

# Estimating the Residual Waiting Time for Binary Stationary Time Series

Appeared in:

ITW 2009. IEEE Information Theory Workshop on Networking and Information Theory, 10-12 June 2009 pp. 67 - 70, 2009.

Gusztáv Morvai

MTA-BME Stochastics Research Group  
1 Egrý József utca , Building H, Budapest,1111, Hungary  
Email: morvai@math.bme.hu

Benjamin Weiss

Hebrew University of Jerusalem  
Jerusalem 91904 Israel  
Email: weiss@math.huji.ac.il

**Abstract**—We present here a universal estimation scheme for the problem of estimating the residual waiting time until the next occurrence of a zero after observing the first  $n$  outputs of a stationary and ergodic binary process. The scheme will involve estimating only at carefully selected stopping times but will be almost surely consistent. In case the process happens to be a genuine renewal process then our stopping times will have asymptotic density one.

## I. 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 waiting time to the state 0 given some previous observations, in particular given  $X_0^n$ .

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

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

Our goal is to estimate  $E(\sigma_n | X_0^n)$  but without prior knowledge of the distribution function of the process. We would like our estimation scheme to be almost surely eventually consistent and for this we must give up the attempt to give an estimate for all time instants  $n$ , this explains why we need to introduce stopping times. For renewal processes we can find such a scheme where the asymptotic density of the stopping times is one, but in order to account for all ergodic processes we will also need to use stopping times that are much rarer. Indeed our scheme will be composed of two different kinds of stopping times and tests, one of which will work for the renewal processes and the other will be a universal one. A third auxiliary test will enable us to decide which scheme should be used.

Here is the first auxiliary algorithm. 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:

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

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

Define  $\psi$  as the position of the first zero, that is,

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

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

$$\xi_n = \min\{k > \xi_{n-1} :$$

$$|\{\psi \leq i < k : \tau_i(X_0^i) = \tau_k(X_0^k)\}| \geq k^{1-\delta}\}.$$

These are the successive times  $i$  when the value  $t = \tau_i(X_0^i)$  has occurred previously enough times so that we can safely estimate the residual renewal time by empirical distributions derived from observations already made. We also need to fix  $\kappa_n$  as the index where reading backwards from  $X_{\xi_n}$  will have seen for the first time  $\geq \xi_n^{1-\delta}$  occurrences of an  $i$  with  $\tau_i(X_0^i) = \tau_n(X_0^{\xi_n})$ . Formally put

$$\kappa_n =$$

$$\max\{K : |\{K \leq k < \xi_n : \tau_k(X_0^k) = \tau_{\xi_n}(X_0^{\xi_n})\}| = \lceil \xi_n^{1-\delta} \rceil\}$$

For  $n > 0$  define our estimator  $h_n(X_0, \dots, X_{\xi_n})$  at time  $\xi_n$  as

$$h_n(X_0, \dots, X_{\xi_n}) = \frac{1}{\lceil (\xi_n)^{1-\delta} \rceil} \sum_{i=\kappa_n}^{\xi_n-1} I_{\{\tau_i(X_0^i) = \tau_{\xi_n}(X_0^{\xi_n})\}} \sigma_i.$$

(Notice that the role of  $\kappa_n$  is rather technical. It ensures that we take into consideration exactly  $\lceil (\xi_n)^{1-\delta} \rceil$  pieces of occurrences.) The above formula is simply the average of the residual waiting times that we have already observed in the data segment  $(X_{\kappa_n}, \dots, X_{\xi_n})$  when we were at the same value of  $\tau$  as we see at time  $\xi_n$ .

Note that both  $h_n$  and  $\xi_n$  depend on  $\delta$ .

