KYBERNETIKA — VOLUME *56* (2020), NUMBER 3, PAGES 410–431

ESTIMATING THE CONDITIONAL EXPECTATIONS FOR CONTINUOUS TIME STATIONARY PROCESSES

Gusztáv Morvai and Benjamin Weiss

One of the basic estimation problems for continuous time stationary processes X_t, is that of estimating $E\{X_{t+\beta}|X_s : s \in [0,t]\}$ based on the observation of the single block $\{X_s : s \in [0,t]\}$ when the actual distribution of the process is not known. We will give fairly optimal universal estimates of this type that correspond to the optimal results in the case of discrete time processes.

Keywords: nonparametric estimation, continuous time stationary processes

Classification: 60G10,60G25, 62G05

1. INTRODUCTION

Tom Cover formulated a number of problems in the Proceedings of the First International IEEE-USSR Information Workshop [6] that have generated a substantial literature. He posed two questions concerning estimation of discrete time stationary and ergodic binary processes without any further prior knowledge of the distribution. In his first question he asked if there exists a universal estimator $\hat{E}_n$ solely depending on the observations $(X_0, X_1, X_2, \ldots, X_n)$ such that for all discrete time stationary and ergodic binary processes

$$\lim_{n\to\infty} |\hat{E}_n(X_0, X_1, X_2, \ldots, X_n) - E(X_{n+1}|X_0, X_1, X_2, \ldots, X_n)| = 0 \text{ almost surely.} \quad (1)$$

This problem is called the 'forward' problem and the estimator a 'forward' estimator because the estimator $\hat{E}_n$ may make use of the data segment of ever increasing length $(X_0, X_1, X_2, \ldots, X_n)$ and tries to estimate an ever moving target $E(X_{n+1}|X_0, X_1, X_2, \ldots, X_n)$ where n tends to $+\infty$, in the positive 'forward' direction. (As for an application, one may consider a river and let zero denote the event that there will not be flood and let one denote the event that there will be flood. In this case $E(X_{n+1}|X_0, X_1, X_2, \ldots, X_n)$ is the probability that there will be flood in year $(n + 1)$ given the past observations of the behaviour of the river from year zero to year n.)

In his second question, Cover asked if there exists a universal estimator $\hat{E}_{-n}$ solely depending on the observations $(X_{-n} \ldots, X_{-2}, X_{-1}, X_0)$ such that for all discrete time

DOI: 10.14736/kyb-2020-3-0410

stationary and ergodic binary processes

$$\lim_{n\to\infty} \hat{E}_{-n}(X_{-n}\ldots,X_{-2},X_{-1},X_0) = E(X_1|\ldots,X_{-2},X_{-1},X_0) \text{ almost surely.} \quad (2)$$

This problem is called the 'backward' problem and the estimator a 'backward' estimator because the estimator $\hat{E}_{-n}$ may make use of the data segment of ever increasing length $(X_{-n}\ldots,X_{-2},X_{-1},X_0)$ and tries to estimate a fixed target $E(X_1|\ldots,X_{-2},X_{-1},X_0)$ where we collect more and more data from the past, in the negative 'backward' direction. (As for an application, one may consider the special case where the infinite past $(\ldots,X_{-2},X_{-1},X_0)$ determines the exact value of X_1. In this case $E(X_1|\ldots,X_{-2},X_{-1},X_0)$ is either zero or one and the goal is to reconstruct the exact value of X_1 from the past observations. This problem is called the reconstruction problem.)

Notice that while $E(X_{n+1}|X_0,X_1,X_2,\ldots,X_n)$ does not converge almost surely in general, $E(X_1|X_{-n}\ldots,X_{-2},X_{-1},X_0)$ does. Namely,

$$\lim_{n\to\infty} E(X_1|X_{-n}\ldots,X_{-2},X_{-1},X_0) = E(X_1|\ldots,X_{-2},X_{-1},X_0) \text{ almost surely.} \quad (3)$$

It turned out that the answers to the 'forward' and the 'backward' problems are far from being the same. Ornstein [20] gave a rather complicated algorithm for the backward estimation problem (2) whereas Bailey [4] provided a proof for the nonexistence of a universal algorithm guaranteeing almost sure convergence in the forward estimation problem (1) . To do this, Bailey in [4], assuming the existence of a universal algorithm, used Ornstein's technique of cutting and stacking [20] for the construction of a "counterexample" process for which the algorithm fails to converge (see Shields [25] for more details on this method).

The problem came to life again in the late eighties with the work of Ryabko [21]. He used a simpler technique, namely - relabelling a countable state Markov chain, in order to prove the nonexistence of a universal estimator for the forward estimation problem (1) (cf. also Györfi, Morvai and Yakowitz [11]).

One approach in an attempt to obtain positive results for the problem of forward estimation in the face of Bailey's theorem modifies the almost sure convergence to almost sure convergence of Cesaro averages. The forward problem for Cesaro averages is this. Does there exist a universal estimator $\hat{E}_n$ solely depending on the observations $(X_0,X_1,X_2,\ldots,X_n)$ such that for all discrete time stationary and ergodic binary processes

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} |\hat{E}_n(X_0,X_1,X_2,\ldots,X_n) - E(X_{n+1}|X_0,X_1,X_2,\ldots,X_n)| = 0 \quad (4)$$

almost surely? (Notice that now one is allowed to make a certain error infinitely many times but not too often so that the errors vanish in the time (Cesaro) average.) This was solved already by Bailey in his thesis [4] who constructed such universal estimator. (Cf. Algoet [2, 3] and Weiss [27] also.) (As for an application, one may consider a certain stock at the stock market and let zero denote the event that the price of the stock goes down and let one denote the event that the price goes up. In this case $E(X_{n+1}|X_0,X_1,X_2,\ldots,X_n)$ is the probability that the price of the stock will go up on day $(n+1)$ given the past observations of the behaviour of the stock. The goal is to estimate this probability well in Cesaro average so that most of the time the predictionn will be correct. One may use this prediction to sell when with higher probability the price will go down and buy when the price will go up according to our estimator.)

In case of the backward estimation problem (2), several authors first have extended the results from discrete time binary processes to discrete time bounded real valued processes using quantization to reduce to the finite valued case see for example Algoet [1], Morvai [16], Morvai Yakowitz and Györfi [17] and later to discrete time real-valued unbounded processes, cf. Györfi et.al. [10] and Algoet [3].

In case of the forward estimation in Cesaro averages problem (4), several authors extended the results from discrete time binary processes to discrete time real-valued bounded processes, for example Algoet [1, 3], Morvai [16], Morvai Yakowitz and Györfi [17]. Even though, some authors using the method of weighted averages of so called 'experts' obtained results for discrete time real-valued unbounded processes, for example Györfi and Ottucsák [12] (cf. Györfi et. al. [13] also) the moment conditions were not optimal. We have given some fairly definitive results for forward estimation in Cesaro averages (4) in [18].

Since if $E\left(|X_0|\log^+(|X_0|)\right)<\infty$ then martingale convergence in (3), Doob's inequality and Breiman's generalized ergodic theorem (cf. [2]) yield

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|E(X_{n+1}|X_0,X_1,\dots,X_n-E(X_{n+1}|\dots,X_{-1},X_0,X_1,\dots,X_n)|=0$$

almost surely and so the Cesaro average problem for discrete time real valued stationary and ergodic processes in (4) is equivalent to the following formulation of the problem. Does there exist a universal estimator $\hat{E}_n$ solely depending on the observations $(X_0,X_1,X_2,\dots,X_n)$ such that for all discrete time stationary and ergodic real valued processes with $E\left(|X_0|\log^+(|X_0|)\right)<\infty$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|\hat{E}_n(X_0,X_1,\dots,X_n)-E(X_{n+1}|\dots,X_{-1},X_0,X_1,\dots,X_n)|=0 \quad (5)$$

almost surely? Note that the estimator $\hat{E}_n$ depends only on $(X_0,X_1,\dots,X_n)$ but the quantity we are trying to estimate $E(X_{n+1}|\dots,X_{-1},X_0,X_1,\dots,X_n)$ depends partly on values $(\dots,X_{-2},X_{-1})$ which the estimator will never observe.

