Kybernetika

Gusztáv Morvai; Benjamin Weiss

Universal rates for estimating the residual waiting time in an intermittent way

Kybernetika, Vol. 56 (2020), No. 4, 601–616

Persistent URL: `http://dml.cz/dmlcz/148374`

KYBERNETIKA — VOLUME *56* (2020), NUMBER 4, PAGES 601–616

UNIVERSAL RATES FOR ESTIMATING THE RESIDUAL WAITING TIME IN AN INTERMITTENT WAY

Gusztáv Morvai and Benjamin Weiss

A simple renewal process is a stochastic process $\{X_n\}$ taking values in $\{0,1\}$ where the lengths of the runs of 1's between successive zeros are independent and identically distributed. After observing $X_0, X_1, \dots X_n$ one would like to estimate the time remaining until the next occurrence of a zero, and the problem of universal estimators is to do so without prior knowledge of the distribution of the process. We give some universal estimates with rates for the expected time to renewal as well as for the conditional distribution of the time to renewal.

Keywords: statistical learning, statistical inference, prediction methods, renewal theory

Classification: 60G25, 60K05

1. INTRODUCTION

A simple renewal process is a stochastic process $\{X_n\}$ taking values in $\{0,1\}$ where the lengths of the runs of 1's between successive zeros are independent. These processes arise in the study of Markov chains since the successive visits to a fixed state form such a renewal process, cf. [4]. There are applications in which the failure times of some system which is renewed after each failure form such a renewal process and so the problem naturally arises of estimating when will the next failure occur cf. Ex. 12.13 in [5]. This may be formalized in the following way, we define σ_n to be the residual waiting to the next occurrence of a zero after observing $\{X_0, X_1, \dots X_n\}$, and we wish to estimate σ_n. If the distribution of the process is known then the best least square estimator for σ_n would be θ_n, its conditional expectation given the observations $\{X_0, X_1, \dots X_n\}$. We plan to give here a universal estimator, i. e., one in which we learn enough about the process in order to give a function of these observations which will almost surely in the limit of large n be as good as θ_n.

Prior works on estimating the parameters of an unknown renewal processes have treated rather different problems. The often quoted paper of Vardi [24] treats the problem of finding a maximum likelihood estimate (MLE) of the discrete lifetime distribution on the basis of data collected from independent identically distributed stationary renewal processes. The key point is that the data is being collected from the outputs of distinct independent sources, whereas we are trying to predict the next event in a single sample

DOI: 10.14736/kyb-2020-4-0601

of a renewal process in the spirit of sequential estimation. (For the idea of sequential estimation see e. g. Ryabko [20], Algoet [1], Györfi and Ottucsák [6] and Nobel [17]). The same applies to more recent papers on estimating the lifetime distribution such as Peña, Strawderman and Hollander [18] and Denby and Vardi [3].

Before stating our results we recall the basic framework. The basic parameters of these renewal processes are the $\{p_k\}_{k=0}^{\infty}$, the conditional probability that a run of k 1's follows a given 0 event. In order that the probability of $X_0 = 0$ be nonzero it is necessary that $\mu = \sum_{k=0}^{\infty} kp_k < \infty$ and then $P(X_0 = 0) = 1/(1 + \mu)$ is positive. (This relation between the mean of the conditional renewal distribution and the stationary probability of the renewal event is well known in ergodic theory as Kac's formula for the expected return time to a set, and in probability theory cf. [4] Ch. XIII and [21] Sec. I.2.c.) If the process distribution is known, then after observing $X_0, X_1, \ldots, X_n$ one may calculate the expected value of the residual waiting time to the occurrence of the next zero as

$$\mu_L = \frac{\sum_{k=L}^{\infty}(k - L)p_k}{\sum_{k=L}^{\infty} p_k}$$

if there is at least one zero among the values of $X_0, X_1, \ldots, X_n$ and the last zero occurs at moment $X_{n-L} = 0$. We denote this L by $\tau(X_0, X_1, \ldots, X_n)$.

Similarly, we define $\tau = \tau(X_{-\infty}^0)$ as that $t \geq 0$ such that $X_{-t} = 0$ and $X_i = 1$ for all $-t < i \leq 0$. It is clear from the stationarity that $P(\tau = L)$ is proportional to $\sum_{k=L}^{\infty} p_k$ and thus for the finiteness of the expectation of the residual waiting time we have to demand that $\sum_{k=0}^{\infty} k^2 p_k < \infty$. The moment conditions that we shall impose are just a bit more - namely $2 + \epsilon$ with $\epsilon > 0$.

In our work [13] we took up the problem of how well can we do when all that we know is that the binary process $\{X_n\}$ is, in fact, a renewal process. Our purpose in this note is to give a simpler scheme which utilizes the observed data in a more efficient way. In addition, we are able to provide a universal rate of decay to zero of the deviation of our scheme from the optimal estimate which uses complete knowledge about the process.

The fact, that we are trying to estimate the time to next occurrence of zero rather than X_{n+1}, takes us out of the framework of most previous investigations [9, 10, 11, 12]. In earlier works such as [7] attention is restricted to those renewal processes which arise from Markov chains with a finite number of states. In that case the probabilities p_k decay exponentially and one can use this information in trying to find not only the distribution but even the hidden Markov chain itself. Since we are considering the general case where the number of hidden states might be infinite this exponential decay no longer holds in general and the problem becomes much more difficult.

