptions since the number of unfavorable events that we have to consider is reduced. Note that neither $h_n(X_0, \dots, X_{\lambda_n})$ nor λ_n depend on the following α.

Theorem 3.1. Assume $\sum_{k=0}^{\infty} k^{\alpha+1} p_k < \infty$ for some $\alpha > 1$. Let $0 < \gamma < 1/3$. Then for the stopping times λ_n and the estimator $h_n^*(X_0, \dots, X_n)$, almost surely,

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = 1, \tag{5}$$

$$\lim_{n\to\infty} I_{\{\exists k \geq 0 : \lambda_k = n\}} |h_n^*(X_0^n) - \mu_{\tau(X_{-\infty}^n)}| = 0 \tag{6}$$

and

$$\lim_{n\to\infty} E\left(|h_n^*(X_0^n) - \theta_n|\right) = 0. \tag{7}$$

P r o o f. The first statement (5) follows from Theorem 2 in [8]. We have to check the conditions of Theorem 2.1. Condition (1) follows from the definition of λ_n. Condition (3) follows from Theorem 2 in [8]. Now we deal with condition (4). Let $k < m$ be fixed. Define $j_0^{(k,m)} = m$ and for $i \geq 0$ let $j_{i+1}^{(k,m)}$ denote the $(i+1)$th occurrence of $\tau(X_{-\infty}^k)$ (reading forward, starting at position m), that is,

$$j_{i+1}^{(k,m)} = \min\left\{ t > j_i^{(k,m)} : \tau(X_{-\infty}^t) = \tau(X_{-\infty}^k) \right\}.$$

Now for $i \geq 1$ define

$$Z_i^{(k,m)} = \sigma_{j_i^{(k,m)}}.$$

Clearly $Z_i^{(k,m)}$ are conditionally independent and identically distributed given $\tau(X_{-\infty}^k) = L$. Without loss of generality we may assume that $1 < \alpha \leq 2$. Apply Theorem 2 of von Bahr and Esseen in [1] to get the inequality between the second and the third terms

$$\begin{aligned}
&E\left(\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \\
&= E\left(\left|\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} \left(\frac{Z_i^{(k,m)} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}}{\lceil (2^m)^{1-\gamma}\rceil}\right)\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \\
&\leq 2\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} E\left(\left|\frac{Z_i^{(k,m)} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}}{\lceil (2^m)^{1-\gamma}\rceil}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \\
&= 2\frac{1}{\lceil (2^m)^{1-\gamma}\rceil^{\alpha}} \sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} E\left(\left|Z_i^{(k,m)} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \\
&= 2\lceil (2^m)^{1-\gamma}\rceil^{1-\alpha} E\left(\left|Z_i^{(k,m)} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right).
\end{aligned}$$

Since for nonnegative a and b, $|a-b|^{\alpha} \leq a^{\alpha} + b^{\alpha}$ we get

$$\begin{aligned}
&2\lceil (2^m)^{1-\gamma}\rceil^{1-\alpha} E\left(\left|Z_i^{(k,m)} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \\
&\leq 2\lceil 2^{m(1-\gamma)}\rceil^{1-\alpha} \left(E\left(\left|Z_i^{(k,m)}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) + \left|\frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha}\right).
\end{aligned}$$

Notice that $E(|Z_1^{(k,m)}|^{\alpha}|\tau(X_{-\infty}^k) = L) = \frac{\sum_{h=0}^{\infty} h^{\alpha} p_{h+L}}{\sum_{h=L}^{\infty} p_h}$ and apply Jensen's inequality to get

$$\begin{aligned}
&2\lceil 2^{m(1-\gamma)}\rceil^{1-\alpha} \left(E\left(\left|Z_i^{(k,m)}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) + \left|\frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha}\right) \\
&\leq 2\lceil (2^m)^{1-\gamma}\rceil^{1-\alpha} \left(E\left(\left|Z_i^{(k,m)}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) + \frac{\sum_{h=0}^{\infty} h^{\alpha} p_{h+L}}{\sum_{h=L}^{\infty} p_h}\right) \\
&\qquad \leq 4(2^m)^{(1-\gamma)(1-\alpha)} \frac{\sum_{h=0}^{\infty} h^{\alpha} p_{h+L}}{\sum_{h=L}^{\infty} p_h}.
\end{aligned}$$