In this paper we take up the corresponding questions of (2) and (5) for continuous time processes. This is of interest because there are many natural phenomena modelled by continuous time processes such as Brownian motion, Poisson point processes and more general continuous time renewal processes, Markov processes in continuous time etc. The backward problem we will consider for a stationary processes $\{X_t\}$ is that of estimating X_β given the past $\{X_t : t\in(-\infty,0]\}$ based on observing finite sections of the past when the distribution of the process is unknown. We shall do this for any fixed value of $\beta>0$. We will also deal with the problem of forward estimation, that is estimating the conditional expectation of $X_{T+\beta}$ given $\{X_t : t\in(-\infty,T]\}$. Here as in the case of discrete time case we will need to evaluate our guesses using Cesaro averaging.

As for an application, one may consider a device measuring the water level of a river, in continuous time. Our task is then that based on the observations, to give estimation for the water level, let's say, a week ahead.

The only prior works for universal estimation in this setting are due to Scarpellini [22, 23, 24] who based his result for the continuous time backward problem on the original universal scheme for discrete time backward estimator due to Ornstein [20]. Scarpellini [24] considered continuous time real-valued bounded stationary and ergodic processes and obtained results for the backward estimation problem under severe restrictions. Using the more recent schemes pioneered by Morvai [16] (cf. also Algoet [3] and Morvai et al. [17]) we will generalize his results in several ways, in particular for unbounded processes (with some integrability restrictions) and get optimal results for the forward estimation in Cesaro averages problem as well.

In the next section we will formulate more precisely our main results.The following section contains the proofs of these results, while several auxiliary facts which are needed for the proofs are relegated to an appendix.

Finally, we would like to thank the referees for several useful remarks which improved our exposition.

2. RESULTS

Before giving the main results we shall describe the processes we shall deal with. The simplest framework for discussing stationary processes with a continuous time parameter is to assume that we have a probability space (Ω, Φ, P) and a one parameter family of measure preserving invertible transformation $T_t : \Omega \to \Omega$ ($-\infty < t < \infty$) which are jointly measurable as a map from $\Omega \times \mathbb{R} \to \Omega$ and has the group property that $T_t T_s = T_{t+s}$ (see Ch. XI in Doob [8] and Ch. III in Neveu [19]). In this situation any real valued measurable function $f : \Omega \to \mathbb{R}$ defines a stochastic process $X_t(\omega) = f(T_t \omega)$. These processes are separable which means that for any dense subset $S \subset \mathbb{R}$ and any interval I, the σ-field generated by $\{X_t : t \in I\}$ equals the σ-field generated by $\{X_t : t \in S \bigcap I\}$ (cf. e.g. Proposition III.4.3 on p. 89 in Neveu [19]). Note that we will not assume that the flow T_t is ergodic.

We follow this formal framework for a stochastic process that we have just explained. To define the basic backward scheme, we shall use a sequence of finer and finer discretizations of the time parameters and quantizations of the real random variables $\{X_s : s \in (-\infty, 0]\}$ that are being observed. (We need discretization and quantization because our scheme will depend on pattern matching and we have to ensure to find a recurrence of the patern.)

We adjust the discrete time scheme in Morvai [16], Morvai, Yakowitz and Györfi [17] , Algoet [3] and Morvai and Weiss [18] to continuous time in the following way:

For $n = 1, 2, \dots$ let $\mathcal{P}_n = \{A_{n,i} : i = 1, 2, \dots\}$ be a nested sequence of countable partitions of the real line by intervals. Let $A_n(x)$ denote the cell of the nth partition $\mathcal{P}_n$ which contains the point x. Assume that

$$\sup_{n=1,2,\dots} \sup_{x \in \mathbb{R}} \sup_{y \in A_n(x)} |y - x| < \infty \quad \text{and} \quad \lim_{n \to \infty} \sup_{x \in \mathbb{R}} \sup_{y \in A_n(x)} |y - x| = 0. \tag{6}$$

Let $[\cdot]^n$ denote the quantizer which is measurable with respect to $\sigma(\mathcal{P}_n)$ and $[x]^n \in$

$A_n(x)$. It is immediate from (6) that

$$\sup_{n=1,2,\dots}\sup_{x\in\mathbb{R}}|[x]^n-x|<\infty. \quad \text{and} \quad \lim_{n\to\infty}\sup_{x\in\mathbb{R}}|[x]^n-x|=0. \tag{7}$$

For example, one may choose $\mathcal{P}_n=\{[\frac{k}{2^n},\frac{k+1}{2^n}) : k=0,\mp1,\mp2,\dots\}$ and $[x]^n=\frac{k}{2^n}$ if $\frac{k}{2^n}\le x<\frac{k+1}{2^n}$.

Let $\beta>0$ be arbitrary, but fixed. Let

$$\hat{T}=T_\beta. \tag{8}$$

Note that $\hat{T}$ is a measure preserving transformation.
Define the sequences λ_{m-1}, τ_m and R_{m-1} recursively ($m=1,2,\dots$). Put $\lambda_0=0$, $R_0=0$ and let τ_1 be the time between the occurrence of the pattern

$$[X_0]^1$$

at time 0 and the last occurrence of the same pattern at times $\dots,-2\beta,-\beta$. Formally let

$$\tau_1=\min\{t\in\{\beta,2\beta,3\beta,\dots\}:[X_{-t}]^1=[X_0]^1\}.$$

Note that since $[X_0]^1$ takes values from a countable set and since $\hat{T}$ in (8) is measure preserving transformation we have $0<\beta\le\tau_1<\infty$ almost surely. Put

$$\lambda_1=\tau_1+\lambda_0=\tau_1+0=\tau_1.$$

Note that $0=\lambda_0<\lambda_0+\beta=\beta\le\lambda_1<\infty$ almost surely. Define the first estimate R_1 as

$$R_1=X_{-\tau_1+\beta}.$$

Note that $-\tau_1+\beta\le 0$ and R_1 depends only on $\{X_s : s\in[-\lambda_1,0]\}$. Now let τ_2 be the time between the occurrence of the pattern

$$([X_{-\lambda_1}]^2,[X_{-\lambda_1+\beta/2^2}]^2,\dots,[X_{-\beta/2^2}]^2,[X_0]^2)$$

at time 0 and the last occurrence of the same pattern at times $\dots,-2\beta,-\beta$. Formally let

$$\tau_2=\min\{t\in\{\beta,2\beta,3\beta,\dots\}:[X_{-j\beta/2^2-t}]^2=[X_{-j\beta/2^2}]^2 \text{ for } j=0,1,\dots,\tfrac{\lambda_1 2^2}{\beta}\}.$$

