KYBERNETIKA — VOLUME *50* (2014), NUMBER 6, PAGES 869–882

INFERRING THE RESIDUAL WAITING TIME FOR BINARY STATIONARY TIME SERIES

Gusztáv Morvai and Benjamin Weiss

For a binary stationary time series define σ_n to be the number of consecutive ones up to the first zero encountered after time n, and consider the problem of estimating the conditional distribution and conditional expectation of σ_n after one has observed the first n outputs. We present a sequence of stopping times and universal estimators for these quantities which are pointwise consistent for all ergodic binary stationary processes. In case the process is a renewal process with zero the renewal state the stopping times along which we estimate have density one.

Keywords: nonparametric estimation, stationary processes

Classification: 62G05, 60G25, 60G10

1. INTRODUCTION AND RESULTS

Let $\{X_n\}$ be a binary-valued stationary and ergodic time series. 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 residual waiting time to the state 0 given some previous observations, in particular given X_0^n.

The "waiting time" is often used for the time between successive occurrences of a fixed state. For this reason we add the adjective "residual" when we want to describe the time until the next occurrence of 0, this is sometimes called the entrance time to 0.

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

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

Our goal is to estimate both the conditional distribution of σ_n given $\{X_0^n\}$ and also the conditional expectation, $E(\sigma_n | X_0^n)$, without prior knowledge of the distribution function of the process. In principle, we can learn something about the finite distributions of the unknown process that is being observed from the observations of X_0^n but it is somewhat problematic to use this information to estimate the future which also depends on these same observations. If we would be content with convergence in probability, then simple averaging to obtain estimates of the finite distributions, usually suffices. However if

DOI: 10.14736/kyb-2014-6-0869

we raise our sights and try to obtain consistent estimates with probability one then more elaborate schemes are needed. In fact for the class of all ergodic processes even the estimation of the conditional expectation of X_{n+1} given X_0^n cannot be carried out universally, i. e. without any knowledge of the processs, cf. [5, 12] and [1] . The residual waiting time until the occurrence of the next zero requires looking ahead on unbounded number of steps – and a fortiori cannot be universally estimated. To get positive results we give up on trying to estimate at all time instants n, and develop stopping times which tell us when the conditions are favourable for us to make on estimate – but we will insist on almost sure consistency.

In our recent paper [11] we gave stopping time estimates for the conditional expectation of a function $f(X_{n+1})$ of the next observation givenX_0^n in case the process takes values from a countably infinite alphabet. These estimates were shown to be almost surely eventually consistent provided that the conditional distribution of X_0 given $X_{-\infty}^{-1}$ was almost surely continuous. In the present paper we will treat arbitrary stationary ergodic $\{0,1\}$ valued process but focus on the residual waiting time until the next 0 occurs. For this random variable we will consistently estimate its conditional expectation and its probability distribution.

Restricting to a smaller class of processes, the binary renewal processes with zero the renewal state, enables us to give estimation schemes that are much better. In fact we will use two schemes – one adapted to general processes and one to the renewal process – and a test which will decide for us at each step which scheme should be applied. For renewal processes the stopping times will have density one – but for general processes they will, of necessity, be much rarer.

Here is the first auxiliary algorithm. It is constructed so that if the process is a renewal process with renewal state 0 then it will give a consistent estimator for the conditional expectation of the residual waiting time. In case of binary renewal processes (with renewal state 0) if a zero occurs then the expected time depends on the location of the zero and so we introduce the notation:

$$\tau_n(X_{-\infty}^n) = \min\{t \geq 0 : X_{n-t} = 0\}. \tag{2}$$

If a zero occurs in X_0^n then $\tau_n(X_{-\infty}^n) \leq n$ and so it depends only on X_0^n and so we will also write for $\tau_n(X_{-\infty}^n)$, $\tau_n(X_0^n)$ in this case. We will ensure that we will use the notation $\tau_n(X_0^n)$ only in the case X_0^n will contain at least one zero.

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

$$\psi = \min\{t \geq 0 : X_t = 0\}. \tag{3}$$

In order to reduce our assumption $E(|\sigma_0|^\alpha) < \infty$ in [10] from $\alpha > 2$ to $\alpha > 1$ a slightly more involved scheme of stopping times is needed.

Let $0 < \delta < 1$ be arbitrary. First define the stopping times ξ_n^* as $\xi_0^* = \psi$ and for $n \geq 1$,

$$\xi_n^* = \min\Big\{ t > \xi_{n-1}^* : \exists i \in (\psi, \log t) \text{ such that } \tau(X_0^i) = \tau(X_0^t) \text{ and}$$
$$\Big|\Big\{\log t \leq j < 2^{\lfloor \log t \rfloor} : \ \tau(X_0^j) = \tau(X_0^t)\Big\}\Big| \geq 2^{\lfloor \log t \rfloor (1-\delta)} \Big\}. \tag{4}$$

(Note that the stopping time refer to the natural filtration and all logarithms are to the base 2.) Put

$$\kappa_n^* = \min\left\{K : \left|\left\{\lfloor\log \xi_n^*\rfloor < j \le K : \tau(X_0^j) = \tau(X_0^{\xi_n^*})\right\}\right| = \lceil 2^{\lfloor\log \xi_n^*\rfloor(1-\delta)}\rceil\right\}. \quad (5)$$

