Intermittent Estimation for Gaussian Processes Appeared in: IEEE Transactions on Information Theory, Vol. 56. No. 6, June, pp. 2778–2782, 2010.

Gábor Molnár-Sáska, Gusztáv Morvai

***Abstract*—Let $\{X_n\}_{n=0}^{\infty}$ be a stationary real-valued Gaussian time series. We estimate the conditional expectation $E(X_{n+1}|X_0,\dots,X_n)$ from a growing number of observations $X_0,\dots,X_n$ in a pointwise consistent way along a sequence of stopping times.**

***Index Terms*—Gaussian process, estimation, conditional expectation, stopping time**

I. Introduction

Suppose $\{X_n\}_{n=0}^{\infty}$ is a stationary real-valued time series with apriori unknown distribution. The goal is to estimate the conditional expectation $E(X_{n+1}|X_0,\dots,X_n)$ from the observations $X_0,\dots,X_n$ such that the difference between the estimate and the conditional expectation should tend to zero almost surely as the number of observations n tends to infinity. The importance of this estimation problem origins from the fact that the conditional expectation minimizes the conditional mean squared error.

This type of problem (for binary time series) was introduced in Cover [4]. When one is obliged to estimate for all n, Bailey [3], Ryabko [27], Györfi,Morvai and Yakowitz [9], Morvai and Weiss [16] proved the nonexistence of such a universal algorithm even over the class of all stationary and ergodic binary time series.

However using Cesaro mean one can estimate the conditional mean for all n which is a much simpler problem, for the discrete case see Ornstein [25], Györfi Lugosi and Morvai [8], and for the real valued case see Algoet [1], [2], Morvai, Yakowitz and Algoet [23], Morvai, Yakowitz and Györfi [24], Györfi and Lugosi [7], Morvai and Weiss [14].

In the intermittent estimation problem we consider the original problem (not the Cesaro mean), but instead of requiring estimaton for all time instances n we estimate merely along a stopping time sequence. That is, looking at the data segment $X_0,\dots,X_n$ our rule will decide if we estimate for this n or not, but anyhow we will definitely estimate for infinitely many n. Algorithms of this kind were proposed for binary time series in Morvai [11], Morvai and Weiss [12]. For a restricted class of real valued processes cf. Morvai and Weiss [13], [17], [15]. (For further reading see [18], [19], [20], [21], [22].)

Schäfer [28] considered stationary and ergodic Gaussian processes. He constructed an algorithm which can estimate the conditional expectation for every time instance n for an extremely restricted and narrow class of Gaussian processes.

In this paper we consider stationary Gaussian (not necessarily ergodic) processes and estimate the conditional mean along a stopping time sequence for a much wider class of processes than in Schäfer [28].

II. Results

Consider a stationary Gaussian process $\{X_n\}$ with autocovariance function $\gamma(k)=E(X_{n+k}X_n)$ and $EX_n=m$. Define the following subclasses of stationary Gaussian processes: In Φ_1 we have Gaussian processes satisfying the condition

$$\sum_{j=0}^{\infty}|\gamma(j)|<\infty \tag{1}$$

and are not Markovian of any order. In Φ_2 we have all Gaussian processes (not necessarily satifying (1)) which are Markov of some order. In this paper we are dealing with processes in $\Phi=\Phi_1\cup\Phi_2$. Note that Φ_2 is not a subset of Φ_1, see Example 2.5. Although estimating the conditional mean in the class Φ_2 is much easier, our algorithm will be valid universally for every process in Φ.

Example 2.1: Consider the class of Gaussian processes given by

$$X_n=\sum_{j=0}^{\infty}\psi_j\epsilon_{n-j}+m,$$

where $\psi_0=1$, $\sum_{j=0}^{\infty}|\psi_j|<\infty$ and ϵ_i-s are independent and identically distributed Gauss innovations distributed as $N(0,\sigma)$. Then condition (1) is satisfied and $\{X_n\}$ is a real-valued stationary and ergodic Gaussian process in Φ, see Hida and Hitsuda [10].

Let

$$\Gamma_n=(\gamma(|i-j|))_{i,j=1,\dots n}=$$

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \dots & & \dots & \dots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}$$

Gábor Molnár-Sáska is with Morgan Stanley Hungary Analytics Ltd, Budapest, Hungary, e-mail: gabor.molnar-saska@morganstanley.com.