Note that since $([X_{-\lambda_1}]^2,[X_{-\lambda_1+\beta/2^2}]^2,\dots,[X_{-\beta/2^2}]^2,[X_0]^2)$ takes values from a countable set and since $\hat{T}$ in (8) is measure preserving transformation we have $0<\beta\le\tau_2<\infty$ almost surely. Put

$$\lambda_2=\tau_2+\lambda_1.$$

Note that $\lambda_1<\lambda_1+\beta\le\lambda_2<\infty$. Define the second estimate R_2 as

$$R_2=\frac{X_{-\tau_1+\beta}+X_{-\tau_2+\beta}}{2}.$$

Note that R_2 depends only on $\{X_s : s \in [-\lambda_2, 0]\}$. Now in general let τ_m be the time between the occurrence of the pattern

$$([X_{-\lambda_{m-1}}]^m, [X_{-\lambda_{m-1}+\beta/2^m}]^m, \dots, [X_{-\beta/2^m}]^m, [X_0]^m)$$

at time 0 and the last occurrence of the same pattern at times $\dots, -2\beta, -\beta$. Formally let

$$\tau_m = \min\{t \in \{\beta, 2\beta, 3\beta, \dots\} : [X_{-j\beta/2^m - t}]^m = [X_{-j\beta/2^m}]^m \text{ for } j = 0, 1, \dots, \tfrac{\lambda_{m-1}2^m}{\beta}\}.$$

Note that since $([X_{-\lambda_{m-1}}]^m, [X_{-\lambda_{m-1}+\beta/2^m}]^m, \dots, [X_{-\beta/2^m}]^m, [X_0]^m)$ takes values from a countable set and since $\hat{T}$ in (8) is measure preserving transformation we have $0 < \beta \le \tau_m < \infty$ almost surely. Put

$$\lambda_m = \tau_m + \lambda_{m-1}.$$

Note that $\lambda_m \uparrow \infty$ since $\tau_m \ge \beta > 0$. Define the mth estimate R_m as

$$R_m = \frac{1}{m} \sum_{j=1}^{m} X_{-\tau_j + \beta}. \tag{9}$$

Note that R_m depends only on $\{X_s : s \in [-\lambda_m, 0]\}$. To obtain a fixed sample size $t \ge 0$ version, let κ_t be the maximum of integers $k = 0, 1, 2, \dots$ for which $\lambda_k \le t$. Formally, for $t \ge 0$

$$\kappa_t = \max\{k : \lambda_k \le t \quad \text{k=0,1,2,}\dots\}.$$

Since $\lambda_0 = 0$ and $\lambda_k \uparrow \infty$ the above formula is well defined. Note that

$$\kappa_t = k \qquad \text{as long as } \lambda_k \le t < \lambda_{k+1} \tag{10}$$

and

$$\kappa_t = \sum_{k=0}^{\infty} k I_{\{\lambda_k \le t < \lambda_{k+1}\}}. \tag{11}$$

For $t \ge 0$ put

$$\hat{R}_{-t} = R_{\kappa_t}. \tag{12}$$

Note that $\hat{R}_0 = R_0 = 0$ and $\hat{R}_{-t}$ depends only on $\{X_s : s \in [-t, 0]\}$. Note also that

$$\hat{R}_{-t} = R_m \qquad \text{as long as } \lambda_m \le t < \lambda_{m+1} \tag{13}$$

and

$$\hat{R}_{-t} = \sum_{k=0}^{\infty} R_k I_{\{\lambda_k \le t < \lambda_{k+1}\}}. \tag{14}$$

Note that since λ_k takes values from $\{0, \beta, 2\beta, 3\beta, \dots\}$, for any $l = 0, 1, 2, \dots$,

$$\hat{R}_{-t} = \hat{R}_{-l\beta} \qquad \text{as long as } l\beta \le t < (l+1)\beta \tag{15}$$

and

$$\hat{R}_{-t} = \sum_{l=0}^{\infty} \hat{R}_{-l\beta} I_{\{l\beta \le t < (l+1)\beta\}} = \sum_{l=0}^{\infty} I_{\{l\beta \le t < (l+1)\beta\}} \sum_{k=0}^{\infty} R_k I_{\{\lambda_k \le l\beta < \lambda_{k+1}\}}. \tag{16}$$

Note that $\hat{R}_{-t}$ is not a continuous function of $t \in [0,\infty)$ (except if it is the constant zero) but it is right semi continuous in $t \in [0,\infty)$ (or in other words, $\hat{R}_t$ is left semi continuous in $t \in (-\infty, 0]$). Now $\hat{R}_{-t}(\omega)$ is jointly measurable in ω and t. Indeed, for a Borel measurable set $A \subseteq \mathbb{R}$,

$$\begin{aligned}
&\{(\omega,t) \ : \ \hat{R}_{-t}(\omega) \in A\} \\
&= \{(\omega,t) \ : \ \sum_{k=0}^{\infty} R_k(\omega) I_{\{\lambda_k(\omega) \le t < \lambda_{k+1}(\omega)\}} \in A\} \\
&= \bigcup_{k=0}^{\infty} \bigcup_{m=0}^{\infty} \bigcup_{n=m+1}^{\infty} \{\omega \ : \ \lambda_k(\omega) = m, \lambda_{k+1}(\omega) = n, R_k(\omega) \in A\} \times [m\beta, n\beta)
\end{aligned}$$

which is a measurable set.

To get a scheme for forward estimation we follow Bailey [4] and shift this backward scheme to give estimations for the future. For $t > 0$ consider the estimator

$$\hat{R}_t(\omega) = \hat{R}_{-t}(T_t\omega) \tag{17}$$

which is defined in terms of $\{X_s : s \in [0,t]\}$ in the same way as $\hat{R}_{-t}(\omega)$ was defined in terms of $\{X_s : s \in [-t,0]\}$. Now $\hat{R}_t(\omega)$ is jointly measurable in (ω,t). Indeed,

$$\begin{aligned}
\hat{R}_t(\omega) &= \hat{R}_{-t}(T_t\omega) \\
&= \sum_{l=0}^{\infty} \hat{R}_{-l\beta}(T_t\omega) I_{\{l\beta \le t < (l+1)\beta\}} \\
&= \lim_{N\to\infty} \sum_{l=0}^{N} \hat{R}_{-l\beta}(T_t\omega) I_{\{l\beta \le t < (l+1)\beta\}}.
\end{aligned}$$

Now for a fixed $l \in \{0,1,2,\dots\}$,

$$\hat{R}_{-l\beta}(T_t\omega) = \sum_{k=0}^{\infty} R_k(T_t\omega) I_{\{\lambda_k \le l\beta < \lambda_{k+1}\}}(T_t\omega)$$

is jointly measurable in (ω,t). Thus for a fixed $N \in \{0,1,2,\dots\}$,

$$\sum_{l=0}^{N} \hat{R}_{-l\beta}(T_t\omega) I_{\{l\beta \le t < (l+1)\beta\}}$$

is jointly measurable in (ω,t). Now the limit of measurable functions

$$\lim_{N\to\infty} \sum_{l=0}^{N} \hat{R}_{-l\beta}(T_t\omega) I_{\{l\beta \le t < (l+1)\beta\}}$$

is also jointly measurable in (ω,t).

The estimator $\hat{R}_t$ may be viewed as an on-line predictor of $X_{t+\beta}$.

Theorem. Let $\{X_t : t \in \mathbb{R}\}$ be a real-valued stationary continuous time process. Let $\beta > 0$ be arbitrary. Assume that

$$E\left(|X_0|\right) < \infty.$$

Then

$$\lim_{t\to\infty} \hat{R}_{-t} = E(X_\beta | X_s : s \in (-\infty, 0]) \quad \text{almost surely.} \tag{18}$$

If in addition

$$E\left(|X_0| \log^+(|X_0|)\right) < \infty$$

then

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \left| \hat{R}_u - E(X_{u+\beta} | X_s : s \in (-\infty, u]) \right| \mathrm{d}u = 0 \quad \text{almost surely} \tag{19}$$

and

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \left| \left| \hat{R}_u - X_{u+\beta} \right| - |E(X_{u+\beta} | X_s : s \in (-\infty, u]) - X_{u+\beta}| \right| \mathrm{d}u = 0 \tag{20}$$

almost surely. If in addition for some $1 < p < \infty$,

$$E(|X_0|^p) < \infty$$

then

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \left| \hat{R}_u - E(X_{u+\beta} | X_s : s \in (-\infty, u]) \right|^p \mathrm{d}u = 0 \quad \text{almost surely} \tag{21}$$

and

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \left| \left| \hat{R}_u - X_{u+\beta} \right|^p - |E(X_{u+\beta} | X_s : s \in (-\infty, u]) - X_{u+\beta}|^p \right| \mathrm{d}u = 0 \tag{22}$$

almost surely.