Note that $\kappa_n^* < 2^{\lfloor\log \xi_n^*\rfloor}$. As in [9], for $n > 0$, define our estimator $h_n^*(X_0, \ldots, X_{\xi_n^*})$ at time ξ_n^* as

$$h_n^*(X_0, \ldots, X_{\xi_n^*}) = \frac{1}{\lceil 2^{\lfloor\log \xi_n^*\rfloor(1-\delta)}\rceil} \sum_{i=\lfloor\log \xi_n^*\rfloor+1}^{\kappa_n^*} I_{\{\tau(X_0^i)=\tau(X_0^{\xi_n^*})\}}\sigma_i. \quad (6)$$

Note that $h_n^*(X_0, \ldots, X_{\xi_n^*})$ will estimate the value $E(\sigma_{\xi_n^*}|X_0^{\xi_n^*})$ consistently in case the process is a renewal process. Notice that κ_n^* ensures that we take into consideration exactly $\lceil 2^{\lfloor\log \xi_n^*\rfloor(1-\delta)}\rceil$ pieces of occurrences. The above formula is simply the average of the residual waiting times that we have already observed in the data segment $X_{\lfloor\log \xi_n^*\rfloor+1}^{\kappa_n^*}$ when we were at the same value of τ as we see at time ξ_n^*. Note that as long as $2^m \le \xi_n^* < 2^{m+1}$ the estimator $h_n^*(X_0, \ldots, X_{\xi_n^*})$ is not refreshed. Keeping the same estimate for many values of n enables us to use weaker moment assumptions than in [10] since the number of unfavorable events that we have to consider is reduced.

In a similar fashion we can define the average of the number of times that the residual waiting time assumed a fixed value and this will provide us with an estimator for the entire conditional distribution of σ_n. Namely, define $\hat{q}_l^*(X_0, \ldots, X_{\xi_n^*})$ for each l as

$$\hat{q}_l^*(X_0, \ldots, X_{\xi_n^*}) = \frac{1}{\lceil 2^{\lfloor\log \xi_n^*\rfloor(1-\delta)}\rceil} \sum_{i=\lfloor\log \xi_n^*\rfloor+1}^{\kappa_n^*} I_{\left\{\tau(X_0^i)=\tau(X_0^{\xi_n^*}),\sigma_i=l\right\}}. \quad (7)$$

Note that $\hat{q}_l^*(X_0, \ldots, X_{\xi_n^*})$ is a probability distribution on the nonnegative integers and neither h_n^*, $\hat{q}_l^*(X_0, \ldots, X_{\xi_n^*})$ nor ξ_n^* depend on α.

Now we define the second type of auxiliary algorithms which will work for general (not necessarily renewal) processes, cf [10].

We define the stopping times $\{\eta_n\}$. (The event $\{\eta_n = s\}$ will be measurable with respect to the sigma-algebra generated by X_0^s.) Set $\eta_0 = 0$. For $n = 1, 2, \ldots$, define η_n recursively. Let

$$\eta_n = \eta_{n-1} + \min\{t > 0 : X_t^{\eta_{n-1}+t} = X_0^{\eta_{n-1}}\}. \quad (8)$$

Note that by ergodicity, the word $X_0^{\eta_{n-1}}$ will appear again with probability one and so η_n is finite almost surely. By construction, $\eta_n \ge n$ and it is a stopping time on $X_0, X_1, \ldots$. The nth estimate m_n for $E(\sigma_{\eta_n}|X_0^{\eta_n})$ is defined as

$$m_n(X_0, \ldots, X_{\eta_n}) = \frac{1}{n}\sum_{j=0}^{n-1} \sigma_{\eta_j} I_{\{\sigma_{\eta_j} \le \eta_{j+1}-\eta_j\}} \quad (9)$$

and $\hat{q}_l(X_0,\ldots,X_{\eta_n})$ for $P(\sigma_{\eta_n}=l|X_0^{\eta_n})$ is defined as

$$\hat{q}_l(X_0,\ldots,X_{\eta_n})=\frac{1}{n}\sum_{j=0}^{n-1} I_{\{\sigma_{\eta_j}=l,\sigma_{\eta_j}\le\eta_{j+1}-\eta_j\}}. \tag{10}$$

Observe that both m_n and $\hat{q}_l(X_0,\ldots,X_{\eta_n})$ depend solely on $X_0^{\eta_n}$.

Now we define our final scheme. This will involve a test to determine whether or not the process that we are observing is indeed a renewal process and tell us which of the auxiliary algorithms we should be using. Define the empirical conditional distributions by

$$\hat{p}_n(0|z_{-k}^0)=\frac{\left|\{k\le t\le n-1 : X_{t-k}^{t+1}=(z_{-k}^0,0)\}\right|}{\left|\{k\le t\le n-1 : X_{t-k}^{t}=z_{-k}^0\}\right|}. \tag{11}$$

where $0/0$ is defined as 0. These empirical distributions are functions of X_0^n, but we suppress this dependence to keep the notation manageable.

Define $\mathcal{X}_{k,i}$ to be the set of words in $\{0,1\}^{k+i+1}$ whose suffix consists of 01^k (where 1^k stands for k pieces of 1's) and then for a fixed $0<\gamma<1$ define $\mathcal{S}_{k,i}^n$ as

$$\mathcal{S}_{k,i}^n=\left\{x_{-k-i}^0\in\mathcal{X}_{k,i} : \left|\{k+i\le t\le n-1 : X_{t-k-i}^t=x_{-k-i}^0\}\right|>n^{1-\gamma}\right\}. \tag{12}$$

These are the strings which occur sufficiently often so that we can rely on their empirical distribution.