Gusztáv Morvai is with MTA-BME Stochastics Research Group, 1 Egry József utca , Building H, Budapest,1111, Hungary, e-mail: morvai@math.bme.hu, The second author was supported by Bolyai János Research Scholarship and OTKA Grant No. K75143 during the revision of this paper.

and
$$\gamma_n = (\gamma(n), \dots, \gamma(1)).$$

For Gaussian processes with nonsingular Γ_n (every process in Φ_1 has this property) we have

$$E(X_n|X_0, \dots X_{n-1}) =$$
$$\gamma_n \Gamma_n^{-1}(X_0 - m, \dots, X_{n-1} - m)^T + m =$$
$$f_n^n(X_0 - m) + \dots + f_1^n(X_{n-1} - m) + m. \qquad (2)$$

For Gaussian processes in Φ with singular Γ_n, which are all in fact in Φ_2 let

$$E(X_n|X_0, \dots X_{n-1}) =$$
$$f_n^n(X_0 - m) + \dots + f_1^n(X_{n-1} - m) + m, \qquad (3)$$

where for $i > k$ (k is the order of the Markov process) $f_i^n = 0$. For references to (2) and (3) see chapters 7-8 in [6].

Schäfer [28] investigated the restricted model class considered in the following example.

Example 2.2: Consider the model class described in Example 2.1 with the very strong additional condition that the Taylor coefficients of

$$\frac{1}{\psi(z)} = \sum_{k=0}^{\infty} \varphi_k z^k \qquad (|z| > 1)$$

satisfy

$$\sum_{k=d_n+1}^{\infty} |\varphi_k|^2 \le \left(\frac{C}{\log n}\right)^r \qquad (4)$$

for sufficiently large n with some $C > 0$ and $r > 1$, where $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ is the transfer function for $|z| < 1$. For this model class Schäfer proved that the difference between his estimate and the conditional expectation $E(X_{n+1}|X_0^n)$ tends to zero as n tends to infinity. For general Gaussian processes it is hard to check condition (4). Two special extremely narrow classes of Gaussian processes have been given in Schäfer [28] where this condition is satisfied.

In this paper we only consider estimation along a sequence of stopping times. For general stationary (not necessarily Gaussian) processes the notion of intermittent estimation was introduced in Morvai and Weiss [13], [15]. There the notion of almost sure continuity was used to prove consistency. For this reason we need some basic definitions. Consider a two-sided stationary (not necessarily Gaussian) real-valued process $\{X_n\}_{n=-\infty}^{\infty}$. A one-sided stationary time series $\{X_n\}_{n=0}^{\infty}$ can always be considered to be a two-sided stationary time series $\{X_n\}_{n=-\infty}^{\infty}$. Let $\Re$ be the set of all real numbers and put $\Re^{*-}$ the set of all one-sided sequences of real numbers, that is,

$$\Re^{*-} = \{(\dots, x_{-1}, x_0) : x_i \in \Re \text{ for all } -\infty < i \le 0\}.$$

Define the metric $d^*(\cdot,\cdot)$ on $\Re^{*-}$ as

$$d^*((\dots, x_{-1}, x_0), (\dots, y_{-1}, y_0)) =$$
$$\sum_{i=0}^{\infty} 2^{-i-1} \frac{|x_{-i} - y_{-i}|}{1 + |x_{-i} - y_{-i}|}.$$

Definition 2.3: The conditional expectation $E(X_1|\dots, X_{-1}, X_0)$ is said to be almost surely continuous if for some set $B \subseteq \Re^{*-}$ which has probability one the conditional expectation $E(X_1|\dots, X_{-1}, X_0)$ restricted to this set B is continuous with respect to metric $d^*(\cdot,\cdot)$.

Morvai and Weiss [13], [15] suggested an algorithm and sequence of stopping times along which the error tends to zero almost surely under the condition that the conditional expectation $E(X_1|\dots, X_{-1}, X_0)$ is almost surely continuous. Unfortunately the conditional expectation $E(X_1|\dots, X_{-1}, X_0)$ is not almost surely continuous in the Gaussian case in general and so this result of Morvai and Weiss [13], [15] is not applicable for Gaussian processes in general. To prove it, first we need the following lemma. For notational convenience, let $X_m^n = (X_m, \dots, X_n)$, where $m \le n$.