Note that (18) generalizes the result of Scarpellini (cf. Scarpellini [24]) in that we have dropped the assumption that the process is bounded and that the time instant β is special. (Scarpellini [24]) assumed that T_β is an ergodic transformation. We do not need such assumption for our results.)

Note that (19) and (21) state that $\hat{R}_u$ is an asymptotically consistent estimator of the conditional expectation $E(X_{u+\beta} | X_s : s \in (-\infty, u])$ in time (Cesaro) average almost surely.

Note that (20) and (22) state that $\hat{R}_u$ is asymptotically as good estimator for $X_{u+\beta}$ as the conditional expectation $E(X_{u+\beta} | X_s : s \in (-\infty, u])$, in time (Cesaro) average almost surely. This is particularly important for $p = 2$ where the conditional expectation mimimizes the least square error.

As for a possible application consider a device measuring the temperature in continuous time. The goal is to give an estimate for the temperature e. g. a month ahead based only on the measurements. According to (20) and (22) our estimate will be as good in time average as the conditional expectation itself which uses prior knowledge of the process distribution.

3. PROOF OF THE THEOREM

Let

$$K = \sup_{n=1,2,\dots} \sup_{x \in \mathbb{R}} |[x]^n - x|.$$

By (7), $K < \infty$. We will follow Algoet [3] to prove (18). For $m = 1, 2, \dots$ define the forward going version of τ_m as

$$\tilde{\tau}_m = \min\{t \in \{\beta, 2\beta, 3\beta, \dots\} : [X_{-j\beta/2^m+t}]^m = [X_{-j\beta/2^m}]^m \quad \text{for } j = 0, 1, \dots, \tfrac{\lambda_{m-1}2^m}{\beta}.\}.$$

Let r be a nonnegative integer and $b_j \in \{[x]^m : x \in \mathbb{R}\}$ for $j = 0, 1, \dots, r$. By stationarity, it follows that for arbitrary $C \subseteq \mathbb{R}$

$$\begin{aligned}
&P(\{\lambda_{m-1} = \frac{r\beta}{2^m}, [X_{-j\beta/2^m}]^m = b_j : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}\} \cap \{X_{-\tau_m+\beta} \in C\}) \\
= &\sum_{l=1}^{\infty} P(\{\lambda_{m-1} = \frac{r\beta}{2^m}, [X_{-j\beta/2^m}]^m = b_j : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}\} \\
&\cap \{\tau_m = l\beta, X_{-\tau_m+\beta} \in C\}) \\
= &\sum_{l=1}^{\infty} P(T_{-l\beta}(\{\lambda_{m-1} = \frac{r\beta}{2^m}, [X_{-j\beta/2^m}]^m = b_j : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}\} \\
&\cap \{\tau_m = l\beta, X_{-\tau_m+\beta} \in C\})) \\
= &\sum_{l=1}^{\infty} P(\{\lambda_{m-1} = \frac{r\beta}{2^m}, [X_{-j\beta/2^m}]^m = b_j : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}\} \\
&\cap \{\tilde{\tau}_m = l\beta, X_\beta \in C\}) \\
= &P(\{\lambda_{m-1} = \frac{r\beta}{2^m}, [X_{-j\beta/2^m}]^m = b_j : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}\} \cap \{X_\beta \in C\}),
\end{aligned}$$

which in turn implies that

$$\begin{aligned}
&P(X_{-\tau_m+\beta} \in C | [X_{-j\beta/2^m}]^m : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}) \\
= &P(X_\beta \in C | [X_{-j\beta/2^m}]^m : j = 0, 1, \dots, 2^m \frac{\lambda_{m-1}}{\beta}). \qquad (23)
\end{aligned}$$

(Cf. Morvai [16] , Morvai et al. [17] , Algoet [3] and Morvai and Weiss [18] .) Thus for $m = 1, 2, \dots$ the random variables $X_{-\tau_m+\beta}$ and X_β are identically distributed. Now we go back to the definition of the R_k in (9) and decompose the quantity we are trying to estimate into several pieces. We will use the decomposition and argument in Algoet [3] to prove (18).

$$R_k = \frac{1}{k} \sum_{1 \le j \le k} \left(X_{-\tau_j+\beta} - [X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} \right)$$

$$
\begin{aligned}
&+ \quad \frac{1}{k} \sum_{1 \le j \le k} ([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} \\
&\quad -E([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})) \\
&+ \quad \frac{1}{k} \sum_{1 \le j \le k} (E([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta}) \\
&\quad -E([X_\beta]^j I_{\{|[X_\beta]^j| \le j\}} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})) \\
&+ \quad \frac{1}{k} \sum_{1 \le j \le k} E([X_\beta]^j I_{\{|[X_\beta]^j| \le j\}} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta}) \\
&= \quad \xi_k + \eta_k + \theta_k + \zeta_k.
\end{aligned}
$$

By (23), (7) and since $E(|X_\beta| + K) < \infty$ we get that

$$
\sum_{j=1}^{\infty} P(|[X_{-\tau_j+\beta}]^j| > j) = \sum_{j=1}^{\infty} P(|[X_\beta]^j| > j) \le \sum_{j=1}^{\infty} P(|X_\beta| + K > j) < \infty
$$

and by the Borel–Cantelli lemma,

$$
I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} = 1 \quad \text{eventually almost surely.} \tag{24}
$$

By (24) and (7),

$$
\lim_{j \to \infty} |X_{-\tau_j+\beta} - [X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}}| = \lim_{j \to \infty} |X_{-\tau_j+\beta} - [X_{-\tau_j+\beta}]^j| = 0
$$

almost surely. Thus

$$
|\xi_k| \to 0 \quad \text{almost surely.} \tag{25}
$$

Toward mastering η_k, one observes that $\{X_{-\tau_j+\beta}\}$ are identically distributed by (23) and by Proposition 4.1 in the Appendix

$$
\begin{aligned}
V_n = \sum_{j=1}^{n} (&\frac{[X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}}}{j} \\
- &\frac{E([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})}{j})
\end{aligned}
$$

is a martingale with

$$
\sup_{1 \le n} E(|V_n|) < \infty.
$$

By Doob's convergence theorem V_n converges almost surely. Then by Kronecker's lemma (cf. Shiryayev [26] p. 365),

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} ([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j| \le j\}}
$$

$$- \quad E([X_{-\tau_j+\beta}]^j I_{\{|[X_{-\tau_j+\beta}]^j|\le j\}}|[X_{-l\beta/2^j}]^j : l=0,1,\dots,2^j\frac{\lambda_{j-1}}{\beta})) = 0$$

almost surely and we have proven that

$$\eta_k \to 0 \text{ almost surely.} \tag{26}$$

(Alternatively, one could use Theorem 2.15 and the results in the proof of Theorem 2.19 in Hall and Heyde [14] as in Algoet [3] to prove (26)).
Now we will deal with θ_k. By (23) we get that

$$\theta_k = 0 \text{ almost surely.} \tag{27}$$

Now we deal with the last term ζ_k. Since

$$\sigma\{[X_{-l\beta/2^j}]^j : l=0,1,\dots,2^j\frac{\lambda_{j-1}}{\beta}\} \uparrow \sigma\{X_{-l\beta/2^m} : m=1,2,\dots,l=0,1,\dots,\},$$

$$[X_\beta]^j I_{\{|[X_\beta]^j|\le j\}} \to X_\beta \text{ almost surely,}$$

$$\sup_{j\ge 1} |[X_\beta]^j I_{\{|[X_\beta]^j|\le j\}}| \le |X_\beta| + K,$$

and

$$E(|X_\beta| + K) < \infty,$$

by Corollary 1 pp. 237–238 in Chow and Teicher [7] (Lemma 3 in Algoet [3]) we get

$$\begin{aligned}&\lim_{j\to\infty} E([X_\beta]^j I_{\{|[X_\beta]^j|\le j\}}|[X_{-l\beta/2^j}]^j : l=0,1,\dots,2^j\frac{\lambda_{j-1}}{\beta})\\ &= \quad E(X_\beta|X_{-l\beta/2^m} : m=1,2,\dots,l=0,1,\dots,) \text{ almost surely.}\end{aligned}$$