Define

$$\hat{\Delta}_n=\max_{0\le k\le n}\max_{1\le i\le n}\max_{z_{-k-i}^0\in\mathcal{S}_{k,i}^n}\left|\hat{p}_n(0|z_{-k}^0)-\hat{p}_n(0|z_{-k-i}^0)\right|. \tag{13}$$

(Note that the maximum over an empty set is considered to be zero.) For a non renewal process we will have that

$$\liminf_{n\to\infty}\hat{\Delta}_n>0 \quad \text{almost surely}$$

while for a renewal process we will be able to give a definite rate of convergence to zero of this estimator. This motivates the stopping times $\{\lambda_n\}$ along which we will estimate.

Set $\lambda_0=0$ and for $n\ge 1$ define

$$\lambda_n=\min\left\{\eta_{\min\{i:\hat{\Delta}_{\eta_i}>\eta_i^{-\beta},\eta_i>\lambda_{n-1}\}},\xi^*_{\min\{i:\hat{\Delta}_{\xi_i^*}\le(\xi_i^*)^{-\beta},\xi_i^*>\lambda_{n-1}\}}\right\}. \tag{14}$$

The nth estimate f_n is defined as

$$f_n(X_0^{\lambda_n}) \tag{15}$$
$$=\begin{cases} h_i(X_0^{\xi_i^*}), & \text{if } \hat{\Delta}_{\lambda_n}\le\lambda_n^{-\beta} \text{ and } \lambda_n=\xi_i^* \text{ for some } i\\ m_i(X_0^{\eta_i}), & \text{if } \hat{\Delta}_{\lambda_n}>\lambda_n^{-\beta} \text{ and } \lambda_n=\eta_i \text{ for some } i\end{cases}$$

and the nth estimate $q_l(X_0,\dots,X_{\lambda_n})$ is defined as

$$q_l(X_0^{\lambda_n}) = \begin{cases} \hat{q}_l^*(X_0^{\xi_i^*}), & \text{if } \hat{\Delta}_{\lambda_n} \le \lambda_n^{-\beta} \text{ and } \lambda_n = \xi_i^* \text{ for some } i \\ \hat{q}_l(X_0^{\eta_i}), & \text{if } \hat{\Delta}_{\lambda_n} > \lambda_n^{-\beta} \text{ and } \lambda_n = \eta_i \text{ for some } i. \end{cases} \tag{16}$$

For our proof of the theorem we will need some moment condition on the residual waiting time. For more details on the necessity of some moment condition see Theorem 4 in Morvai and Weiss [9]. In our previous paper [10] we considered only the problem of estimating the conditional expectation of the residual waiting time and our condition $\alpha > 2$ in that paper is stronger than the condition $\alpha > 1$ used here.

Theorem. Let $\{X_n\}$ be a binary-valued stationary and ergodic time series. Assume $E(|\sigma_0|^\alpha) < \infty$ for some $\alpha > 1$. Let $0 < \gamma < 1$, $0 < \beta < \frac{1-\gamma}{2}$, $0 < \delta < \frac{1}{3}$ be arbitrary. Then almost surely

$$\lim_{n\to\infty} \left| f_n(X_0^{\lambda_n}) - E(\sigma_{\lambda_n} | X_0^{\lambda_n}) \right| = 0 \tag{17}$$

and

$$\lim_{n\to\infty} \sum_{l=0}^{\infty} \left| q_l(X_0,\dots,X_{\lambda_n}) - P(\sigma_{\lambda_n} = l | X_0^{\lambda_n})) \right| = 0. \tag{18}$$

If in addition the process is a binary renewal process then almost surely,

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = 1. \tag{19}$$

Remark 1. Note that neither f_n, $q_l(X_0,\dots,X_{\lambda_n})$ nor λ_n depend on α.

Remark 2. If the process is not a renewal process then (17) and (18) hold for $\alpha = 1$, (for the backward scheme cf. Algoet [1]).

2. PROOF OF THE THEOREM

The proof of theorem will be divided into several steps. The first few steps construct another version of the stochastic process. It plays an important role in analyzing the estimators that are consistent for any ergodic binary process.

We construct two schemes – one for general stationary processes and a more efficient one designed for renewal processes. One of these schemes is based on universal so called "backward schemes" which provide estimators for functions of future X_0^∞ given that one learns more and more of the past. The general backward scheme that we use is closely related to the one used by Algoet [1]. To apply this we first construct an auxiliary process – this occupies steps $0-3$. In step 45 we show the consistency of the general scheme. In step 6 we show the consistency of schemes for renewal processes. Step 7 shows that our test for deciding which of the two schemes should be applied is eventually correct, and finally step 8 gives the proof of the theorem.

Step 0. *We define some auxiliary processes.*

It will be useful to define $\{\hat{X}_n^{(k)}\}_{n=-\infty}^{\infty}$ for $k \geq 0$ as follows. Let

$$\hat{X}_{-n}^{(k)} = X_{\eta_k - n} \text{ for } -\infty < n < \infty.$$

For an arbitrary stationary time series $\{Y_n\}$, let $\hat{\eta}_0(Y_{-\infty}^0) = 0$ and for $n \geq 1$ define

$$\hat{\eta}_n(Y_{-\infty}^0) = \hat{\eta}_{n-1}(Y_{-\infty}^0) - \min\{t > 0 : Y_{\hat{\eta}_{n-1}-t}^{-t} = Y_{\hat{\eta}_{n-1}}^0\}.$$

Let T denote the left shift operator, that is, $(Tx_{-\infty}^{\infty})_i = x_{i+1}$. It is easy to see that if $\eta_n(x_{-\infty}^{\infty}) = l$ then $\hat{\eta}_n(T^l x_{-\infty}^{\infty}) = -l$.