For the estimator itself it is most natural to use the empirical distribution observed in the data segment $X_0, X_1, \ldots, X_n$. However if there were an insufficient number of occurrences of 1-blocks of length at least $\tau(X_0, X_1, \ldots, X_n)$ then we do not expect the empirical distribution to be close to the true distribution. In particular, if no block of that length has occurred yet, clearly no intelligent estimate can be given cf. Theorem 4.1 in [16]. For this reason we will estimate only along stopping times $\lambda_1, \lambda_2, \ldots$ and our main positive result is that there is a sequence of universally defined stopping times λ_n with density one, estimators $h_n(X_0, X_1, \ldots, X_{\lambda_n})$, and a sequence r_n that converges to zero, so that eventually almost surely:

$$|\mu_{\tau(X_0,X_1,\ldots X_{\lambda_n})} - h_n(X_0, X_1, \ldots, X_{\lambda_n})| < r_n.$$

We denote by $q_l(X_0, X_1, \ldots X_n)$ the conditional probability $P(\sigma_n = l|X_0, X_1, \ldots \ldots, X_n)$. We will also define universal estimators $\hat{p}_l(X_0, X_1, \ldots, X_{\lambda_n})$, so that eventually almost surely:

$$\sum_{l=0}^{\infty} |\hat{p}_l(X_0, X_1, \ldots, X_{\lambda_n}) - q_l(X_0, X_1, \ldots X_{\lambda_n})| < r_n.$$

These results will require a suitable higher moment condition on the $\{p_k\}$ distribution. The estimators themselves are simply the averages of what we observe in a piece of the data segment $X_{\kappa_n}, \ldots, X_{\lambda_n}$ where κ_n is chosen so that there is a large fixed number of occurrences of the relevant pattern. The reason for these stopping times λ_n is that we want to estimate only at those times when we feel that we have enough data.

For further reading on the topics see [14], [15] and [16]. For some limitations on universal stopping time estimators with universal convergence rates see Takahashi [23].

2. RESULTS

It is easiest to formally define a renewal process in terms of an underlying Markov chain. Consider a Markov chain on the state space $\{0, 1, 2, \ldots\}$ with transition probabilities $p_{i,i-1} = 1$ for all $i \geq 1$ and $p_{0,i} = p_i$ a probability distribution π on $\{0, 1, 2 \ldots\}$, cf. [5] Ex. 12.13. This chain is positive recurrent exactly when $\sum_{i=0}^{\infty} ip_{0,i} = \mu < \infty$ and the unique stationary probability assigns mass $\frac{1}{1+\mu}$ to the state 0, cf. [4] Ch. XIII and [21] 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 $X_i^j = (X_i, \ldots, X_j)$ 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:

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

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_n as the length of runs of 1's starting at position n. Formally put

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

Now for the classical binary renewal process $\{X_n\}$ define θ_n as the conditional expectation of the residual waiting time to renewal. given what we have seen at X_0^n. Formally, put

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

(Note that $\theta_n = \sum_{k=0}^{\infty} k p_{k+\tau(X_0^n)} / \sum_{l=\tau(X_0^n)}^{\infty} p_l$ as soon as there is at least one zero in X_0^n and θ_n minimizes the conditional mean square error.) Our goal is to estimate both θ_n and the distribution of the time to renewal given X_0^n but without prior knowledge of the distribution function of the process.

It will be useful to know when the renewal event occurs at the first time. Define the auxiliary stopping time ψ as the position of the first zero, that is,

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

Since we are interested in pointwise results and it was proved in [16] that no estimate can be given for all n which is pointwise consistent in a universal manner, we will give estimates only for carefully selected time instances. We will estimate only for those time instances when we feel we have enough data. Now we define these carefully selected stopping times.

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

$$\lambda_n = \min\{\ t > \lambda_{n-1} : \left(\exists \psi < i \leq t^{1/\beta} : \tau(X_0^i) = \tau(X_0^t)\right) \text{ and }$$

$$\left|\left\{t^{1/\beta} < j < \lfloor t^{1/\beta} \rfloor^{\beta} :\ \tau(X_0^j) = \tau(X_0^t)\right\}\right| \geq \lfloor t^{1/\beta} \rfloor^{\beta(1-\gamma)}\ \}.$$

Remark 2.1. Note that $i \leq t^{1/\beta}$ is the same as $i \leq \lfloor t^{1/\beta} \rfloor$ and $t^{1/\beta} < j$ is the same as $\lfloor t^{1/\beta} \rfloor < j$ in the above definition since i and j are integers.

Remark 2.2. Note that if $\lambda_n = t$ then the pattern $\tau(X_0^t)$ occurred at least once in the first part of the data segment and sufficiently many times in the second part of the data segment. In the proofs, we will use the values in the first part of the data segment as conditions and the many values in the second part to have reliable upper bounds for the conditional probability of the unfavourable events. Note also that as long as $m^\beta \leq t < (m+1)^\beta$ the requirements for the first and second part of the data segment in the definition of λ_n do not change (except the value of $\tau(X_0^t)$). In the proofs, this will ensure that the upper bounds for the probability of the unfavourable events will be summable and we will be able to apply the Borel–Cantelli lemma to show that the unfavourable events can not happen infinitely often. These facts will make it possible to give reliable estimates at the stopping times λ_n.

Remark 2.3. Note that λ_n is not smaller than n and λ_n tends to infinity.

The next theorem says that by using stopping times we skip a negligible portion of the time instances.

Theorem 2.4. Let $0 < \gamma < 1$ and $\beta > 1$. Then for the stopping times λ_n defined above,

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

almost surely.

Remark 2.5. Note that in the time segment $1, 2, \ldots, \lambda_n$ we have the stopping times $1 \le \lambda_1 < \lambda_2 < \cdots < \lambda_n$. Thus the number of time instances which are skipped is $(\lambda_n - n)$. By Theorem 2.4,

$$0 \le \frac{\lambda_n - n}{\lambda_n} = \left(1 - \frac{n}{\lambda_n}\right) \to 0$$

almost surely which means that the density of time instances which are skipped is zero.

In the data segment $\lfloor(\lambda_n)^{1/\beta}\rfloor + 1 \le j \le \lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta}\rceil - 1$ there are at least $\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil$ occurrences of the same value of τ as we see at time λ_n, but may be more. It will be useful to know the exact position where the $\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil$th occurrence is since we will not use more than that for our estimate.