Thus

$$\zeta_k \to E(X_\beta|X_{-l\beta/2^m} : m=1,2,\dots,l=0,1,\dots,) \text{ almost surely.}$$

The set

$$\{-l\beta/2^j : j=1,2,\dots,l=0,1,\dots,\}$$

is a dense subset of the interval $(-\infty,0]$ and this implies that the sigma-algebra generated by the random variables

$$\{X_{-l\beta/2^m} : m=1,2,\dots,l=0,1,\dots,\}$$

coincides up to null sets with the sigma-algebra generated by the random variables

$$\{X_s : s\in(-\infty,0]\}$$

(cf. e.g. Proposition III.4.3 on p. 89 in Neveu [19]) and this yields

$$E(X_\beta|X_{-l\beta/2^m} : m=1,2,\dots,l=0,1,\dots,) = E(X_\beta|X_s : s\in(-\infty,0])$$

almost surely and in turn

$$\zeta_k \to E(X_\beta | X_s : s \in (-\infty, 0]) \ \text{ almost surely.} \tag{28}$$

By (25), (26), (27) and (28) we get

$$\lim_{k\to\infty} R_k = E(X_\beta | X_s : s \in (-\infty, 0]) \ \text{ almost surely.} \tag{29}$$

Now (13) and (29) together imply (18).

Now assume that

$$E\left(|X_0| \log^+(|X_0|)\right) < \infty.$$

We go back to the definition of the R_k in (9) and give a different decomposition.

$$\begin{aligned} R_k &= \frac{1}{k} \sum_{1\le j\le k} \left(X_{-\tau_j+\beta} - [X_{-\tau_j+\beta}]^j\right) \\ &+ \frac{1}{k} \sum_{1\le j\le k} \left([X_{-\tau_j+\beta}]^j - E([X_{-\tau_j+\beta}]^j | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})\right) \\ &+ \frac{1}{k} \sum_{1\le j\le k} (E([X_{-\tau_j+\beta}]^j | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta}) \\ &\quad -E(X_{-\tau_j+\beta} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})) \\ &+ \frac{1}{k} \sum_{1\le j\le k} (E(X_{-\tau_j+\beta} | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta}) \\ &\quad -E(X_\beta | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})) \\ &+ \frac{1}{k} \sum_{1\le j\le k} E(X_\beta | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta}) \\ &= A_k + B_k + C_k + D_k + E_k. \end{aligned}$$

By (7), we get

$$|A_k| + |C_k| \le 2K < \infty \ \text{ almost surely.} \tag{30}$$

Now we will deal with D_k. Using (23) we get that

$$D_k = 0 \ \text{ almost surely.} \tag{31}$$

Toward mastering B_k, one observes that $\{X_{-\tau_j+\beta}\}$ are identically distributed by (23) and by Proposition 4.1 in the Appendix

$$U_n = \sum_{j=1}^{n} \frac{[X_{-\tau_j+\beta}]^j - E([X_{-\tau_j+\beta}]^j | [X_{-l\beta/2^j}]^j : l = 0, 1, \ldots, 2^j \frac{\lambda_{j-1}}{\beta})}{j}$$

is a martingale with

$$E(\sup_{1\le n} |U_n|) < \infty$$

and since for any sequence of real numbers $\{a_i\}$,

$$\sup_{1\le n}\frac{1}{n}\left|\sum_{i=1}^{n} a_i\right| \le 2\left(\sup_{1\le n}\left|\sum_{i=1}^{n}\frac{1}{i}a_i\right|\right),$$

(cf. Lemma 7 in Elton [9]), we get

$$E(\sup_{1\le k}|B_k|) \le 2E(\sup_{1\le n}|U_n|) < \infty. \tag{32}$$

Furthermore, by Doob's inequality,

$$E(\sup_{1\le k}|E_k|) \le E(\sup_{1\le j}E(|X_\beta||[X_{-l\beta/2^j}]^j : l=0,1,\dots,2^j\frac{\lambda_{j-1}}{\beta})) < \infty. \tag{33}$$

By (30), (31), (32) and (33)

$$E\left(\sup_{1\le k}|R_k|\right) \le E\left(\sup_{1\le k}|A_k|+|B_k|+|C_k|+|D_k|+|E_k|\right) < \infty. \tag{34}$$

By (13),

$$\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}| = \sup_{k=0,1,2,\dots}|R_k| = \sup_{k=1,2,\dots}|R_k| \ \text{ almost surely.} \tag{35}$$

Now (35) and (34) together yield

$$E\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|\right) < \infty. \tag{36}$$

For $t\in[0,\infty)$ let $f_t(\omega):\Omega\times[0,\infty)\to\mathbb{R}$ be

$$\begin{aligned} f_t(\omega) &= |\hat{R}_{-t} - E(X_\beta|X_s : s\in(-\infty,0])| \\ &= \sum_{l=0}^{\infty}|\hat{R}_{-l\beta} - E(X_\beta|X_s : s\in(-\infty,0])|I_{\{l\beta\le t<(l+1)\beta\}}. \end{aligned}$$

Now $f_t(\omega)$ is nonnegative and jointly measurable in t and ω, cf. (14). For a fixed ω, $f_t(\omega)$ is right semi continuous in t, cf. (13). By (18) it is then immediate that

$$\lim_{l\to\infty}|\hat{R}_{-l\beta} - E(X_\beta|X_s : s\in(-\infty,0])| = 0 \ \text{ almost surely.} \tag{37}$$

By (36)

$$\begin{aligned} &E\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta} - E(X_\beta|X_s : s\in(-\infty,0])|\right) \\ &\le E\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|\right) + E\left(E(|X_\beta||X_s : s\in(-\infty,0])\right) \end{aligned}$$

$$
\begin{aligned}
&= E\left(\sup_{l=0,1,2,\ldots} |\hat{R}_{-l\beta}|\right) + E\left(|X_\beta|\right) \\
&< \infty.
\end{aligned}
$$