Define the time series $\{\tilde{X}_n\}_{n=-\infty}^{0}$ as

$$\tilde{X}_{\hat{\eta}_k(\tilde{X}_{-\infty}^0)}^0 = X_0^{\eta_k} \text{ for } k \geq 0.$$

Since $X_{\eta_{k+1}-\eta_k}^{\eta_{k+1}} = X_0^{\eta_k}$ process $\{\tilde{X}_n\}_{n=-\infty}^{0}$ is well defined.

Step 1. *We show that for arbitrary $k \geq 0$, the time series $\{\hat{X}_n^{(k)}\}_{n=-\infty}^{\infty}$ and $\{X_n\}_{n=-\infty}^{\infty}$ have identical distribution.*

It is enough to show that for all $k \geq 0$, $m \geq n \geq 0$, and Borel set $F \subseteq \{0,1\}^{n+1}$,

$$P((\hat{X}_{m-n}^{(k)}, \ldots, \hat{X}_m^{(k)}) \in F) = P(X_{m-n}^m \in F).$$

This is immediate by stationarity of $\{X_n\}$ and by the fact that for all $k \geq 0$, $m \geq n \geq 0$, $l \geq 0$, $F \subseteq \{0,1\}^{n+1}$,

$$T^l\{X_{\eta_k+m-n}^{\eta_k+m} \in F, \eta_k = l\} = \{X_{m-n}^m \in F, \hat{\eta}_k(X_{-\infty}^0) = -l\}.$$

Step 2. *We show that for $k \geq 0$, almost surely,*

$$\hat{\eta}_k(\ldots, \hat{X}_{-1}^{(k)}, \hat{X}_0^{(k)}) = \hat{\eta}_k(\tilde{X}_{-\infty}^0)$$

and

$$\tilde{X}_{\hat{\eta}_k(\tilde{X}_{-\infty}^0)}^0 = \hat{X}_{\hat{\eta}_k(\ldots,\hat{X}_{-1}^{(k)},\hat{X}_0^{(k)})}^{(k)}, \ldots, \hat{X}_0^{(k)}.$$

This statement follows immediately from Step 0.

Step 3. *We show that the distributions of $\{\tilde{X}_n\}_{n=-\infty}^{0}$ and $\{X_n\}_{n=-\infty}^{0}$ are the same.*

This is immediate from Step 1 and Step 2.

The time series $\{\tilde{X}_n\}_{n=-\infty}^{0}$ is stationary, since $\{X_n\}_{n=-\infty}^{0}$ is stationary, and it can be extended to be a two-sided time series $\{\tilde{X}_n\}_{n=-\infty}^{\infty}$. We will use this fact only for the purpose of defining the conditional expectation $E(\sigma_0(\tilde{X}_1^{\infty})|\tilde{X}_{-\infty}^0)$.

Step 4. *We show the consistency of the auxiliary algorithms m_n and $\hat{q}_l(X_0^{\eta_j})$.*

We proceed in a manner analogous to the one that we used in our earlier works: [6, 10]. Consider

$$
\begin{aligned}
& m_n - E(\sigma_{\eta_n}|X_0^{\eta_n}) \\
&= \frac{1}{n}\sum_{j=0}^{n-1}\left(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j\}} - \sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}}\right) \\
&+ \frac{1}{n}\sum_{j=0}^{n-1}\left(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} - E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}}|X_0^{\eta_j})\right) \\
&+ \frac{1}{n}\sum_{j=0}^{n-1}\left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}}|X_0^{\eta_j}) - E(\sigma_{\eta_j}|X_0^{\eta_j})\right) \\
&+ \left(\frac{1}{n}\sum_{j=0}^{n-1} E(\sigma_{\eta_j}|X_0^{\eta_j})\right) - E(\sigma_{\eta_n}|X_0^{\eta_n})
\end{aligned}
$$

and denote the four terms in the right hand side by

$$A_n + B_n + C_n + D_n.$$

Next we truncate like Algoet did in [1] and observe that since the σ_{η_j} are identically distributed by Step 0,

$$P(\sigma_{\eta_j} > j) \le P(\sigma_0 > j)$$

and

$$\sum_{j=0}^{\infty} P(\sigma_0 > j) = E(\sigma_0) < \infty.$$

The Borel–Cantelli lemma yields that eventually almost surely

$$\sigma_{\eta_j} \le j$$

and so

$$A_n \to 0$$

almost surely. Now we will deal with B_n. Put

$$\Gamma_{j+1} = \sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} - E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}}|X_0^{\eta_j}).$$

We will consider the following martingale U_n

$$U_n = \sum_{j=1}^{n} \frac{\Gamma_j}{j}$$

with respect to the sigma-field generated by $X_0^{\eta_n}$ To apply martingale convergence theorem for L^2-bounded martingales (cf. Theorem 2 p. 242 in [4])we have to show that

$$\sum_{j=1}^{\infty} \frac{E(\Gamma_j^2)}{j^2} < \infty.$$

Now

$$E\left((\Gamma_{j+1})^2\right) \leq E\left((\sigma_{\eta_j})^2 I_{\{\sigma_{\eta_j} \leq \eta_{j+1}-\eta_j, \sigma_{\eta_j} \leq j\}}\right) \leq E\left((\sigma_{\eta_j})^2 I_{\{\sigma_{\eta_j} \leq j\}}\right).$$