Define the auxiliary random variable κ_n as

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

Notice that κ_n is smaller than $\lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta}$ since at least $\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil$ times $\tau(X_0^j) = \tau(X_0^{\lambda_n})$ for $\lfloor(\lambda_n)^{1/\beta}\rfloor + 1 \le j \le \lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta}\rceil - 1$.

Now we are ready to define our estimate $h_n(X_0^{\lambda_n})$ for the conditional expectation of the residual waiting time θ_{λ_n} at stopping time λ_n.

For $n > 0$ define our estimator $h_n(X_0^{\lambda_n})$ at time λ_n as

$$h_n(X_0^{\lambda_n}) = \frac{\sum_{i=\lfloor(\lambda_n)^{1/\beta}\rfloor+1}^{\kappa_n} I_{\{\tau(X_0^i)=\tau(X_0^{\lambda_n})\}}\sigma_i}{\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil}.$$

(Notice that κ_n ensures that we take into consideration exactly $\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil$ pieces of occurrences.) The above formula is simply the average of the residual waiting times that we have already observed in the data segment $X_{\lfloor(\lambda_n)^{1/\beta}\rfloor+1}^{\kappa_n}$ when we were at the same value of τ as we see at time λ_n. Note that as long as $m^\beta \le \lambda_n < (m+1)^\beta$ the estimator $h_n(X_0^{\lambda_n})$ 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.

Theorem 2.6. Assume $\sum_{k=0}^{\infty} k^{\alpha+1} p_k < \infty$ for some $\alpha > 1$. Let $0 < \gamma < 1$ and $\delta \ge 0$ be arbitrary. Choose

$$\beta > \max\{\frac{2+\delta\alpha}{(1-\gamma)0.5\alpha}, 1, \frac{2+\delta\alpha}{(1-\gamma)(\alpha-1)}\}.$$

Then for the stopping times λ_n and the estimator $h_n(X_0^{\lambda_n})$, defined above, for arbitrary $\epsilon > 0$,

$$\left| h_n(X_0^{\lambda_n}) - \theta_{\lambda_n} \right| \le \epsilon \lfloor (\lambda_n)^{1/\beta} \rfloor^{-\delta} \tag{2}$$

and

$$E\left(\left| \sigma_{\lambda_n} - h_n(X_0^{\lambda_n}) \right|^2 | X_0^{\lambda_n} \right) - E\left(\left| \sigma_{\lambda_n} - \theta_{\lambda_n} \right|^2 | X_0^{\lambda_n} \right) \le \epsilon^2 \lfloor (\lambda_n)^{1/\beta} \rfloor^{-2\delta} \tag{3}$$

eventually almost surely.

Remark 2.7. Note that in (3)

$$E\left(\left| \sigma_{\lambda_n} - h_n(X_0^{\lambda_n}) \right|^2 | X_0^{\lambda_n} \right) - E\left(\left| \sigma_{\lambda_n} - \theta_{\lambda_n} \right|^2 | X_0^{\lambda_n} \right) \ge 0$$

almost surely since θ_{λ_n} minimizes the conditional mean square error. Thus (3) says that the conditional mean square error for our estimate $h_n(X_0^{\lambda_n})$ is $\epsilon^2 \lfloor (\lambda_n)^{1/\beta} \rfloor^{-2\delta}$ close to the optimum eventually almost surely.

Remark 2.8. Note that with the choice of $\delta = 0$, (2) and (3) state that for arbitrary $\epsilon > 0$,

$$\left| h_n(X_0^{\lambda_n}) - \theta_{\lambda_n} \right| \le \epsilon$$

and

$$E\left(\left| \sigma_{\lambda_n} - h_n(X_0^{\lambda_n}) \right|^2 | X_0^{\lambda_n} \right) - E\left(\left| \sigma_{\lambda_n} - \theta_{\lambda_n} \right|^2 | X_0^{\lambda_n} \right) \le \epsilon^2$$

eventually almost surely. Thus we still have pointwise convergence, namely,

$$\left| h_n(X_0^{\lambda_n}) - \theta_{\lambda_n} \right| \to 0$$

and

$$E\left(\left| \sigma_{\lambda_n} - h_n(X_0^{\lambda_n}) \right|^2 | X_0^{\lambda_n} \right) - E\left(\left| \sigma_{\lambda_n} - \theta_{\lambda_n} \right|^2 | X_0^{\lambda_n} \right) \to 0$$

almost surely.

Remark 2.9. The reason for our use of the stopping times λ_n is that they enable us to guarantee that eventually we are doing “almost” as well as the best predictor with an explicit bound on this “almost”.

Remark 2.10. Note that for $1 < \alpha < 2$, the condition on β reduces to

$$\beta > \frac{2+\delta\alpha}{(1-\gamma)(\alpha-1)}$$

and for $2 \le \alpha$,

$$\beta > \max\{1, \frac{2+\delta\alpha}{(1-\gamma)0.5\alpha}\}.$$

Remark 2.11. Note that the choice of β depends on α. Thus we have to know α in advance in order to choose a suitable β which is not an advantage.

Remark 2.12. Note that λ_n depends on α through the choice of β. Thus $h_n(X_0^{\lambda_n})$ depends on α through λ_n. We have to know α in advance in order to choose a suitable β for λ_n, which is not an advantage.

Remark 2.13. For an unpractical scheme (where the estimate is not refreshed for exponentially long time) which however does not make use of prior knowledge of α see [13].

Remark 2.14. If instead of almost sure convergence one considers convergence in probability then one can estimate for all time instances in a non intermittent way, cf. [16].