Now we define the second auxiliary algorithm which will work for general ( not necessarily renewal ) processes. We

define the stopping times  $\{\eta_n\}$ . Set  $\eta_0 = 0$ . For  $n = 1, 2, \dots$ , define  $\eta_n$  recursively. Let

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

Note that  $\eta_n \geq n$  and it is a stopping time on  $X_0^\infty$ . The  $n$ th estimate  $m_n$  is defined as

$$m_n(X_0, \dots, X_{\eta_n}) = \frac{1}{n} \sum_{j=0}^{n-1} \min(\sigma_{\eta_j}, \eta_{j+1} - \eta_j).$$

Observe that  $m_n$  depends 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. Define the empirical conditional distributions by

$$\hat{p}_n(0|z_{-k}^0) = \frac{\#\{k \leq t \leq n-1 : X_{t-k}^{t+1} = (z_{-k}^0, 0)\}}{\#\{k \leq t \leq n-1 : X_{t-k}^t = z_{-k}^0\}}.$$

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  $01^k$  and then for a fixed  $0 < \gamma < 1$  define  $\mathcal{S}_{k,i}^n$  as

$$\mathcal{S}_{k,i}^n = \{x_{-k-i}^0 \in \mathcal{X}_{k,i} : \#\{k+i \leq t \leq n-1 : X_{t-k-i}^t = x_{-k-i}^0\} > n^{1-\gamma}\}.$$

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

Define

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

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

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

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  $n \geq 1$  define

$$\lambda_n = \min \{ \eta_{\min\{i: \hat{\Delta}_{\eta_i} > \eta_i^{-\beta}, \eta_i > \lambda_{n-1}\}}, \xi_{\min\{i: \hat{\Delta}_{\xi_i} \leq \xi_i^{-\beta}, \xi_i > \lambda_{n-1}\}} \}.$$

The  $n$ th estimate  $f_n$  is defined as

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

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 [14].

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

$$\lim_{n \rightarrow \infty} |f_n(X_0^{\lambda_n}) - E(\sigma_{\lambda_n} | X_0^{\lambda_n})| = 0.$$

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

$$\lim_{n \rightarrow \infty} \frac{\lambda_n}{n} = 1.$$

## II. PROOF OF THE THEOREM

We will divide the proof into several steps. In the first four steps, as in Morvai and Weiss [6], a stationary process is defined in order to reduce the analysis of the  $m_n$ 's to a backward scheme. The next two steps are devoted to showing why our algorithm is consistent for non-renewal processes while in the last step we do the same for renewal processes.

**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  $F \subseteq \{0, 1\}^{n+1}$ ,

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

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,

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

and

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

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 algorithm  $m_n$ .

Similarly to Morvai and Weiss [6], consider

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

Observe that

$$\{\Gamma_j = \min(\sigma_{\eta_j}, \eta_{j+1} - \eta_j) - E(\min(\sigma_{\eta_j}, \eta_{j+1} - \eta_j)|X_0^{\eta_j})\}$$

is a sequence of orthogonal random variables with  $E\Gamma_j = 0$  and  $E(\Gamma_j^2) \leq 2E(|\sigma_0|^2)$  since  $E(\Gamma_j^2) \leq 2E(|\sigma_{\eta_j}|^2)$  and, by Step 1,  $\sigma_{\eta_j}$  has the same distribution as  $\sigma_0$ . Now by Theorem 3.2.2 in Révész [17],

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

Now we deal with the second term. Since  $\sigma_{\eta_j} \leq \eta_{j+1} - \eta_j$  as soon as there is at least one zero in  $X_0^{\eta_j}$

$$\frac{1}{n} \sum_{j=0}^{n-1} |E(\min(\sigma_{\eta_j}, \eta_{j+1} - \eta_j)|X_0^{\eta_j}) - E(\sigma_{\eta_j}|X_0^{\eta_j})| \rightarrow 0$$

almost surely.

Now we deal with the last term. By Step 2, Step 1 and Step 3,

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

The latter forms a martingale and so

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

almost surely.

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

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

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

which in turn implies that

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

and so

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

Thus eventually, we will use  $m_n$  on the stopping times  $\eta_n$  and by Step 4 the scheme is consistent.