By Step 1 the σ_{η_j}'s are identically distributed therefore

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{E(\Gamma_j^2)}{j^2} &\leq \sum_{j=1}^{\infty} \frac{1}{j^2} E\left((\sigma_{\eta_{j-1}})^2 I_{\{\sigma_{\eta_{j-1}} \leq j-1\}}\right) \\ &= \sum_{j=1}^{\infty} \left(E\left((\sigma_0)^2 I_{\{\sigma_0 = j-1\}}\right) \left(\sum_{l=j}^{\infty} \frac{1}{l^2} \right) \right) \\ &\leq KE(\sigma_0) \end{aligned}$$

where K is a suitable constant. We conclude that the sequence U_n converges almost surely and Kronecker's lemma (cf. Shiryayev [14] p. 365) yields,

$$B_n = \frac{1}{n} \sum_{j=1}^{n} \Gamma_j \to 0 \text{ almost surely.}$$

Now we deal with the third term C_n. Clearly,

$$E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j} \leq \eta_{j+1}-\eta_j, \sigma_{\eta_j} \leq j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \leq 0$$

and

$$\limsup_{j\to\infty} \left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j} \leq \eta_{j+1}-\eta_j, \sigma_{\eta_j} \leq j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \right) \leq 0.$$

Since $\sigma_{\eta_j} \leq \eta_{j+1} - \eta_j$ as soon as there is at least one zero in $X_0^{\eta_j}$ for an arbitrary L, eventually almost surely,

$$E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j} \leq L\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \leq E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j} \leq \eta_{j+1}-\eta_j, \sigma_{\eta_j} \leq j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}).$$

By Step 2, Step 1 and Step 3,

$$E(\sigma_{\eta_j} | X_0^{\eta_j}) = E\Big(\sigma_0(\tilde{X}_1^{\infty}) | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}\Big).$$

The latter forms a martingale and so

$$E(\sigma_{\eta_j} | X_0^{\eta_j}) = E(\sigma_0(\tilde{X}_1^{\infty}) | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}) \to E(\sigma_0(\tilde{X}_1^{\infty}) | \tilde{X}^0_{-\infty}) \tag{20}$$

almost surely. Similarly,

$$E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j} \leq L\}} | X_0^{\eta_j}) = E(\sigma_0 I_{\{\sigma_{\eta_0}(\tilde{X}_1^{\infty}) \leq L\}}(\tilde{X}_1^{\infty}) | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}).$$

The latter forms a martingale and so

$$\begin{aligned} & E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le L\}} | X_0^{\eta_j}) \\ = \; & E(\sigma_0(\tilde{X}_1^{\infty}) I_{\{\sigma_{\eta_0}(\tilde{X}_1^{\infty})\le L\}} | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}) \to E(\sigma_0(\tilde{X}_1^{\infty}) I_{\{\sigma_{\eta_0}(\tilde{X}_1^{\infty})\le L\}} | \tilde{X}^0_{-\infty}) \end{aligned}$$

almost surely. Now, almost surely,

$$\begin{aligned} & \liminf_{j\to\infty} \left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \right) \\ & \quad \ge E(\sigma_0(\tilde{X}_1^{\infty}) I_{\{\sigma_{\eta_0}(\tilde{X}_1^{\infty})\le L\}} | \tilde{X}^0_{-\infty}) - E(\sigma_0(\tilde{X}_1^{\infty}) | \tilde{X}^0_{-\infty}). \end{aligned}$$

Since L was arbitrary,

$$\liminf_{j\to\infty} \left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \right) \ge 0$$

almost surely. Combining the upper and lower bounds we get that almost surely

$$\lim_{j\to\infty} \left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \right) = 0.$$

Thus by the Toeplitz lemma

$$C_n = \frac{1}{n} \sum_{j=0}^{n-1} \left(E(\sigma_{\eta_j} I_{\{\sigma_{\eta_j}\le \eta_{j+1}-\eta_j, \sigma_{\eta_j}\le j\}} | X_0^{\eta_j}) - E(\sigma_{\eta_j} | X_0^{\eta_j}) \right) \to 0$$

almost surely.

Now we deal with the last term D_n. By (20) and the Toeplitz lemma

$$D_n \to 0$$

almost surely.

The proof of the result for $\hat{q}_l(X_0^{\eta_j})$ is even simpler.

$$\begin{aligned} & \hat{q}_l(X_0^{\eta_j}) - P(\sigma_{\eta_n} = l | X_0^{\eta_n}) \\ = \; & \frac{1}{n} \sum_{j=0}^{n-1} \left(I_{\{\sigma_{\eta_j}=l, \sigma_{\eta_j}\le \eta_{j+1}-\eta_j\}} - P(\sigma_{\eta_j} = l, \sigma_{\eta_j} \le \eta_{j+1} - \eta_j | X_0^{\eta_j}) \right) \\ + \; & \frac{1}{n} \sum_{j=0}^{n-1} \left(P(\sigma_{\eta_j} = l, \sigma_{\eta_j} \le \eta_{j+1} - \eta_j | X_0^{\eta_j}) - P(\sigma_{\eta_j} = l | X_0^{\eta_j}) \right) \\ + \; & \left(\frac{1}{n} \sum_{j=0}^{n-1} P(\sigma_{\eta_j} = l | X_0^{\eta_j}) \right) - P(\sigma_{\eta_n} = l | X_0^{\eta_n}) \\ = \; & E_n + F_n + G_n. \end{aligned}$$