In a similar fashion we can define the average of the number of times that the residual waiting time assumed a fixed value. Namely, define $\hat{p}_l(X_0^{\lambda_n})$ for each l as

$$\hat{p}_l(X_0^{\lambda_n}) = \frac{\sum_{i=\lfloor(\lambda_n)^{1/\beta}\rfloor+1}^{\kappa_n} I_{\{\tau(X_0^i)=\tau(X_0^{\lambda_n}),\sigma_i=l\}}}{\lceil \lfloor(\lambda_n)^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil}.$$

Note that $\hat{p}_l(X_0^{\lambda_n})$ is a probability distribution on the nonnegative integers.

Theorem 2.15. Assume $\sum_{k=0}^{\infty} k^{\alpha+1}p_k < \infty$ for some $\alpha > 1$. Let $0 < \gamma < 1/3$ be arbitrary. Choose

$$\beta > \max\{\frac{2}{(1-\gamma)(\alpha-1)}, \frac{4}{1-3\gamma}\}.$$

Then for the stopping times λ_n and the estimator $\hat{p}_l(X_0^{\lambda_n})$ defined above, for arbitrary $0 < \epsilon < 1$,

$$\sum_{l=0}^{\infty} \left| \hat{p}_l(X_0^{\lambda_n}) - \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} \right| \le \frac{\epsilon}{\lfloor(\lambda_n)^{(1/\beta)}\rfloor} \tag{4}$$

eventually almost surely.

Remark 2.16. Note that $\hat{p}_l(X_0^{\lambda_n})$ depends on α through λ_n via the choice of β. For an unpractical scheme (where the estimate is not refreshed for exponentially long time) which however does not make use of prior knowledge of α see [13].

Remark 2.17. Note that the conditions in Theorem 2.4 are weaker than the ones in Theorem 2.15 or in Theorem 2.6. Note that the conditions in Theorem 2.15 are stronger than the conditions in Theorem 2.6 with the choice of $\delta = 0$. This is due to the fact that the results in Theorem 2.6 (with the choice of $\delta = 0$, cf. Remark 2.8) are used in the proof of Theorem 2.15 and that for the total variation in (4) we need a different technique for upperbounding the probability of the unfavourable event because of the summation.

3. PROOF OF THEOREM 2.4.

Since if a block of 1's has positive probability it will appear with that frequency which is eventually greater than

$$\frac{\lfloor t^{1/\beta}\rfloor + \lceil \lfloor t^{1/\beta}\rfloor^{\beta(1-\gamma)}\rceil}{\lfloor t^{1/\beta}\rfloor^{\beta}}$$

(which tends to zero). The proof of Theorem 2.4 is complete. $\square$

4. PROOF OF THEOREM 2.6.

Let $1 \leq k \leq m$ 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,

$$j_{i+1}^{(k,m)} = \min\left\{t > j_i^{(k,m)} : \tau(X_{-\infty}^t) = \tau(X_{-\infty}^k)\right\}. \tag{5}$$

Now for $i \geq 1$ define

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

The outline of the proof is this. First we will consider the bad events, that is, where for some $1 \leq k \leq m$ the difference between the average $\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil} Z_i^{(k,m)}/\lceil (m^\beta)^{1-\gamma}\rceil$ and its conditional expectation is greater than $\epsilon m^{-\delta}$. (The events are indexed by m.) We will give an upper bound for the probability of these events. We will show then that these upper bounds are summable in m. Then using the Borel–Cantelli lemma we will conclude that these events can not happen infinitely often. Finally we will show that this implies (2).

Clearly $\{Z_i^{(k,m)}\}_{i=1}^{\infty}$ are conditionally independent and identically distributed given the event $\mathcal{E}_L^k$ where

$$\mathcal{E}_L^k = \{\tau(X_{-\infty}^k) = L\}. \tag{6}$$

Notice that $E(Z_i^{(k,m)}|\mathcal{E}_L^k) = \sum_{h=0}^{\infty} h p_{h+L} / \sum_{h=L}^{\infty} p_h$.

We will divide the proof into two cases. In the first case we assume that $1 < \alpha \leq 2$.

Apply Markov inequality to get that

$$
\begin{aligned}
& P\left(\left|\frac{\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^\beta)^{1-\gamma}\rceil} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right| > \epsilon m^{-\delta} | \mathcal{E}_L^k\right) \\
= \quad & P\left(\left|\frac{\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^\beta)^{1-\gamma}\rceil} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right|^\alpha > \frac{\epsilon^\alpha}{m^{\delta\alpha}} | \mathcal{E}_L^k\right) \\
\le \quad & \frac{E\left(\left|\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil}\left(Z_i^{(k,m)} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right)\right|^\alpha | \mathcal{E}_L^k\right)}{\epsilon^\alpha m^{-\delta\alpha}\lceil (m^\beta)^{(1-\gamma)}\rceil^\alpha}.
\end{aligned}
$$

Now applying Lemma 6.1 in the Appendix (Theorem 2 of von Bahr and Esseen in [2]) to upperbound the αth conditional moments of sums of conditionally independent and identically distributed random variables with zero conditional means (here we use that $1 < \alpha \le 2$) we get that

$$
\frac{E\left(\left|\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil}\left(Z_i^{(k,m)} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right)\right|^\alpha | \mathcal{E}_L^k\right)}{\epsilon^\alpha m^{-\delta\alpha}\lceil (m^\beta)^{(1-\gamma)}\rceil^\alpha}
$$

is less than or equal to

$$
\frac{2E\left(\left|Z_1^{(k,m)} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right|^\alpha | \mathcal{E}_L^k\right)}{\epsilon^\alpha m^{-\delta\alpha}\lceil (m^\beta)^{(1-\gamma)}\rceil^{(\alpha-1)}}.
$$