Now apply Proposition 4.3 in the Appendix to conclude that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f_u(T_u\omega)\,\mathrm{d}u = 0$$

almost surely. Thus

$$
\begin{aligned}
&\lim_{t\to\infty} \frac{1}{t}\int_0^t \left|\hat{R}_u - E(X_{u+\beta}|X_s : s \in (-\infty, u])\right| \mathrm{d}u \\
&\quad= \lim_{t\to\infty} \frac{1}{t}\int_0^t \left(\left|\hat{R}_{-u} - E(X_\beta|X_s : s \in (-\infty, 0])\right| (T_u\omega)\right) \mathrm{d}u \\
&\quad= 0
\end{aligned}
$$

almost surely and the proof of (19) is complete. Similarly,

$$
\begin{aligned}
&\left||\hat{R}_{-t} - X_\beta| - |E(X_\beta|X_s : s \in (-\infty, 0]) - X_\beta|\right| \\
&\quad= \sum_{l=0}^{\infty} \left||\hat{R}_{-l\beta} - X_\beta| - |E(X_\beta|X_s : s \in (-\infty, 0]) - X_\beta|\right| I_{\{l\beta \le t < (l+1)\beta\}}
\end{aligned}
$$

and by (18)

$$\lim_{l\to\infty} \left||\hat{R}_{-l\beta} - X_\beta| - |E(X_\beta|X_s : s \in (-\infty, 0]) - X_\beta|\right| = 0$$

almost surely and by (36)

$$
\begin{aligned}
&E\left(\sup_{l=0,1,2,\ldots} \left||\hat{R}_{-l\beta} - X_\beta| - |E(X_\beta|X_s : s \in (-\infty, 0]) - X_\beta|\right|\right) \\
&\quad\le E\left(\sup_{l=0,1,2,\ldots} |\hat{R}_{-l\beta}|\right) + 3E\left(|X_\beta|\right) \\
&\quad< \infty
\end{aligned}
$$

and Proposition 4.3 in the Appendix gives

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \left||\hat{R}_{-t} - X_\beta| - |E(X_\beta|X_s : s \in (-\infty, 0]) - X_\beta|\right| \mathrm{d}u = 0$$

almost surely and the proof of (20) is complete.
Now we assume that for some $1 < p < \infty$, $E(|X_0|^p) < \infty$, and we prove (21).

Observe that by (31) and (30)

$$|R_k|^p = |A_k + B_k + C_k + D_k + E_k|^p \le 3^p\left[(2K)^p + |B_k|^p + |E_k|^p\right]. \tag{38}$$

By Proposition 4.2 in the Appendix

$$E(\sup_{1\le k}|B_k|^p) < \infty \tag{39}$$

and by Doob's inequality, (cf. Theorem 1 on p. 464, §3 Ch. VII in Shiryayev [26]),

$$E(\sup_{1\le k}|E_k|^p) < \infty. \tag{40}$$

By (38), (39) and (40),

$$E\left(\sup_{1\le k}|R_k|^p\right) < \infty. \tag{41}$$

Now (41) and (35) together imply

$$E\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|^p\right) < \infty. \tag{42}$$

$$|\hat{R}_{-t} - E(X_\beta|X_s : s\in(-\infty,0])|^p = \sum_{l=0}^{\infty}|\hat{R}_{-l\beta} - E(X_\beta|X_s : s\in(-\infty,0])|^p I_{\{l\beta\le t<(l+1)\beta\}}$$

and by (18)

$$\lim_{l\to\infty}|\hat{R}_{-l\beta} - E(X_\beta|\{X_s : s\in(-\infty,0]\})|^p = 0 \ \text{ almost surely}$$

and by (42)

$$\begin{aligned}
&E\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta} - E(X_\beta|X_s : s\in(-\infty,0])|^p\right)\\
&\le \ 2^pE\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|^p\right)\\
&+ \ 2^pE\left(|E(X_\beta|X_s : s\in(-\infty,0])|^p\right)\\
&\le \ 2^pE\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|^p\right)\\
&+ \ 2^pE\left(|X_\beta|^p\right)\\
&< \ \infty
\end{aligned}$$

and by Proposition 4.3 in the Appendix one gets (21). Similarly,

$$\begin{aligned}
&\Big||\hat{R}_{-t} - X_\beta|^p - |E(X_\beta|X_s : s\in(-\infty,0]) - X_\beta|^p\Big|\\
&= \ \sum_{l=0}^{\infty}\Big||\hat{R}_{-l\beta} - X_\beta|^p - |E(X_\beta|X_s : s\in(-\infty,0]) - X_\beta|^p\Big| I_{\{l\beta\le t<(l+1)\beta\}}
\end{aligned}$$

and by (37)

$$\lim_{l\to\infty}\left||\hat{R}_{-l\beta}-X_\beta|^p-|E(X_\beta|X_s:s\in(-\infty,0])-X_\beta|^p\right|=0 \text{ almost surely}$$

and by (42)

$$\begin{aligned} &E\left(\sup_{l=0,1,2,\dots}\left||\hat{R}_{-l\beta}-X_\beta|^p-|E(X_\beta|X_s:s\in(-\infty,0])-X_\beta|^p\right|\right)\\ \leq\ & 2^pE\left(\sup_{l=0,1,2,\dots}|\hat{R}_{-l\beta}|^p\right)+3(2^p)E\left(|X_\beta|^p\right)\\ <\ & \infty. \end{aligned}$$

Now apply Proposition 4.3 in the Appendix to prove (22). The proof of the Theorem is complete. □

4. APPENDIX

The next result is a generalization of a result due to Elton, cf. Theorems 2 and 4 in Elton [9].

Proposition 4.1. (Cf. Elton [9], Hall and Heyde [14], Algoet [3] and Morvai and Weiss [18]) For $n=0,1,2,\dots$ let X_n be random variables identically distributed with

$$E(|X_0|)<\infty$$

and let $\mathcal{G}_n$ be an increasing sequence of σ-algebras. For $n=1,2,\dots$ let g_n be a real valued functions such that

$$\sup_{n=1,2,\dots}\ \sup_{x\in\mathbb{R}}|g_n(x)-x|<\infty$$

and $g_n(X_n)$ is measurable with respect to $\mathcal{G}_n$. Then

$$E\left(\sup_{1\leq n}\left|\sum_{i=1}^n\frac{g_i(X_i)I_{\{|g_i(X_i)|\leq i\}}-E(g_i(X_i)I_{\{|g_i(X_i)|\leq i\}}|\mathcal{G}_{i-1})}{i}\right|\right)<\infty. \tag{43}$$

If in addition

$$E(|X_0|\log^+(|X_0|))<\infty$$

then

$$E\left(\sup_{1\leq n}\left|\sum_{i=1}^n\frac{g_i(X_i)-E(g_i(X_i)|\mathcal{G}_{i-1})}{i}\right|\right)<\infty. \tag{44}$$

Proof. Write

$$Y_n=g_n(X_n),$$

$$Y_n'=Y_nI_{\{|Y_n|\leq n\}}$$

and

$$Y_n'' = Y_n I_{\{|Y_n|>n\}}.$$

By Davis' inequality (valid for all martingale differences cf. e.g. Shiryayev [26] p. 470), we get

$$\begin{aligned} E\left(\sup_{1\le n}\left|\sum_{i=1}^{n}\frac{Y_i'-E(Y_i'|\mathcal{G}_{i-1})}{i}\right|\right) &\le BE\left[\left(\sum_{i=1}^{\infty}\frac{(Y_i'-E(Y_i'|\mathcal{G}_{i-1}))^2}{i^2}\right)^{0.5}\right] \\ &\le B\left[E\left(\sum_{i=1}^{\infty}\frac{(Y_i'-E(Y_i'|\mathcal{G}_{i-1}))^2}{i^2}\right)\right]^{0.5} \\ &= B\left[\sum_{i=1}^{\infty}\frac{E\left((Y_i'-E(Y_i'|\mathcal{G}_{i-1}))^2\right)}{i^2}\right]^{0.5}. \end{aligned}$$

Now

$$\begin{aligned} E\left((Y_i'-E(Y_i'|\mathcal{G}_{i-1}))^2\right) &= E\left((Y_i')^2\right)+E\left(E(Y_i'|\mathcal{G}_{i-1})^2\right) \\ &- 2E\left(Y_i'E(Y_i'|\mathcal{G}_{i-1})\right) \\ &= E\left((Y_i')^2\right)-E\left(E(Y_i'|\mathcal{G}_{i-1})^2\right) \\ &\le E\left((Y_i')^2\right). \end{aligned}$$