Lemma 2.4: Consider a stationary Gaussian process $\{X_n\}$ and assume that it is not Markov of any order. Then for any $K \ge 0$ the conditional expectation $E(X_1|X_{-\infty}^0)$ has a non degenerate Gaussian conditional distribution given X_{-K}^0 almost surely.

Proof: Observe that $E(X_1|X_{-\infty}^0)$ is a normally distributed random variable. Let us fix a $K \ge 0$ and assume that the conditional expectation $E(X_1|X_{-\infty}^0)$ has degenerate conditional distribution given X_{-K}^0 on a set with positive probablity. We prove that this assumption leads to a contradiction.

The above assumption is equivalent with $P(D) > 0$, where

$$D = \{x_{-\infty}^0 : E(X_1|X_{-\infty}^0 = x_{-\infty}^0) = C(x_{-K}, \dots, x_0)\},$$

and $C(x_{-K}, \dots, x_0)$ is a function depending merely on $x_{-K}, \dots, x_0$.

Consider an arbitrary $k \ge K$. On one hand on D

$$E(X_1|X_{-k}^0) = E(E(X_1|X_{-\infty}^0)|X_{-k}^0) = C(X_{-K}, \dots, X_0),$$

on the other hand

$$E(X_1|X_{-k}^0) = f_1^{k+1}(X_0 - m) + \dots f_{k+1}^{k+1}(X_{-k} - m) + m.$$

Thus we have that for all $k \ge K$ on D

$$f_1^{k+1}(X_0 - m) + \dots f_{k+1}^{k+1}(X_{-k} - m) =$$
$$f_1^{K+1}(X_0 - m) + \dots f_{K+1}^{K+1}(X_{-K} - m).$$

Thus we have two linear functions which are equal on a set with positive Lebesgue measure. (Indeed, since the process is not Markov of any order, the $(X_0, \dots X_k)$ is a non-degenerate $k+1$ dimensional normal distribution.) This implies that the coefficients are equal. Hence we get that

$$f_i^K = f_i^k \quad \text{if} \quad i \le K$$

and $f_i^k = 0$ if $i > K$ for all $k \ge K$. Thus we have

$$E(X_1|X_{-k}^0) = f_1^{K+1}(X_0 - m) + \dots f_{K+1}^{K+1}(X_{-K} - m) + m$$

for all $k \ge K$. Considering the limit of the left hand side we get

$$\lim_{k\to\infty} E(X_1|X_{-k}^0) = E(X_1|X_{-\infty}^0) =$$
$$f_1^{K+1}(X_0 - m) + \dots f_{K+1}^{K+1}(X_{-K} - m) + m,$$

which implies that the Gaussian process is a Markov process with order at most $K+1$, which contradicts with the assumption of the lemma. ∎

Example 2.5: Let X_0 be a standard normal random variable. For $n \geq 0$ let $X_{n+1} = X_n$. Then $\{X_n\}$ is a stationary, non-ergodic, first order Gauss-Markov process in Φ_2. Since $\gamma_n = 1$ (1) is not satisfied. The distribution of X_{n+1} given X_0^n is degenerate.

Theorem 2.6: Consider a stationary Gaussian process $\{X_n\}$. The conditional expectation $E(X_1|X_{-\infty}^0)$ is not almost surely continuous if the process $\{X_n\}$ is not Markov of any order.

Proof: Assume that the statement of the theorem is not true and consider a Gaussian process $\{X_n\}$ which is not Markov and the conditional expectation $E(X_1|X_{-\infty}^0)$ is almost surely continuous. The convergence

$$E(X_1|X_{-\infty}^0) = \lim_{n\to\infty} E(X_1|X_{-n},\ldots,X_0) \tag{5}$$

is fulfilled almost surely. Let $\tilde{\Omega} \subset \Omega$ be such that on $\tilde{\Omega}$ the limit (5) is fulfilled, the conditional expectation $E(X_1|X_{-\infty}^0)$ is continuous and the conditional expectation $E(X_1|X_{-\infty}^0)$ has non-degenerate conditional distribution given X_{-K}^0 almost surely for all K. By Lemma 2.4 we have $P(\tilde{\Omega}) = 1$.