The first term is an average of bounded martingale differences and so $E_n \to 0$ almost surely. Since $\sigma_{\eta_j} \le \eta_{j+1} - \eta_j$ as soon as there is at least one zero in $X_0^{\eta_j}$ so $F_n \to 0$ almost surely. Concerning the last term, by Step 2, Step 1 and Step 3,

$$P(\sigma_{\eta_j} = l | X_0^{\eta_j}) = P(\sigma_0(\tilde{X}_1^{\infty}) = l | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}).$$

The latter forms a martingale and so

$$P(\sigma_{\eta_j} = l | X_0^{\eta_j}) = P(\sigma_0(\tilde{X}_1^{\infty}) = l | \tilde{X}^0_{\hat{\eta}_j(\tilde{X}^0_{-\infty})}) \to P(\sigma_0(\tilde{X}_1^{\infty}) = l | \tilde{X}^0_{-\infty}) \tag{21}$$

almost surely and so $G_n \to 0$ almost surely. Since both $\hat{q}_l(X_0^{\eta_j})$ (as soon as there is at least one zero in $X_0^{\eta_j}$)and $P(\sigma_0(\tilde{X}_1^{\infty}) = l | \tilde{X}^0_{-\infty})$ are probability distributions

$$\lim_{n\to\infty} \sum_{l=0}^{\infty} |\hat{q}_l(X_0, \ldots, X_{\eta_n}) - P(\sigma_{\eta_n} = l | X_0^{\eta_n}))| = 0$$

almost surely.

Step 5. *We show that if the process is not a renewal process then our schemes are consistent.*

If the process is not a renewal process then for some $k \ge 0$, $i \ge 1$, $z_{-k-i}^{-k-1} \in \{0,1\}^i$, $P(X_{-k-i}^{-k-1} = z_{-k-i}^{-k-1}, X_{-k}^0 = 01^k) > 0$:

$$\begin{aligned} &P(X_1 = 0 | X_{-k}^0 = 01^k) \\ &\quad \neq \; P(X_1 = 0 | X_{-k-i}^{-k-1} = z_{-k-i}^{-k-1}, X_{-k}^0 = 01^k) \end{aligned}$$

which in turn implies that

$$\liminf_{n\to\infty} \hat{\Delta}_n > 0 \quad \text{almost surely}$$

and so

$$\hat{\Delta}_n > n^{-\beta} \quad \text{eventually almost surely.}$$

Thus eventually, we will use m_n and $\hat{q}_l(X_0^{\eta_j})$ on the stopping times η_n and by Step 4 the scheme is consistent.

Step 6. *For renewal processes we show that our schemes $h_i^*(X_0^{\xi_i^*})$, $\hat{q}_l^*(X_0^{\xi_i^*})$ are consistent and that $\lim_{n\to\infty} \frac{\xi_n}{n} = 1$ almost surely.*

We have to check the conditions of Theorem 2 in Morvai and Weiss [9]. For shorthand, let $p_h = P(\sigma_0 = h | X_0 = 0)$. Then it is easy to see that

$$E(|\sigma_0|^\alpha) = \sum_{L=0}^{\infty} \frac{\sum_{h=0}^{\infty} h^\alpha p_{h+L}}{\sum_{h=L}^{\infty} p_h} \frac{\sum_{h=L}^{\infty} p_h}{1 + \sum_{h=0}^{\infty} h p_h}$$

(cf. the proof of Theorem 2 in Morvai and Weiss [9]). Now, by assumption, for $\alpha > 1$

$$\begin{aligned}
\infty > \sum_{L=0}^{\infty}\sum_{h=0}^{\infty} h^{\alpha} p_{h+L} &= \sum_{h=0}^{\infty}(1+2^{\alpha}+\dots h^{\alpha})p_h \\
&\geq \sum_{h=2}^{\infty}(2^{\alpha}+\dots h^{\alpha})p_h \\
&\geq \sum_{h=2}^{\infty}\Big(\int_1^2 x^{\alpha}dx+\dots\int_{h-1}^{h} x^{\alpha}\,\mathrm{d}x\Big)p_h \\
&= \sum_{h=2}^{\infty}\Big(\int_1^h x^{\alpha}\,\mathrm{d}x\Big)p_h \\
&= \sum_{h=2}^{\infty}\Big(\frac{h^{\alpha+1}p_h}{\alpha+1}-\frac{p_h}{\alpha+1}\Big) \\
&\geq \sum_{h=2}^{\infty}\frac{h^{\alpha+1}p_h}{\alpha+1}-1.
\end{aligned}$$

Thus the condition in Theorem 2 in Morvai and Weiss [9]

$$\sum_{h=0}^{\infty} h^{\alpha+1}p_h < \infty$$

is satisfied. By Theorem 2 in Morvai and Weiss [9] our schemes $h_i^*(X_0^{\xi_i^*})$, $\hat{q}_l^*(X_0^{\xi_i^*})$ are consistent and

$$\limsup_{n\to\infty}\frac{\xi_n}{n}=1.$$

Step 7. *We prove that $\hat{\Delta}_n \leq n^{-\beta}$ eventually almost surely for renewal processes.*