Thus

$$E\left(\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty} hp_{h+L}}{\sum_{h=L}^{\infty} p_h}\right|^{\alpha} |\tau(X_{-\infty}^k) = L\right) \leq 4(2^m)^{(1-\gamma)(1-\alpha)} \frac{\sum_{h=0}^{\infty} h^{\alpha} p_{h+L}}{\sum_{h=L}^{\infty} p_h}.$$

Multiply both sides of the last inequality by

$$P(\tau(X_{-\infty}^k)=L)=\frac{1}{1+\sum_{h=0}^\infty hp_h}\sum_{h=L}^\infty p_h$$

(note that by Kac's theorem

$$P(X_{k-L}=0)=\frac{1}{1+\sum_{h=0}^\infty hp_h}$$

cf. [4] Ch. XIII and [13] Sec. I.2.c) and sum over L. It is easy to see that

$$\sum_{L=0}^\infty \frac{\sum_{h=0}^\infty h^\alpha p_{h+L}}{\sum_{h=L}^\infty p_h}\frac{\sum_{h=L}^\infty p_h}{1+\sum_{h=0}^\infty hp_h}\le\frac{\sum_{h=0}^\infty h^{\alpha+1}p_h}{1+\sum_{h=0}^\infty hp_h}$$

and we get the following estimate

$$E\left(\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|^\alpha\right)\le\frac{4}{(2^m)^{(1-\gamma)(\alpha-1)}}\frac{\sum_{h=0}^\infty h^{\alpha+1}p_h}{1+\sum_{h=0}^\infty hp_h}$$

and in turn

$$\begin{aligned}&E\left(\max_{0\le k\le m-1}\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|^\alpha\right)\\ \le\ &\frac{4m}{(2^m)^{(1-\gamma)(\alpha-1)}}\frac{\sum_{h=0}^\infty h^{\alpha+1}p_h}{1+\sum_{h=0}^\infty hp_h}.\end{aligned}$$

Now

$$\begin{aligned}&E\left(\max_{0\le k\le m-1}\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|\right)\\ &=E\left(\max_{0\le k\le m-1}\left(\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|^\alpha\right)^{\frac{1}{\alpha}}\right)\\ &\le\left(E\left(\max_{0\le k\le m-1}\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|^\alpha\right)\right)^{\frac{1}{\alpha}}\\ &\qquad\le\left(\frac{4m}{(2^m)^{(1-\gamma)(\alpha-1)}}\frac{\sum_{h=0}^\infty h^{\alpha+1}p_h}{1+\sum_{h=0}^\infty hp_h}\right)^{\frac{1}{\alpha}}\end{aligned}$$

and the right hand side tends to zero. Now let $m=\lfloor\log n\rfloor$ and observe that

$$\begin{aligned}&E\left(I_{\{\exists k\ge 0:\lambda_k=n\}}\left|h_{\chi_n}(X_0^{\lambda_{\chi_n}})-\mu_{\tau(X_{-\infty}^n)}\right|\right)\\ \le\ &E\left(\max_{0\le k\le m-1}\left|\frac{\sum_{i=1}^{\lceil (2^m)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (2^m)^{1-\gamma}\rceil}-\frac{\sum_{h=0}^\infty hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^\infty p_h}\right|\right)\end{aligned}$$

which tends to zero. Now we prove that (2) is satisfied. Indeed,

$$\lim_{n\to\infty} E\left(I_{\{\tau(X^n_{-\infty})\le L\}} I_{\{\forall k\ge 0:\lambda_k\ne n\}}\mu_{\tau(X^n_{-\infty})}\right)$$
$$\le \lim_{n\to\infty}\sum_{l=1}^{L}\mu_l(P(\forall\psi < i\le \lfloor\log n\rfloor \ :\ \tau(X_0^i)\ne\tau(X_0^n))$$
$$+P(\left|\left\{\lfloor\log n\rfloor < j < 2^{\lfloor\log n\rfloor}\ :\ \tau(X_0^j)=\tau(X_0^n)\right\}\right| < 2^{\lfloor\log n\rfloor(1-\gamma)})).$$

By ergodicity,

$$\lim_{n\to\infty} P(\forall\psi < i\le \lfloor\log n\rfloor \ :\ \tau(X_0^i)\ne\tau(X_0^n))=0$$

and

$$\lim_{n\to\infty} P(\left|\left\{\lfloor\log n\rfloor < j < 2^{\lfloor\log n\rfloor}\ :\ \tau(X_0^j)=\tau(X_0^n)\right\}\right| < 2^{\lfloor\log n\rfloor(1-\gamma)})=0.$$