Let us choose an $\tilde{\omega} \in \tilde{\Omega}$. By the almost sure continuity assumption for a fix $\varepsilon > 0$ there exists a $\delta > 0$ such that if $d^*(\tilde{\omega},\omega') < \delta$, $\omega' \in \tilde{\Omega}$ then

$$|E(X_1|X_{-\infty}^0)(\tilde{\omega}) - E(X_1|X_{-\infty}^0)(\omega')| < \varepsilon. \tag{6}$$

Identify $\tilde{\omega} = (\ldots,\tilde{x}_{-1},\tilde{x}_0)$ and $\omega' = (\ldots,x'_{-1},x'_0)$. Due to the definition of the metric d^* if $K = \lceil \log_2 \frac{1}{\delta} \rceil$ and $\tilde{x}_{-K} = x'_{-K},\ldots,\tilde{x}_0 = x'_0$, then $d^*(\tilde{\omega},\omega') < \delta/2$. Since $\tilde{\omega} \in \tilde{\Omega}$ by Lemma 2.4 we have that $E(X_1|X_{-\infty}^0)$ has a non-degenerate conditional distribution given $X_{-K}^0 = \tilde{x}_{-K}^0$. Thus there exists an $\omega' \in \tilde{\Omega}$ with $\omega' = (\ldots,x'_{-K-1},\tilde{x}_{-K},\ldots,\tilde{x}_0)$ such that

$$|E(X_1|X_{-\infty}^0)(\tilde{\omega}) - E(X_1|X_{-\infty}^0)(\omega')| > \varepsilon,$$

which contradicts (6). ∎

Remark 2.7: We note that for Gauss-Markov processes the conditional expectation $E(X_1|X_{-\infty}^0)$ is continuous.

Now we consider an extension of the algorithm discussed in Morvai and Weiss [13], [15].

Define the nested sequence of partitions $\{\mathcal{P}_k\}_{k=0}^{\infty}$ of the real line as follows. Let

$$\mathcal{P}_k = \{[i2^{-(k+1)^3},(i+1)2^{-(k+1)^3}) : \text{ for } i = 0,1,-1,\ldots\}.$$

The choice of $\{\mathcal{P}_k\}_{k=0}^{\infty}$ in such form has technical reasons, see (11) in the proof of Theorem 2.8. Let $x \to [x]^k$ denote a quantizer that assigns to any point $x \in \Re$ the unique interval in $\mathcal{P}_k$ that contains x. Let

$$[X_m^n]^k = ([X_m]^k,\ldots,[X_n]^k).$$

We define the stopping times $\{\lambda_n\}$ along which we will estimate. Set $\lambda_0 = 0$. For $n = 1,2,\ldots$, define λ_n recursively. Let λ_n be

$$\lambda_{n-1} + \min\{t > 0 : [X_t^{\lambda_{n-1}+t}]^{\lambda_{n-1}} = [X_0^{\lambda_{n-1}}]^{\lambda_{n-1}}\}. \tag{7}$$

Note that the quantizer used at step n depends on the data and $\lambda_n \geq n$. Let the nth estimate m_n be defined as

$$m_n = \frac{1}{n}\sum_{j=0}^{n-1} X_{\lambda_j+1}. \tag{8}$$

The main result of the paper is the following.

Theorem 2.8: Consider a stationary Gaussian process from the model class Φ. Then

$$\lim_{n\to\infty}\left|m_n - E(X_{\lambda_n+1}|X_0^{\lambda_n})\right| = 0 \text{ almost surely.} \tag{9}$$

Proof: Consider the following decomposition

$$\frac{1}{n}\sum_{j=0}^{n-1} X_{\lambda_j+1} - E(X_{\lambda_n+1}|X_0^{\lambda_n}) =$$
$$\frac{1}{n}\sum_{j=0}^{n-1}\Big(X_{\lambda_j+1} - E(X_{\lambda_j+1}|X_0^{\lambda_j})\Big)$$
$$+\frac{1}{n}\sum_{j=0}^{n-1} E(X_{\lambda_j+1}|X_0^{\lambda_j}) - E(X_{\lambda_n+1}|X_0^{\lambda_n}).$$