Define

$$K := \sup_{n=1,2,\ldots}\sup_{x\in\mathbb{R}}|g_n(x)-x| < \infty.$$

But since $|Y_i - X_i| \le K$ we get

$$E\left((Y_i')^2\right) = E\left((Y_i)^2 I_{\{|Y_i|\le i\}}\right) \le E\left((|X_i|+K)^2 I_{\{|X_i|\le i+K\}}\right)$$

and the X_i's are identically distributed therefore

$$\begin{aligned} &\sum_{i=1}^{\infty}\frac{1}{i^2}E\left((|X_i|+K)^2 I_{\{|X_i|\le i+K\}}\right) \\ &= \sum_{i=1}^{\infty}\frac{1}{i^2}E\left((|X_0|+K)^2 I_{\{|X_0|\le i+K\}}\right) \\ &\le \sum_{i=1}^{\infty}\frac{1}{i^2}E\left(4|X_0|^2 I_{\{|X_0|\le i+K\}}\right) \\ &+ \sum_{i=1}^{\infty}\frac{4K^2}{i^2} \end{aligned}$$

where $4K^2\sum_{i=1}^{\infty}\frac{1}{i^2}$ is finite. Now

$$\sum_{i=1}^{\infty}\frac{1}{i^2}E\left(|X_0|^2 I_{\{|X_0|\le i+K\}}\right) = \sum_{i=1}^{\infty}\frac{1}{i^2}E\left(|X_0|^2 I_{\{|X_0|\le i\}}\right)+\sum_{i=1}^{\infty}\frac{1}{i^2}E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{\infty} \frac{1}{i^2} \sum_{j=1}^{i} E\left(|X_0|^2 I_{\{j-1<|X_0|\le j\}}\right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&= \sum_{i=1}^{\infty} \left(E\left(|X_0|^2 I_{\{i-1<|X_0|\le i\}}\right) \left(\sum_{j=i}^{\infty} \frac{1}{j^2} \right) \right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&= \sum_{i=1}^{\infty} \left(E\left(|X_0|^2 I_{\{i-1<|X_0|\le i\}}\right) \left(\frac{1}{i^2} + \sum_{j=i+1}^{\infty} \frac{1}{j^2} \right) \right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&\le \sum_{i=1}^{\infty} \left(E\left(|X_0|^2 I_{\{i-1<|X_0|\le i\}}\right) \left(\frac{1}{i^2} + \int_{i}^{\infty} \frac{1}{z^2} dz \right) \right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&\le \sum_{i=1}^{\infty} \left(E\left(|X_0|^2 I_{\{i-1<|X_0|\le i\}}\right) \left(\frac{1}{i^2} + \frac{1}{i} \right) \right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&\le \sum_{i=1}^{\infty} \left(E\left(|X_0|^2 I_{\{i-1<|X_0|\le i\}}\right) \frac{2}{i} \right) \\
&+ \sum_{i=1}^{\infty} \frac{1}{i^2} E\left(|X_0|^2 I_{\{i<|X_0|\le i+K\}}\right) \\
&= 2\sum_{i=1}^{\infty} \left(E\left(\frac{|X_0|}{i} |X_0| I_{\{i-1<|X_0|\le i\}} \right) \right) \\
&+ \sum_{i=1}^{\infty} E\left(\frac{|X_0|^2}{i^2} I_{\{i<|X_0|\le i+K\}} \right) \\
&\le 2\sum_{i=1}^{\infty} \left(E\left(|X_0| I_{\{i-1<|X_0|\le i\}}\right) \right) \\
&+ \sum_{i=1}^{\infty} E\left((K+1)^2 I_{\{i<|X_0|\le i+K\}}\right)
\end{aligned}
$$

$$\leq \quad 2E(|X_0|) + (K+1)^2 K < \infty.$$

Combining all these we get (43), (cf. Theorem 2.19 in Hall and Heyde [14] also). Now we assume that $E(|X_0| \log^+(|X_0|)) < \infty$.

$$\begin{aligned} E|Y_n'' - E(Y_n''|\mathcal{G}_{n-1})| &\leq 2E|Y_n''| \\ &\leq 2E\left((K + |X_n|) I_{\{|X_n| > n-K\}}\right) \\ &= 2E\left((K + |X_0|) I_{\{|X_0| > n-K\}}\right) \end{aligned}$$

since X_n' are identically distributed. Now

$$\begin{aligned} E\left(\sum_{n=1}^{\infty} \frac{|Y_n'' - E(Y_n''|\mathcal{G}_{n-1})|}{n}\right) &\leq 2\sum_{n=1}^{\infty} \frac{1}{n} E\left((K + |X_0|) I_{\{|X_0| > n-K\}}\right) \\ &= 2\sum_{n=1}^{\infty} \frac{1}{n} E\left((K + |X_0|) I_{\{|X_0| + K > n\}}\right). \end{aligned}$$

Since $E((|X_0| + K) \log^+(|X_0| + K) < \infty$, Lemma 2 in Elton [9] implies that

$$\sum_{n=1}^{\infty} \frac{1}{n} E\left((|X_0| + K) I_{\{|X_0| + K > n\}}\right) < \infty$$

and so

$$E\left(\sup_{1\leq n}\left|\sum_{i=1}^{n} \frac{Y_i'' - E(Y_i''|\mathcal{G}_{i-1})}{i}\right|\right) \leq E\left(\sum_{n=1}^{\infty} \frac{|Y_n'' - E(Y_n''|\mathcal{G}_{n-1})|}{n}\right) < \infty. \tag{45}$$

Now by (43) and (45) we get (44). The proof of Proposition 4.1 is complete. □

Proposition 4.2. (Proposition 2 in Morvai and Weiss [18]) Let ϕ_n be a martingale difference sequence. If, for some $1 < p < \infty$,

$$\sup_{1\leq n} E(|\phi_n|^p) < \infty$$

then

$$E\left(\sup_{1\leq n}\left|\frac{1}{n}\sum_{i=1}^{n} \phi_i\right|^p\right) < \infty. \tag{46}$$

Now we adapt the method of proofs in Maker [15], Breiman [5] and Algoet [2] to our needs.

Proposition 4.3. (Cf. Maker [15], Breiman [5] and Algoet [2]) Let (Ω, Φ, P) be a probability space with a family of measure preserving invertible transformations T_t, $-\infty < t < \infty$ with the group property $(T_s T_r = T_{s+r})$ such that $T : \Omega \times [0, \infty) \to \mathbb{R}$ is jointly measurable in (ω, t). For $l = 0, 1, 2, \ldots$ let h_l be measurable real valued functions such that h_l is nonnegative, $E(\sup_{l=0,1,2,\ldots} h_l) < \infty$ and $\lim_{l\to\infty} h_l = 0$ almost surely. For a given fixed $\beta > 0$ let $f_t(\omega) : \Omega \times [0,\infty) \to \mathbb{R}$ be such that $f_t(\omega) = \sum_{l=0}^{\infty} h_l(\omega) I_{\{l\beta \le t < (l+1)\beta\}}$. Then

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f_u(T_u\omega) \mathrm{d}u = 0 \text{ almost surely.} \tag{47}$$