The proof of Theorem 3.1 is complete. □

4. JUSTIFYING INTERMITTENT SCHEMES

In this section we will give an explicit construction that justifies the intuitive fact that when the quantity that is being estimate is potentially infinite one shouldn't expect to have a pointwise consistent scheme valid for all large n. The construction is given in a positive form, we will start from any scheme with certain properties and show how to construct a binary renewal process for which the scheme will fail infinitely often to give a good estimate for the conditional expectation of the residual waiting time.
Consider the class of those stationary and ergodic binary renewal processes which are arbitrary finite order Markov chains. Then it is clear that one can estimate the conditional expectation of the residual waiting time for all time instances. Indeed, estimate the order of the Markov chain by a consistent order estimation scheme $ORDEST(X_0^n)$ (you may use e. g. [2, 3, 6, 7]) or [10] and then calculate the conditional expectation of the residual waiting time by using frequency count. Since there are finitely many possible strings with length equal to the order the ergodic theorem yields consistency of the estimator.

Theorem 4.1. For any estimation scheme $g_n(X_0^n)$ such that for all those stationary and ergodic binary renewal processes $\{Y_n\}$ which are Markov with some finite but unknown order,

$$\lim_{n\to\infty}\left|g_n(Y_0^n)-E(\sigma_n(Y_{n+1}^\infty)|Y_0^n)\right|=0$$

almost surely, there exists a stationary and ergodic binary renewal process $\{X_n\}$ with $E(\sigma_0^m(X_1^\infty))<\infty$ for all $m>0$ such that with positive probability,

$$\limsup_{n\to\infty}\left|g_n(X_0^n)-E(\sigma_n(X_{n+1}^\infty)|X_0^n)\right|>0.75.$$

Proof. Let $0 = E_0 < E_1 < \ldots$ be an increasing sequence of nonnegative integers which will be specified later. Put

$$S = \{E_i : \ 0 \le i < \infty\}$$

and

$$S_k = \{E_i : \ 0 \le i < k\}.$$

Let

$$f(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{otherwise.} \end{cases}$$

Define the transition probabilities of the Markov chain $\{M_n^{(k)}\}$ as

$$p_{i,j}^{(k)} = \begin{cases} 1 & \text{if } i \in S_k \text{ and } j = i+1 \\ 0.5 & \text{if } i \notin S_k \text{ and } j = 0 \\ 0.5 & \text{if } i \notin S_k \text{ and } j = i+1 \\ 0 & \text{otherwise} \end{cases}$$

and the transition probabilities of the Markov chain $\{M_n\}$ as

$$p_{i,j} = \begin{cases} 1 & \text{if } i \in S \text{ and } j = i+1 \\ 0.5 & \text{if } i \notin S \text{ and } j = 0 \\ 0.5 & \text{if } i \notin S \text{ and } j = i+1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $X_n^{(k)} = f(M_n^{(k)})$ and $X_n = f(M_n)$. (Each process $X_n^{(k)}$ will be Markov with some finite order since $|S_k| < \infty$ but process X_n will not be Markov of any order.) Now we start with S_0 as the empty set. Let $L_0 = N_{-1} = 1$. Given the Markov chain $\{M_n^{(k)}\}$ we proceed to define $\{M_n^{(k+1)}\}$. We will make our modification in such a manner that the probabilities of strings of length up to some L_k conditioned on starting at 0 will be changed only slightly so that the behavior of the estimators g_n on strings up to that length will be maintained. Define the event

$$\begin{aligned} C_{k,N} \quad = \quad & \{ \left| g_n(X_0^{(k)}, \ldots, X_n^{(k)}) - E\left(\sigma_n(X_{n+1}^{(k)}, \ldots,)|X_0^{(k)}, \ldots, X_n^{(k)}\right) \right| < 0.25 \\ & \text{for all } n \ge N\}. \end{aligned}$$

Let $N_k \ge 2N_{k-1}$ be so large that

$$P(C_{k,N_k}|X_0^{(k)} = 0) > 1 - 1/2^k.$$

Now let $E_k = N_k + L_k$. Observe that since

$$E\left(\sigma_n(X_{n+1}^{(k)}, \ldots)|X_{n-E_k}^{(k)} = 0, X_{n-E_k+1}^{(k)} = 1, \ldots, X_n^{(k)} = 1\right) = 2$$

whereas

$$E\left(\sigma_n(X_{n+1}^{(k+1)}, \ldots)|X_{n-E_k}^{(k+1)} = 0, X_{n-E_k+1}^{(k+1)} = 1, \ldots, X_n^{(k+1)} = 1\right) = 3$$

so that the first time that we will see a string of ones of length E_k with probability greater than $1 - 1/2^k$ the scheme g_n will be in error by at least 0.75. Define L_{k+1} to be large enough so that with probability greater than $1 - 1/2^k$ starting at 0 we will encounter a string with such a length by the time we have observed the first L_{k+1} X_i's.