Set $\theta^+_{l,j,0}=0$, $\theta^-_{l,j,0}=0$ and define

$$\theta^+_{l,j,i}=\theta^+_{l,j,i-1}+\min\Big\{t>0: X_{l+\theta^+_{l,j,i-1}-j+1+t}^{l+\theta^+_{l,j,i-1}+t}=X_{l+\theta^+_{l,j,i-1}-j+1}^{l+\theta^+_{l,j,i-1}}\Big\}$$

and

$$\theta^-_{l,j,i}=\theta^-_{l,j,i-1}+\min\Big\{t>0: X_{l-\theta^-_{l,j,i-1}-j+1-t}^{l-\theta^-_{l,j,i-1}-t}=X_{l-\theta^-_{l,j,i-1}-j+1}^{l-\theta^-_{l,j,i-1}}\Big\}.$$

Define $p(0|w^0_{-q})=P(X_1=0|X^0_{-q}=w^0_{-q})$. Assume that the process is a binary renewal process. Then for $r,s\geq 0$,

$$X_{l-\theta^-_{l,j,r}+1},\dots,X_{l-\theta^-_{l,j,1}+1},X_{l+\theta^+_{l,j,0}+1},\dots,X_{l+\theta^+_{l,j,s}+1}$$

are conditionally independent and identically distributed random variables for any fixed $j\geq k$ given $\tau_l=k$, where the identical distribution is $p(0|01^k)$.

The following argument is identical to the one in [10] but we repeat it for the sake of completeness. Observe that

$$
\begin{aligned}
&P(\hat{\Delta}_n > n^{-\beta})\\
&\quad = P\Big(\max_{0\le k\le n-1}\max_{1\le i\le n-k}\max_{z^0_{-k-i}\in\mathcal{S}^n_{k,i}}\left|\hat{p}_n(0|z^0_{-k})-\hat{p}_n(0|z^0_{-k-i})\right| > n^{-\beta}\Big)\\
&\quad \le P\Big(\max_{0\le k\le n-1}\max_{1\le i\le n-k}\max_{z^0_{-k-i}\in\mathcal{S}^n_{k,i}}\left|\hat{p}_n(0|z^0_{-k})-p(0|z^0_{-k})\right| > 0.5n^{-\beta}\Big)\\
&\quad + P\Big(\max_{0\le k\le n-1}\max_{1\le i\le n-k}\max_{z^0_{-k-i}\in\mathcal{S}^n_{k,i}}\left|p(0|z^0_{-k})-\hat{p}_n(0|z^0_{-k-i})\right| > 0.5n^{-\beta}\Big)\\
&\quad \le P\Big(\max_{0\le k\le n-1}\max_{z^0_{-k}\in\mathcal{S}^n_{k,0}}\left|\hat{p}_n(0|z^0_{-k})-p(0|z^0_{-k})\right| > 0.5n^{-\beta}\Big)\\
&\quad + P\Big(\max_{0\le k\le n}\max_{1\le i\le n-k}\max_{z^0_{-k-i}\in\mathcal{S}^n_{k,i}}\left|p(0|z^0_{-k})-\hat{p}_n(0|z^0_{-k-i})\right| > 0.5n^{-\beta}\Big)\\
&\quad \le P\left(\exists k\ge 0, k\le l\le n-1 : X^l_{l-k}\in\mathcal{S}^n_{k,0}, \left|\hat{p}_n(0|X^l_{l-k})-p(0|X^l_{l-k})\right| > 0.5n^{-\beta}\right)\\
&\quad + P\left(\exists k\ge 0, i\ge 1, k+i\le l< n : X^l_{l-k-i}\in\mathcal{S}^n_{k,i}, \left|\hat{p}_n(0|X^l_{l-k-i})-p(0|X^l_{l-k})\right| > 0.5n^{-\beta}\right)\\
&\quad \le \sum_{i=0}^{n-1} P\left(\exists k\ge 0, k+i\le l< n : X^l_{l-k-i}\in\mathcal{S}^n_{k,i}, \left|\hat{p}_n(0|X^l_{l-k-i})-p(0|X^l_{l-k})\right| > 0.5n^{-\beta}\right)\\
&\quad \le \sum_{i=0}^{n-1}\sum_{l=0}^{n-1} P\left(X^l_{l-\tau_l-i}\in\mathcal{S}^n_{\tau_l,i}, \left|\hat{p}_n(0|X^l_{l-\tau_l-i})-p(0|X^l_{l-\tau_l})\right| > 0.5n^{-\beta}\right).
\end{aligned}
$$

For a given $0\le l\le n-1$ assume that $\tau_l = k$. By Hoeffding's inequality for sums of bounded independent random variables,

$$
\begin{aligned}
P\Big(\Big|\frac{1}{r+s+1}\Big(\sum_{h=1}^{r} 1_{\{X_{l-\theta^-_{l,k+i,h}+1}=0\}} + \sum_{h=0}^{s} 1_{\{X_{l+\theta^+_{l-1,k+i,h}+1}=0\}}\Big)\\
-p(0|01^k)\Big| \ge 0.5n^{-\beta} \Big| \tau_l = k\Big) \le 2e^{-0.5n^{-2\beta}(r+s+1)}.
\end{aligned}
$$