Since $\alpha > 1$ we can apply Jensen's inequality in order to get that

$$
\begin{aligned}
\frac{2E\left(\left|Z_1^{(k,m)} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right|^\alpha | \mathcal{E}_L^k\right)}{\epsilon^\alpha m^{-\delta\alpha}\lceil (m^\beta)^{(1-\gamma)}\rceil^{(\alpha-1)}} \quad & \le \quad \frac{2\left(E\left(\left|Z_1^{(k,m)}\right|^\alpha | \mathcal{E}_L^k\right) + \left|\frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right|^\alpha\right)}{\epsilon^\alpha m^{-\delta\alpha}(m^\beta)^{(1-\gamma)(\alpha-1)}} \\
& \le \quad \frac{2\left(E\left(\left(Z_1^{(k,m)}\right)^\alpha | \mathcal{E}_L^k\right) + \frac{\sum_{h=0}^\infty h^\alpha p_{h+L}}{\sum_{h=L}^\infty p_h}\right)}{\epsilon^\alpha m^{-\delta\alpha}(m^\beta)^{(1-\gamma)(\alpha-1)}}.
\end{aligned}
$$

Since the conditional αth moment of the random variable $Z_1^{(k,m)}$ can be calculated as $E\left(\left(Z_1^{(k,m)}\right)^\alpha | \mathcal{E}_L^k\right) = \sum_{h=0}^\infty h^\alpha p_{h+L} / \sum_{h=L}^\infty p_h$ we get that

$$
\frac{2\left(E\left(\left(Z_1^{(k,m)}\right)^\alpha | \mathcal{E}_L^k\right) + \frac{\sum_{h=0}^\infty h^\alpha p_{h+L}}{\sum_{h=L}^\infty p_h}\right)}{\epsilon^\alpha m^{-\delta\alpha}(m^\beta)^{(1-\gamma)(\alpha-1)}} \quad = \quad \frac{4}{\epsilon^\alpha m^{-\delta\alpha}(m^\beta)^{(1-\gamma)(\alpha-1)}} \frac{\sum_{h=0}^\infty h^\alpha p_{h+L}}{\sum_{h=L}^\infty p_h}.
$$

Combining all these we get that the conditional probability

$$
P\left(\left|\frac{\sum_{i=1}^{\lceil (m^\beta)^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^\beta)^{1-\gamma}\rceil} - \frac{\sum_{h=0}^\infty h p_{h+L}}{\sum_{h=L}^\infty p_h}\right| > \epsilon m^{-\delta} | \mathcal{E}_L^k\right)
$$

is less than or equal to

$$\frac{4}{\epsilon^{\alpha}m^{-\delta\alpha}(m^{\beta})^{(1-\gamma)(\alpha-1)}}\frac{\sum_{h=0}^{\infty}h^{\alpha}p_{h+L}}{\sum_{h=L}^{\infty}p_h}.$$

Multiply both sides of the inequality by the probability of the condition $P(\mathcal{E}_L^k) = (\sum_{h=L}^{\infty}p_h)/(1+\sum_{h=0}^{\infty}hp_h)$ (note that by Kac's theorem $P(X_{k-L}=0) = 1/(1+\sum_{h=0}^{\infty}hp_h)$ cf. [4] Ch. XIII and [21] 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 that the probability of the event

$$\left|\frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil}Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty}hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty}p_h}\right| > \epsilon m^{-\delta}$$

is less than or equal to

$$\frac{4}{\epsilon^{\alpha}m^{-\delta\alpha}(m^{\beta})^{(1-\gamma)(\alpha-1)}}\frac{\sum_{h=0}^{\infty}h^{\alpha+1}p_h}{1+\sum_{h=0}^{\infty}hp_h}$$

and in turn the probability of the unfavourable event

$$\max_{1\le k\le m}\left|\frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil}Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty}hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty}p_h}\right| > \epsilon m^{-\delta}$$

is less than or equal to

$$4\epsilon^{-\alpha}m^{(1+\delta\alpha-\beta(1-\gamma)(\alpha-1))}\frac{\sum_{h=0}^{\infty}h^{\alpha+1}p_h}{1+\sum_{h=0}^{\infty}hp_h}$$

which is summable since β is greater than $\frac{2+\delta\alpha}{(1-\gamma)(\alpha-1)}$.

For $\alpha > 2$ apply Markov inequality and Lemma 6.2 in the Appendix (Theorem 2.10 of Petrov [19]) to get that the conditional probability of the unfavourable event

$$P\left(\left|\frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil}Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty}hp_{h+L}}{\sum_{h=L}^{\infty}p_h}\right| > \epsilon m^{-\delta}|\mathcal{E}_L^k\right)$$

is less than or equal to

$$\frac{2C(\alpha)}{\epsilon^{\alpha}m^{-\delta\alpha}m^{\beta(1-\gamma)\alpha/2}}\frac{\sum_{h=0}^{\infty}h^{\alpha}p_{h+L}}{\sum_{h=L}^{\infty}p_h}$$

where $C(\alpha)$ depends only on α. Integrating both sides, just as in the previous case above, we get that the probability of the event

$$\left|\frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil}Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty}hp_{h+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty}p_h}\right| > \epsilon m^{-\delta}$$

is less than or equal to

$$\frac{2C(\alpha)}{\epsilon^{\alpha} m^{-\delta\alpha} m^{\beta(1-\gamma)\alpha/2}} \frac{\sum_{h=0}^{\infty} h^{\alpha+1} p_h}{1+\sum_{h=0}^{\infty} h p_h}$$

and in turn the probability that the random variable

$$\max_{1\le k\le m} \left| \frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty} h p_{h+\tau(X_{-\infty}^{k})}}{\sum_{h=\tau(X_{-\infty}^{k})}^{\infty} p_h} \right|$$

is greater than $\epsilon m^{-\delta}$ is less than or equal to

$$2C(\alpha)\epsilon^{-\alpha} m^{(1+\delta\alpha-\beta(1-\gamma)0.5\alpha)} \frac{\sum_{h=0}^{\infty} h^{\alpha+1} p_h}{1+\sum_{h=0}^{\infty} h p_h}$$

which is summable since β is greater than $\frac{2+\delta\alpha}{(1-\gamma)0.5\alpha}$. Applying the Borel–Cantelli lemma in both cases one gets that

$$\max_{1\le k\le m} \left| \frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil} - \frac{\sum_{h=0}^{\infty} h p_{h+\tau(X_{-\infty}^{k})}}{\sum_{h=\tau(X_{-\infty}^{k})}^{\infty} p_h} \right| \le \frac{\epsilon}{m^{\delta}}$$

eventually almost surely.