**Step 6.** We show the consistency of our scheme for renewal processes.

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

$$\theta_{l,j,i}^+ = \theta_{l,j,i-1}^+ +$$

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

and

$$\theta_{l,j,i}^- = \theta_{l,j,i-1}^- +$$

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

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

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

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

Observe that

$$\begin{aligned}
& P(\hat{\Delta}_n > n^{-\beta}) \\
&= P\left(\max_{0 \leq k \leq n-1} \max_{1 \leq i \leq n-k} \max_{z_{k-i}^0 \in \mathcal{S}_{k,i}^n} \left| \hat{p}_n(0|z_{-k}^0) - \hat{p}_n(0|z_{-k-i}^0) \right| > n^{-\beta}\right) \\
&\leq P\left(\max_{0 \leq k \leq n-1} \max_{1 \leq i \leq n-k} \max_{z_{k-i}^0 \in \mathcal{S}_{k,i}^n} \left| \hat{p}_n(0|z_{-k}^0) - p(0|z_{-k}^0) \right| > 0.5n^{-\beta}\right) \\
&+ P\left(\max_{0 \leq k \leq n-1} \max_{1 \leq i \leq n-k} \max_{z_{k-i}^0 \in \mathcal{S}_{k,i}^n} \left| p(0|z_{-k}^0) - \hat{p}_n(0|z_{-k-i}^0) \right| > 0.5n^{-\beta}\right) \\
&\leq P\left(\max_{0 \leq k \leq n-1} \max_{z_{k,0}^0 \in \mathcal{S}_{k,0}^n} \left| \hat{p}_n(0|z_{-k}^0) - p(0|z_{-k}^0) \right| > 0.5n^{-\beta}\right) \\
&+ P\left(\max_{0 \leq k \leq n-1} \max_{1 \leq i \leq n-k} \max_{z_{k-i}^0 \in \mathcal{S}_{k,i}^n} \left| p(0|z_{-k}^0) - \hat{p}_n(0|z_{-k-i}^0) \right| > 0.5n^{-\beta}\right) \\
&\leq P\left(\text{For some } k \geq 0, k \leq l \leq n-1 : X_{l-k}^l \in \mathcal{S}_{k,0}^n, \left| \hat{p}_n(0|X_{l-k}^l) - p(0|X_{l-k}^l) \right| > 0.5n^{-\beta}\right) \\
&+ P\left(\text{For some } k, i \geq 0, k+i \leq l \leq n-1 : X_{l-k-i}^l \in \mathcal{S}_{k,i}^n, \left| \hat{p}_n(0|X_{l-k-i}^l) - p(0|X_{l-k-i}^l) \right| > 0.5n^{-\beta}\right) \\
&\leq \sum_{i=0}^{n-1} P\left(\text{For some } k \geq 0, k+i \leq l \leq n-1 : X_{l-k-i}^l \in \mathcal{S}_{k,i}^n, \left| \hat{p}_n(0|X_{l-k-i}^l) - p(0|X_{l-k-i}^l) \right| > 0.5n^{-\beta}\right) \\
&\leq \sum_{i=0}^{n-1} \sum_{l=0}^{n-1} P\left(X_{l-\tau_l-i}^l \in \mathcal{S}_{\tau_l,i}^n, \left| \hat{p}_n(0|X_{l-\tau_l-i}^l) - p(0|X_{l-\tau_l-i}^l) \right| > 0.5n^{-\beta}\right). \tag{1}
\end{aligned}$$

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

$$P\left(\frac{\sum_{h=1}^r \mathbf{1}_{\{X_{l-\theta_{l,k+i,h}^-} = 0\}} + \sum_{h=0}^s \mathbf{1}_{\{X_{l+\theta_{l-1,k+i,h}^+} = 0\}}}{r+s+1} - p(0|01^k)\right) \geq 0.5n^{-\beta} | \tau_l = k \leq 2e^{-0.5n^{-2\beta}(r+s+1)}.$$

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

$$\begin{aligned}
& P\left(X_{l-\tau_l-i}^l \in \mathcal{S}_{\tau_l,i}^n, \left| \frac{\sum_{h=1}^r \mathbf{1}_{\{X_{l-\theta_{l,\tau_l+i,h}^-} = 0\}} + \sum_{h=0}^s \mathbf{1}_{\{X_{l+\theta_{l-1,\tau_l+i,h}^+} = 0\}}}{r+s+1} - p(0|X_{l-\tau_l}^l) \right| > 0.5n^{-\beta}\right) \\
&\leq P\left(\frac{\sum_{h=1}^r \mathbf{1}_{\{X_{l-\theta_{l,\tau_l+i,h}^-} = 0\}} + \sum_{h=0}^s \mathbf{1}_{\{X_{l+\theta_{l-1,\tau_l+i,h}^+} = 0\}}}{r+s+1} - p(0|X_{l-\tau_l}^l) \right) > 0.5n^{-\beta} \\
&\leq 2e^{-0.5n^{-2\beta}(r+s+1)}.
\end{aligned}$$