Multiplying both sides by $P(\tau_l = k)$ and summing over all possible k we get that

$$
\begin{aligned}
&P\Big(X^l_{l-\tau_l-i}\in\mathcal{S}^n_{\tau_l,i},\\
&\qquad \Big|\frac{1}{r+s+1}\Big(\sum_{h=1}^{r} 1_{\{X_{l-\theta^-_{l,\tau_l+i,h}+1}=0\}} + \sum_{h=0}^{s} 1_{\{X_{l+\theta^+_{l,\tau_l+i,h}+1}=0\}}\Big)\\
&\quad - \quad p(0|X^l_{l-\tau_l})\Big| > 0.5n^{-\beta}\Big)\\
&\quad \le \quad P\Big(\Big|\frac{1}{r+s+1}\Big(\sum_{h=1}^{r} 1_{\{X_{l-\theta^-_{l,\tau_l+i,h}+1}=0\}} + \sum_{h=0}^{s} 1_{\{X_{l+\theta^+_{l,\tau_l+i,h}+1}=0\}}\Big)\\
&\quad - \quad p(0|X^l_{l-\tau_l})\Big| > 0.5n^{-\beta}\Big)\\
&\quad \le \quad 2e^{-0.5n^{-2\beta}(r+s+1)}.
\end{aligned}
$$

Summing over all $0 \le l \le n-1$ and over all pairs (r,s) such that $r \ge 0$, $s \ge 0$, $r+s+1 \ge \lfloor n^{1-\gamma} \rfloor$ we get that

$$\begin{aligned}&\sum_{l=0}^{n-1} P\left(X_{l-\tau_l-i}^{l} \in \mathcal{S}_{\tau_l,i}^{n},\right.\\ &\qquad\left.\left|\hat{p}_n(0|X_{l-\tau_l-i}^{l}) - p(0|X_{l-\tau_l}^{l})\right| > 0.5n^{-\beta}\right)\\ &\le \quad n \sum_{h=\lfloor n^{1-\gamma}\rfloor}^{\infty} h2e^{-0.5n^{-2\beta}h}.\end{aligned}$$

Applying this final inequality to (22) we get that

$$P(\hat{\Delta}_n > n^{-\beta}) \le 2n^2 \sum_{h=\lfloor n^{1-\gamma}\rfloor}^{\infty} he^{-0.5n^{-2\beta}h}.$$

The sum on the right hand side is bounded by a constant times the first term and since $0 < \beta < \frac{1-\gamma}{2}$ and thus as n varies the right hand side is a convergent series and by the Borel–Cantelli lemma eventually almost surely we will have that:

$$\hat{\Delta}_n \le n^{-\beta}.$$

Step 8. *We prove the statements of our Theorem.*

By Step 5 and 7 we stick to the right estimators and stopping times and so by Step 5, 6 and 7 our schemes are consistent and if the process is a renewal process then

$$1 \le \limsup_{n\to\infty} \frac{\lambda_n}{n} = \limsup_{n\to\infty} \frac{\xi_n}{n} = 1.$$

This completes the proof of the Theorem.

For further reading on related results we refer the interested reader to [1, 2, 3, 7, 8, 9, 12, 13, 15, 16].

ACKNOWLEDGEMENT

The first author was supported by OTKA Grant No. K75143 and Bolyai János Research Scholarship.

(Received August 9, 2013)

REFERENCES

[1] P. Algoet: Universal schemes for learning the best nonlinear predictor given the infinite past and side information. IEEE Trans. Inform. Theory *45* (1999), 1165–1185.

[2] D. H. Bailey: Sequential Schemes for Classifying and Predicting Ergodic Processes. Ph. D. Thesis, Stanford University 1976.

[3] F. Bunea and A. Nobel: Sequential procedures for aggregating arbitrary estimators of a conditional mean. IEEE Trans. Inform. Theory *54* (2008), 4, 1725–1735.

[4] W. Feller: An Introduction to Probability Theory and its Applications. Vol. II. Second edition. John Wiley and Sons, New York – London – Sydney 1971.

[5] L. Györfi, G. Morvai, and S. Yakowitz: Limits to consistent on-line forecasting for ergodic time series. IEEE Trans. Inform. Theory *44* (1998), 886–892.

[6] G. Morvai and B. Weiss: Inferring the conditional mean. Theory Stoch. Process. *11* (2005), 112–120.

[7] G. Morvai and B. Weiss: Order estimation of Markov chains. IEEE Trans. Inform. Theory *51* (2005), 1496–1497.

[8] G. Morvai and B. Weiss: On sequential estimation and prediction for discrete time series. Stoch. Dyn. *7* (2007), 4, 417–437.

[9] G. Morvai and B. Weiss: On Universal estimates for binary renewal processes. Ann. Appl. Probab. *18* (2008), 5, 1970–1992.

[10] G. Morvai and B. Weiss: Estimating the residual waiting time for binary stationary time series. In: Proc. ITW2009, Volos 2009, pp. 67–70.

[11] G. Morvai and B. Weiss: A note on prediction for discrete time series. Kybernetika *48* (2012), 4, 809–823.

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

[13] D. Ryabko and B. Ryabko: Nonparametric statistical inference for ergodic processes. IEEE Trans. Inform. Theory *56* (2010), 3, 1430–1435.

[14] A. N. Shiryayev: Probability. Springer-Verlag, New York 1984.

[15] H. Takahashi: Computational limits to nonparametric estimation for ergodic processes. IEEE Trans. Inform. Theory *57* (2011), 10, 6995–6999.

[16] Z. Zhou, Z. Xu, and W. B. Wu: Long-term prediction intervals of time series. IEEE Trans. Inform. Theory *56* (2010), 3, 1436–1446.

Gusztáv Morvai, Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences Reáltanoda utca 13-15, 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