Observe that for $k \ge \psi$, $\tau(X_{-\infty}^{k}) = \tau(X_0^k)$. Now for suitable $\psi < k \le \lfloor (\lambda_n)^{1/\beta} \rfloor$ and $m = \lfloor (\lambda_n)^{1/\beta} \rfloor$:

$$h_n(X_0^{\lambda_n}) = \frac{\sum_{i=1}^{\lceil (m^{\beta})^{1-\gamma}\rceil} Z_i^{(k,m)}}{\lceil (m^{\beta})^{1-\gamma}\rceil}$$

and

$$\theta_{\lambda_n} = \frac{\sum_{h=0}^{\infty} h p_{h+\tau(X_{-\infty}^{k})}}{\sum_{h=\tau(X_{-\infty}^{k})}^{\infty} p_h}.$$

Since $\lambda_n \uparrow \infty$ we get that

$$\left| h_n(X_0^{\lambda_n}) - \theta_{\lambda_n} \right| \le \epsilon \lfloor (\lambda_n)^{1/\beta} \rfloor^{-\delta}$$

eventually almost surely which gives (2). The first part of Theorem 2.6 is complete. Finally we prove (3). Since

$$E\left(\left| \sigma_{\lambda_n} - h_n(X_0^{\lambda_n}) \right|^2 | X_0^{\lambda_n} \right) - E\left(|\sigma_{\lambda_n} - \theta_{\lambda_n}|^2 | X_0^{\lambda_n} \right) = \left(h_n(X_0^{\lambda_n}) - \theta_{\lambda_n} \right)^2$$

apply (2) to get (3). The proof of Theorem 2.6 is complete. □

5. PROOF OF THEOREM 2.15.

Let $0 < \eta < 1$ be arbitrary. The outline of the proof is this. We will decompose the error $\sum_{l=0}^{\infty} \left|\hat{p}_l(X_0^{\lambda_n}) - p_{l+\tau(X_0^{\lambda_n})} / \sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i\right|$ into the sum of two terms $A_n + B_n$ where $A_n = \sum_{l=0}^{\lceil \eta^{-1} \lfloor (\lambda_n)^{(1/\beta)} \rfloor^{\beta\gamma+1} \rceil - 1} \left|\hat{p}_l(X_0^{\lambda_n}) - p_{l+\tau(X_0^{\lambda_n})} / \sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i\right|$ and B_n is the rest of the sum. First, using Hoeffding's inequality we will prove that A_n is at most $2\eta / \lfloor (\lambda_n)^{(1/\beta)} \rfloor$ eventually almost surely. Here the reasoning will be similar to the argument in the proof of Theorem 2.6. For B_n we will use the trivial upper bound, that is, B_n is less than or equal to the sum $\sum_{l=\lceil \eta^{-1} \lfloor (\lambda_n)^{(1/\beta)} \rfloor^{\gamma\beta+1} \rceil}^{\infty} \left(\hat{p}_l(X_0^{\lambda_n}) + p_{l+\tau(X_0^{\lambda_n})} / \sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i\right)$ and by using the results in Theorem 2.6 we will show that the sum is at most $4\eta / \lfloor (\lambda_n)^{(1/\beta)} \rfloor$ eventually almost surely. By choosing $\eta = \epsilon/6$ we will conclude that the sum $A_n + B_n$ is at most $\epsilon / \lfloor (\lambda_n)^{(1/\beta)} \rfloor$ eventually almost surely.

In order to deal with A_n we define the indicator random variables

$$Z_{i,l}^{(k,m)} = I_{\left\{\sigma_{j_i^{(k,m)}} = l\right\}}$$

where $j_i^{(k,m)}$ is the same as in (5). Clearly, for fixed $k < m$ and l, $\{Z_{i,l}^{(k,m)}\}_{i=1}^{\infty}$ are conditionally independent and identically distributed given $\mathcal{E}_L^k$, see (6). Apply Hoeffding's inequality to get that

$$P\left(\left|\frac{\sum_{i=1}^{\lceil m^{\beta(1-\gamma)} \rceil} Z_{i,l}^{(k,m)}}{\lceil m^{\beta(1-\gamma)} \rceil} - \frac{p_{l+L}}{\sum_{h=L}^{\infty} p_h}\right| > \frac{\eta^2}{m^{\beta\gamma+2}} \middle| \mathcal{E}_L^k\right) \le 2e^{-\frac{\eta^4 m^{\beta(1-\gamma)}}{2m^{2\beta\gamma+4}}}.$$

After integrating both sides with respect to the conditioning we get that

$$P\left(\left|\frac{\sum_{i=1}^{\lceil m^{\beta(1-\gamma)} \rceil} Z_{i,l}^{(k,m)}}{\lceil m^{\beta(1-\gamma)} \rceil} - \frac{p_{l+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty} p_h}\right| > \eta^2 m^{-\beta\gamma-2}\right) \le 2e^{-\frac{\eta^4 m^{\beta(1-\gamma)}}{2m^{2\beta\gamma+4}}}.$$