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

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

Applying this final inequality to (1) we get that

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

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:

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

Thus eventually we stick to estimator  $h_n$  and stopping times  $\xi_n$  and so by Theorem 1 in Morvai and Weiss [14] our scheme is consistent and

$$1 \leq \limsup_{n \rightarrow \infty} \frac{\lambda_n}{n} = \limsup_{n \rightarrow \infty} \frac{\xi_n}{n} = 1.$$

This completes the proof of the Theorem.

For more on estimation on stopping times see [2], [3], [4], [6], [10], [12], [14]. For further reading on related results we refer the interested reader to [1], [5], [7], [8], [11], [9], [13], [18], [15], [16].

#### ACKNOWLEDGMENT

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

#### REFERENCES

- [1] D. H. Bailey, *Sequential Schemes for Classifying and Predicting Ergodic Processes*. Ph. D. thesis, Stanford University, 1976.
- [2] G. Morvai "Guessing the output of a stationary binary time series" In: *Foundations of Statistical Inference*, (Eds. Y. Haitovsky, H.R.Lerche, Y. Ritov), Physika-Verlag, pp. 207-215, 2003.
- [3] G. Morvai, B. Weiss, "Forecasting for stationary binary time series" *Acta Applicandae Mathematicae*, vol. 79, 25-34, 2003.
- [4] G. Morvai, B. Weiss, "Intermittent estimation of stationary time series" *Test*, vol. 13, 525-542, 2004.
- [5] G. Morvai, B. Weiss, "Forward estimation for ergodic time series" *Ann. I.H.Poincaré Probabilités et Statistiques*, vol. 41, 859-870, 2005.
- [6] G. Morvai, B. Weiss, "Inferring the conditional mean" *Theory of Stochastic Processes*, vol. 11, 112-120, 2005.
- [7] G. Morvai, B. Weiss, "On classifying processes" *Bernoulli*, vol. 11, 523-532, 2005.
- [8] G. Morvai, B. Weiss, "Order estimation of Markov chains," *IEEE Transactions on Information Theory*, vol. 51, pp. 1496-1497, 2005.
- [9] G. Morvai, B. Weiss, "Limitations on intermittent forecasting" *Statistics and Probability Letters*, vol. 72, 285-290, 2005.
- [10] G. Morvai, B. Weiss, "Prediction for discrete time series" *Probability Theory and Related Fields*, vol. 132, 1-12, 2005.
- [11] G. Morvai, B. Weiss, "Estimating the memory for finitarily Markovian processes" *Ann. I.H.Poincaré Probabilités et Statistiques*, vol. 43, pp. 15-30, 2007.

- [12] G. Morvai, B. Weiss, "On sequential estimation and prediction for discrete time series " *Stochastics and Dynamics*, vol. 7, no. 4, pp. 417-437, 2007.
- [13] G. Morvai, B. Weiss, "Estimating the Lengths of Memory Words," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3804-3807, 2008.
- [14] G. Morvai, B. Weiss, "On Universal Estimates for Binary Renewal Processes," *Annals of Applied Probability*, vol. 18, No. 5, pp. 1970-1992, 2008.
- [15] G. Morvai, S. Yakowitz, and P. Algoet, "Weakly convergent non-parametric forecasting of stationary time series," *IEEE Transactions on Information Theory*, vol. 43, pp. 483-498, 1997.
- [16] G. Morvai, S. Yakowitz, and L. Györfi, "Nonparametric inferences for ergodic, stationary time series," *Annals of Statistics.*, vol. 24, pp. 370-379, 1996.
- [17] P. Révész, *The Law of Large Numbers*, Academic Press, 1968.
- [18] B. Ya. Ryabko, "Prediction of random sequences and universal coding," *Problems of Inform. Trans.*, vol. 24, pp. 87-96, Apr.-June 1988.