More formally if we introduce the set D_{k+1} defined by:

$$X_{n-E_k}^{(k+1)} = 0, X_{n-E_k+1}^{(k+1)} = 1, \dots, X_n^{(k+1)} = 1$$

for some $L_k < n < L_{k+1}$ then its probability conditioned on $X_0^{(k+1)} = 0$ is greater than $1 - 1/2^k$.

Continuing by induction we define the final process X_n and with probability one the scheme will be in error by a definite amount infinitely many times given $X_0 = 0$. . What remains to be proved that $E(\sigma_n^m(X_{n+1}^\infty)) < \infty$ for all $m > 0$.

This is so since

$$E(\sigma_0^{m+1}(X_1^\infty)|X_0 = 0) \leq \sum_{j=0}^{\infty} j^{m+1}(0.5)^{j-\rho_j} \leq \sum_{j=0}^{\infty} j^{m+1}(0.5)^{j-\log(j)} < \infty$$

where $\rho_j = \#\{N_i < j\}$. This completes the proof of Theorem 4.1. □

(Received February 26, 2015)

REFERENCES

[1] B. von Bahr and C. G. Esseen: Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$. Annals Math. Statist. *36* (1965), 299–303. DOI:10.1214/aoms/1177700291

[2] I. Csiszár and P. Shields: The consistency of the BIC Markov order estimator. Annals Statist. *28* (2000), 1601–1619. DOI:10.1214/aos/1015957472

[3] I. Csiszár: Large-scale typicality of Markov sample paths and consistency of MDL order estimators. IEEE Trans. Inform. Theory *48* (2002), 1616–1628. DOI:10.1109/tit.2002.1003842

[4] W. Feller: An Introduction to Probability Theory and its Applications Vol. I. Third edition. John Wiley and Sons, New York – London – Sydney 1968.

[5] S. Ghahramani: Fundamentals of Probability with Stochastic Processes. Third edition. Pearson Prentice Hall, Upper Saddle River NJ 2005.

[6] G. Morvai and B. Weiss: Order estimation of Markov chains. IEEE Trans. Inform. Theory *51* (2005), 1496–1497. DOI:10.1109/tit.2005.844093

[7] G. Morvai and B. Weiss: Estimating the lengths of memory words. IEEE Trans. Inform. Theory *54* (2008), 8, 3804–3807. DOI:10.1109/tit.2008.926316

[8] G. Morvai and B. Weiss: On universal estimates for binary renewal processes. Ann. Appl. Probab. *18* (2008), 5, 1970–1992. DOI:10.1214/07-aap512

[9] G. Morvai and B. Weiss: Estimating the residual waiting time for binary stationary time series. In: Proceedings of ITW2009, Volos 2009, pp. 67–70. DOI:10.1109/itwnit.2009.5158543

[10] G. Morvai and B. Weiss: Universal tests for memory words. IEEE Trans. Inform. Theory *59* (2013), 6873–6879. DOI:10.1109/tit.2013.2268913

[11] G. Morvai and B. Weiss: Inferring the residual waiting time for binary stationary time series. Kybernetika *50* (2014), 869–882. DOI:10.14736/kyb-2014-6-0869

[12] B. Ya. Ryabko: Prediction of random sequences and universal coding. Probl. Inform. Transmiss. *24* (1988), 87–96.

[13] P. C. Shields: The Ergodic Theory of Discrete Sample Paths. Graduate Studies in Mathematics, American Mathematical Society, Providence *13* 1996. DOI:10.1090/gsm/013

Gusztáv Morvai, Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, 13-15 Reáltanoda utca , H-1053, Budapest, Hungary and MTA-BME Stochastics Research Group, 1 Egry József utca , Building H, Budapest,1111. Hungary.
e-mail: morvai@math.bme.hu

Benjamin Weiss, Hebrew University of Jerusalem, Jerusalem 91904. Israel.
e-mail: weiss@math.huji.ac.il