P r o o f. We follow Algoet's proof for the discrete time case (Theorem 12 in [2]) and adapt to our needs in the continuous time. First note that

$$f_t(T_t\omega) = \sum_{l=0}^{\infty} h_l(T_t\omega) I_{\{l\beta \le t<(l+1)\beta\}} = \lim_{N\to\infty} \sum_{l=0}^{N} h_l(T_t\omega) I_{\{l\beta \le t<(l+1)\beta\}}$$

is jointly measurable in (ω, t) since it is a pointwise limit of sums of measurable functions (Cf. Maker [15]) . For $k = 0, 1, 2, \ldots$ define

$$G_k(\omega) = \sup_{l=k,k+1,\ldots} h_l(\omega).$$

Note that G_k is nonnegative, monotone decreasing and $E(G_0) < \infty$. Furthermore, by the Fubini–Tonelli theorem and stationarity

$$E \int_0^t f_u(T_u\omega)\, \mathrm{d}u = \int_0^t E f_u(T_u\omega) \mathrm{d}u = \int_0^t E f_u \, \mathrm{d}u \le \int_0^t E G_0 \, \mathrm{d}u = t E G_0 < \infty$$

and thus the integrals exist. Now

$$\begin{aligned}
&\frac{1}{(j+1)\beta} \int_0^{(j+1)\beta} f_u(T_u\omega)\, \mathrm{d}u \\
&= \frac{1}{(j+1)\beta} \int_0^{(j+1)\beta} \sum_{l=0}^{\infty} h_l(T_u\omega) I_{\{l\beta \le u<(l+1)\beta\}} \, \mathrm{d}u \\
&= \frac{1}{(j+1)\beta} \int_0^{(j+1)\beta} \sum_{l=0}^{j} h_l(T_u\omega) I_{\{l\beta \le u<(l+1)\beta\}} \, \mathrm{d}u \\
&= \frac{1}{(j+1)\beta} \sum_{l=0}^{j} \int_0^{(j+1)\beta} h_l(T_u\omega) I_{\{l\beta \le u<(l+1)\beta\}} \, \mathrm{d}u \\
&= \frac{1}{(j+1)\beta} \sum_{l=0}^{j} \int_{l\beta}^{(l+1)\beta} h_l(T_u\omega)\, \mathrm{d}u \\
&= \frac{1}{(j+1)\beta} \sum_{l=0}^{k} \int_{l\beta}^{(l+1)\beta} h_l(T_u\omega)\, \mathrm{d}u + \frac{1}{(j+1)\beta} \sum_{l=k+1}^{j} \int_{l\beta}^{(l+1)\beta} h_l(T_u\omega)\, \mathrm{d}u \\
&\le \frac{1}{(j+1)\beta} \sum_{l=0}^{k} \int_{l\beta}^{(l+1)\beta} G_0(T_u\omega)\, \mathrm{d}u + \frac{1}{(j+1)\beta} \sum_{l=k+1}^{j} \int_{l\beta}^{(l+1)\beta} G_k(T_u\omega)\, \mathrm{d}u \\
&\le \frac{1}{(j+1)\beta} \sum_{l=0}^{k} \int_{l\beta}^{(l+1)\beta} G_0(T_u\omega)\, \mathrm{d}u + \frac{1}{(j+1)\beta} \sum_{l=0}^{j} \int_{l\beta}^{(l+1)\beta} G_k(T_u\omega)\, \mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
&\leq \quad \frac{1}{(j+1)\beta}\int_0^{(k+1)\beta} G_0(T_u\omega)\,\mathrm{d}u + \frac{1}{(j+1)\beta}\int_0^{(j+1)\beta} G_k(T_u\omega)\,\mathrm{d}u \\
&\to \quad 0 + E(G_k|\mathcal{I})
\end{aligned}
$$

where $\mathcal{I}$ is the sigma algebra of the invariant sets. (Cf. Maker [15], Breiman [5] and Algoet [2]) Since G_k is nonnegative monotone decreasing and $E(G_0) < \infty$ we get that $E(G_k|\mathcal{I}) \to 0$ almost surely. Thus

$$\frac{1}{(j+1)\beta}\int_0^{(j+1)\beta} f_u(T_u\omega)\,\mathrm{d}u \to 0$$

almost surely. Now for $j\beta < t < (j+1)\beta$,

$$0 \leq \frac{1}{t}\int_0^t f_u(T_u\omega)\,\mathrm{d}u \leq \frac{j+1}{j}\frac{1}{(j+1)\beta}\int_0^{(j+1)\beta} f_u(T_u\omega)\,\mathrm{d}u$$

and the right hand side tends to zero almost surely which yields (47). This completes the proof of Proposition 4.3. □

ACKNOWLEDGEMENT

This first author was supported by Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, 13-15 Reáltanoda utca , H-1053, Budapest, Hungary.

(Received April 15, 2019)

REFERENCES

[1] P. Algoet: Universal schemes for prediction, gambling and portfolio selection. Ann. Probab. *20* (1992), 901–941. DOI:10.1214/aop/1176989811

[2] P. Algoet: The strong law of large numbers for sequential decisions under uncertainty. IEEE Trans. Inform. Theory *40* (1994), 609–633. DOI:10.1109/18.335876

[3] 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. DOI:10.1109/18.761258

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

[5] L. Breiman: The individual ergodic theorem of information theory, Ann. Math. Statist. *28* (1957), 809–811.

[6] T. Cover: Open problems in information theory. In: 1975 IEEE-USSR Joint Workshop on Information Theory 1975, pp. 35–36.

[7] Y. S. Chow and H. Teicher: Probability Theory: Independence, Interchangeability, Martingales. Second edition. Springer-Verlag, New York 1978.

[8] J. L. Doob: Stochastic Processes. Wiley, 1990

[9] J. Elton: A law of large numbers for identically distributed martingale differences. Ann. Probab. *9* (1981), 405–412. DOI:10.1214/aop/1176994414

[10] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk: A Distribution-Free Theory of Nonparametric Regression. Springer Series in Statistics, Springer-Verlag, New York 2002. DOI:10.1007/b97848

[11] 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. DOI:10.1109/18.661540

[12] L. Györfi and Gy. Ottucsák: Sequential prediction of unbounded stationary time series. IEEE Trans. Inform. Theory *53* (2007), 1866–1872. DOI:10.1109/tit.2007.894660

[13] L. Györfi, Gy. Ottucsák, and H. Walk: Machine Learning for Financial Engineering. Imperial College Press, London 2012. DOI:10.1142/p818

[14] P. Hall and C. C. Heyde: Martingale Limit Theory and Its Application. Academic Prress, 1975.

[15] Ph. T. Maker: The ergodic theorem for a sequence of functions. Duke Math. J. *6* (1940), 27–30. DOI:10.1215/s0012-7094-40-00602-0

[16] G. Morvai: Estimation of Conditional Distribution for Stationary Time Series. Ph.D. Thesis, Technical University of Budapest 1994.

[17] G. Morvai, S. Yakowitz, and L. Györfi: Nonparametric inferences for ergodic, stationary time series. Ann. Statist. *24* (1996), 370–379. DOI:10.1214/aos/1033066215

[18] G. Morvai and B. Weiss: Nonparametric sequential prediction for stationary processes. Ann. Prob. *39* (2011), 1137–1160. DOI:10.1214/10-aop576

[19] J. Neveu: Mathematical Foundations of the Calculus of Probability. Holden-Day, 1965.

[20] D. Ornstein: Guessing the next output of a stationary process. Israel J. of Math. *30* (1978), 292–296. DOI:10.1007/bf02761077

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

[22] B. Scarpellini: Predicting the future of functions on flows. Math. Systems Theory *12* (1979), 281–296. DOI:10.1007/bf01776579

[23] B. Scarpellini: Entropy and nonlinear prediction. Probab. Theory Related Fields *50* (1079, 2, 165–178. DOI:10.1007/bf00533638

[24] B. Scarpellini: Conditional expectations of stationary processes. Z. Wahrsch. Verw. Gebiete *56* (1981), 4, 427–441. DOI:10.1007/bf00531426

[25] P. C. Shields: Cutting and stacking: a method for constructing stationary processes. IEEE Trans. Inform. Theory *37* (1991), 1605–1614. DOI:10.1109/18.104321

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

[27] B. Weiss: Single Orbit Dynamics. American Mathematical Society, 2000.

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

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