Now the probability that the random variable

$$\sum_{l=0}^{\lceil \eta^{-1} m^{\beta\gamma+1} \rceil - 1} \left|\frac{\sum_{i=1}^{\lceil m^{\beta(1-\gamma)} \rceil} Z_{i,l}^{(k,m)}}{\lceil m^{\beta(1-\gamma)} \rceil} - \frac{p_{l+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty} p_h}\right|$$

is greater than $\frac{\lceil \eta^{-1} m^{\beta\gamma+1} \rceil \eta^2}{m^{\beta\gamma+2}}$ can be at most $2\lceil \eta^{-1} m^{\beta\gamma+1} \rceil e^{-\frac{\eta^4 m^{\beta(1-\gamma)}}{2m^{2\beta\gamma+4}}}$ and in turn the probability that the random variable

$$\max_{1 \le k \le m} \sum_{l=0}^{\lceil \eta^{-1} m^{\beta\gamma+1} \rceil - 1} \left|\frac{\sum_{i=1}^{\lceil m^{\beta(1-\gamma)} \rceil} Z_{i,l}^{(k,m)}}{\lceil m^{\beta(1-\gamma)} \rceil} - \frac{p_{l+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty} p_h}\right|$$

is greater than $2\eta/m$ can not be greater than $2m\lceil\eta^{-1}m^{\beta\gamma+1}\rceil e^{-0.5\eta^4 m^{\beta(1-3\gamma)-4}}$ which is summable and so by the Borel–Cantelli lemma,

$$\max_{1\le k\le m}\sum_{l=0}^{\lceil\eta^{-1}m^{\beta\gamma+1}\rceil-1}\left|\frac{\sum_{i=1}^{\lceil m^{\beta(1-\gamma)}\rceil}Z_{i,l}^{(k,m)}}{\lceil m^{\beta(1-\gamma)}\rceil}-\frac{p_{l+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty}p_h}\right|$$

is not greater than $2\eta/m$ eventually almost surely.

Observe first that after the first appearance of the zero, that is for $k\ge\psi$, $\tau(X_{-\infty}^k)=\tau(X_0^k)$. Now for suitable $\psi<k\le\lfloor(\lambda_n)^{(1/\beta)}\rfloor$ and $m=\lfloor(\lambda_n)^{(1/\beta)}\rfloor$:

$$\hat{p}_l(X_0^{\lambda_n})=\frac{\sum_{i=1}^{\lceil(m^\beta)^{1-\gamma}\rceil}Z_{i,l}^{(k,m)}}{\lceil(m^\beta)^{1-\gamma}\rceil}$$

and

$$\frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty}p_i}=\frac{p_{l+\tau(X_{-\infty}^k)}}{\sum_{h=\tau(X_{-\infty}^k)}^{\infty}p_h}.$$

Since $\lambda_n\uparrow\infty$ we get that

$$A_n=\sum_{l=0}^{\lceil\eta^{-1}\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{\beta\gamma+1}\rceil-1}\left|\hat{p}_l(X_0^{\lambda_n})-\frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty}p_i}\right|\le\frac{2\eta}{\lfloor(\lambda_n)^{(1/\beta)}\rfloor}\tag{7}$$

eventually almost surely.
Now we deal with B_n. Note that by the Markov inequality,

$$\frac{\sum_{l=\lceil\eta^{-1}(\mu_L+0.5)\lfloor k^{(1/\beta)}\rfloor\rceil}^{\infty}p_{l+L}}{\sum_{l=L}^{\infty}p_l}\le\frac{\eta\mu_L}{(\mu_L+0.5)\lfloor k^{(1/\beta)}\rfloor}=\frac{\eta}{\lfloor k^{(1/\beta)}\rfloor}\tag{8}$$

where $\mu_L=\sum_{i=L}^{\infty}(i-L)p_i/\sum_{i=L}^{\infty}p_i$.
Now observe that almost surely for sufficiently large n:

$$\mu_{\tau(X_0^{\lambda_n})}+0.5\le\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{\beta\gamma}.\tag{9}$$

Indeed, since in the data segment $X_{\lfloor(\lambda_n)^{(1/\beta)}\rfloor+1},\dots,X_{\lfloor\lfloor(\lambda_n)^{(1/\beta)}\rfloor^\beta\rfloor}$ there are at least $\lceil\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{(1-\gamma)\beta}\rceil$ zeros, we can give an upperbound on the estimate for the conditional expectation of the residual waiting time as

$$\begin{aligned}h_n(X_0^{\lambda_n})&=\frac{\sum_{i=\lfloor(\lambda_n)^{(1/\beta)}\rfloor+1}^{\kappa_n}I_{\{\tau(X_0^i)=\tau(X_0^{\lambda_n})\}}\sigma_i}{\lceil\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{\beta(1-\gamma)}\rceil}\\&\le\frac{\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{\beta}-\lceil\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{(1-\gamma)\beta}\rceil}{\lceil\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{(1-\gamma)\beta}\rceil}\le\lfloor(\lambda_n)^{(1/\beta)}\rfloor^{(\gamma\beta)}-1,\end{aligned}$$

and by Theorem 2.6 (with $\delta=0$, cf. Remark 2.8), $|h_n(X_0^{\lambda_n})-\mu_{\tau(X_0^{\lambda_n})}|\to 0$ almost surely.

By (9) and (8),

$$\sum_{l=\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil}^{\infty} \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} \leq \sum_{l=\lceil \eta^{-1}(\mu_{\tau(X_0^{\lambda_n})}+0.5)\lfloor (\lambda_n)^{(1/\beta)}\rfloor\rceil}^{\infty} \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} \leq \frac{\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \tag{10}$$

eventually almost surely.
Now apply (10), in order to give an upperbound on B_n

$$\begin{aligned} B_n &= \sum_{l=\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil}^{\infty} \left| \hat{p}_l(X_0^{\lambda_n}) - \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} \right| \\ &\leq \sum_{l=\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil}^{\infty} \hat{p}_l(X_0^{\lambda_n}) + \sum_{l=\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil}^{\infty} \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} \\ &\leq 1 - \sum_{l=0}^{\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil-1} \hat{p}_l(X_0^{\lambda_n}) + \frac{\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \end{aligned}$$

eventually almost surely.
Now apply (7) and (10) in order to further upperbound B_n

$$\begin{aligned} B_n &\leq 1 - \sum_{l=0}^{\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil-1} \hat{p}_l(X_0^{\lambda_n}) + \frac{\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \\ &\leq 1 - \sum_{l=0}^{\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil-1} \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} + A_n + \frac{\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \\ &\leq \sum_{l=\lceil \eta^{-1}\lfloor (\lambda_n)^{(1/\beta)}\rfloor^{\gamma\beta+1}\rceil}^{\infty} \frac{p_{l+\tau(X_0^{\lambda_n})}}{\sum_{i=\tau(X_0^{\lambda_n})}^{\infty} p_i} + \frac{3\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \\ &\leq \frac{4\eta}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor} \end{aligned}$$

eventually almost surely. Choose $\eta = \epsilon/6$ in order to get the upperbound on the error

$$A_n + B_n \leq \frac{\epsilon}{\lfloor (\lambda_n)^{(1/\beta)}\rfloor}$$

eventually almost surely. The proof of Theorem 2.15 is complete. $\square$

6. APPENDIX

Lemma 6.1. (Theorem 2 in von Bahr and Esseen [2]) Let $X_1, X_2, \ldots, X_n$ be random varables satisfying

$$E(X_{m+1}|\sum_{i=1}^{m} X_i) = 0 \quad \text{for all } 1 \le m \le n-1.$$

Let $1 \le r \le 2$. If

$$E(|X_k|^r < \infty \quad \text{for all } 1 \le k \le n$$

then

$$E(|\sum_{i=1}^{n} X_i|^r) \le 2 \sum_{i=1}^{n} E(|X_i|^r).$$

The next lemma is Theorem 2.10 in Petrov [19].

Lemma 6.2. (Theorem 2.10 in Petrov [19]) Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables with zero means and let $p \ge 2$. Then

$$E|\sum_{i=1}^{n} Z_i|^p \le C(p) n^{p/2-1} \sum_{i=1}^{n} E|Z_i|^p.$$

where $C(p)$ is a positive constant depending only on p.

As Petrov [19] pointed out, this is an immediate consequence of the Marcinkiewicz-Zygmund inequality, cf. Marcinkiewicz and Zygmund [8] (p. 498 in Shiryayev [22]).

(Received July 26, 2019)

REFERENCES

[1] P. Algoet: The strong low of large numbers for sequential decisions under uncertainity. IEEE Trans. Inform. Theory *40* (1994), 609–634. DOI:10.1109/18.335876

[2] B. von Bahr and C. G. Esseen: Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$. Ann. Math. Statist. *36* (1965), 299–303. DOI:10.1214/aoms/1177700291

[3] L. Denby and Y. Vardi: A short-cut method for estimation in renewal processes. Technometrics *27* (1985), 4, 361–373. DOI:10.1080/00401706.1985.10488075

[4] W. Feller: An Introduction to Probability Theory and its Applications Vol. I. Third edition. John Wiley and Sons, Inc., 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] L. Györfi and G. Ottucsák: Sequential prediction of unbounded stationary time series. IEEE Trans. Inform. Theory *53* (2007), 1866–1872. DOI:10.1109/tit.2007.894660

[7] S. Khudanpur and P. Narayan: Order estimation for a special class of hidden Markov sources and binary renewal processses. IEEE Trans. Inform. Theory *48* (2002), 1704–1713. DOI:10.1109/tit.2002.1003850

[8] J. Marcinkiewicz and A. Zygmund: Sur les foncions independantes. Fund. Math. *28* (1937), 60–90.

[9] G. Morvai: Guessing the output of a stationary binary time series. In: Foundations of Statistical Inference (Y. Haitovsky, H. R. Lerche and Y. Ritov, eds.), Physika-Verlag 2003, pp. 207–215.

[10] G. Morvai and B. Weiss: Forecasting for stationary binary time series. Acta Applic. Math. *79* (2003), 25–34. DOI:10.1023/a:1025862222287

[11] G. Morvai and B. Weiss: Prediction for discrete time series. Probab. Theory Related Fields *132* (2005), 1–12. DOI:10.1007/s00440-004-0386-3

[12] G. Morvai and B. Weiss: On sequential estimation and prediction for discrete time series. Stoch. Dynam. *7* (2007), 4, 417–437. DOI:10.1142/s021949370700213x

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

[14] G. Morvai and B. Weiss: Estimating the residual waiting time for binary stationary time series. In: ITW 2009, IEEE Information Theory Workshop on Networking and Information Theory, 2009 pp. 67–70. DOI:10.1109/itwnit.2009.5158543

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

[16] G. Morvai and B. Weiss: A versatile scheme for predicting renewal times. Kybernetika *52* (2016), 3, 348–358. DOI:10.14736/kyb-2016-3-0348

[17] A. B. Nobel: On optimal sequential prediction for general processes. IEEE Trans. Inform. Theory *49* (2003), 83–98. DOI:10.1109/tit.2002.806141

[18] E. A. Peña, R. Strawderman, and M. Hollander: Nonparametric estimation with recurrent event data. J. Amer. Statist. Assoc. *96* (2001), 456, 1299–1315. DOI:10.1198/016214501753381922

[19] V. V. Petrov: Limit Theorems of Probability Theory. Clarendon Press, Oxford 1995. DOI:10.1017/s001309150002335x

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

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

[22] A. N. Shiryayev: Probability. Second edition. Springer-Verlag, New York 1996.

[23] H. Takahashi: Computational limits to nonparametric estimation for ergodic processes. IEEE Trans. Inform. Theory *57* (2011), 10, 6995–6999. DOI:10.1109/tit.2011.2165791

[24] Y. Vardi: Nonparametric estimation in renewal processes. Ann. Statist. (1982), 772–785. DOI:10.1214/aos/1176345870

Gusztáv Morvai, Budapest, Hungary.
e-mail: morvaigusztav@gmail.com

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