First we prove that

$$P(X_{\lambda_{n-1}+1} \in C|[X_0^{\lambda_{n-1}}]^{\lambda_{n-1}}) = P(X_{\lambda_n+1} \in C|[X_0^{\lambda_{n-1}}]^{\lambda_{n-1}}) \tag{10}$$

for any Borel set $C \subseteq \mathbf{R}$. Indeed,

$$P\big(X_{\lambda_n+1} \in C, [X_0^{\lambda_{n-1}}]^{\lambda_{n-1}} = A_0^{\lambda_{n-1}}, \lambda_{n-1} = l\big) =$$
$$\sum_{t\geq l} P\big(X_{\lambda_n+1} \in C, [X_0^{\lambda_{n-1}}]^{\lambda_{n-1}} = A_0^{\lambda_{n-1}}, B_n(t,l)\big) =$$
$$\sum_{t\geq l} P\big(T^{t-l}\{X_{\lambda_n+1} \in C, [X_0^{\lambda_{n-1}}]^{\lambda_{n-1}} = A_0^{\lambda_{n-1}}, B_n(t,l)\big) =$$
$$P\big(X_{\lambda_{n-1}+1} \in C, [X_0^{\lambda_{n-1}}]^{\lambda_{n-1}} = A_0^{\lambda_{n-1}}, \lambda_{n-1} = l\big)$$

where $B_n(t,l)$ denotes the event

$$\{\lambda_{n-1} = l, \lambda_n = t\},$$

T is the left shift operator, and we have used the stationarity property of the process $\{X_n\}$.

It follows that X_{λ_j+1} has the same distribution as X_1. Now observe that $X_{\lambda_j+1} - E(X_{\lambda_j+1}|X_0^{\lambda_j})$ is a sequence of orthogonal random variables with zero mean and the variance is less than or equal to $E\left(|X_{\lambda_j+1}|^2\right) = E\left(|X_1|^2\right)$. Now by Theorem 3.2.2 in Révész [26], the average of orthogonal random variables tends to zero almost surely.

What remains to prove is that

$$\frac{1}{n}\sum_{j=0}^{n-1} E(X_{\lambda_j+1}|X_0^{\lambda_j}) - E(X_{\lambda_n+1}|X_0^{\lambda_n})$$

tends to zero. Observe that $E(X_{\lambda_j+1}|[X_0^{\lambda_j}]^{\lambda_j})$ converges with probability 1 since it forms a martingale by (10). Thus to finish the proof it is enough to show that

$$E(X_{\lambda_n+1}|[X_0^{\lambda_n}]^{\lambda_n}) - E(X_{\lambda_n+1}|X_0^{\lambda_n}) \to 0 \quad \text{almost surely.}$$

Indeed, using the linearity of the autoregression function we have that

$$\lim_{n\to\infty} |E(X_n|[X_0^{n-1}]^{n-1}) - E(X_n|X_0^{n-1})| \le$$

$$\lim_{n\to\infty} \sup_{y_0^{n-1}\in[X_0^{n-1}]^{n-1}} |\sum_{i=0}^{n-1} f_{n-i}^n (X_i - y_i)| \le$$

$$\lim_{n\to\infty} \sup_{y_0^{n-1}\in[X_0^{n-1}]^{n-1}} \sum_{i=0}^{n-1} |f_{n-i}^n||X_i - y_i| \le$$

$$\lim_{n\to\infty} 2^{-n^3} \sum_{i=1}^{n} |f_i^n| =$$

$$\lim_{n\to\infty} 2^{-n^3} \sup_{\underline{e}^n\in\mathcal{E}_n} \sum_{i=1}^{n} f_i^n e_i^n,$$

where $\mathcal{E}_n = \{\underline{e}^n : \underline{e}^n = (e_1^n, \dots, e_n^n), e_i^n \in \{-1, +1\}\}$. If $\{X_n\}$ belongs to the class Φ_2, i.e. the process is Markov of some order k, then for $n, i > k$ $f_i^n = 0$, and therefore the limit

$$\lim_{n\to\infty} 2^{-n^3} \sup_{\underline{e}^n\in\mathcal{E}_n} \sum_{i=1}^{n} f_i^n e_i^n = 0.$$

We have to deal with the case when the process belongs to Φ_1. Since

$$\sum_{i=1}^{n} f_i^n e_i^n = \gamma_n \Gamma_n^{-1} \underline{e}^n,$$

(cf. (2)) we get

$$\lim_{n\to\infty} |E(X_n|[X_0^{n-1}]^{n-1}) - E(X_n|X_0^{n-1})| \le$$

$$\lim_{n\to\infty} \left(2^{-n^3} \sup_{\underline{e}^n\in\mathcal{E}_n} \gamma_n \Gamma_n^{-1} \underline{e}^n \right) \le$$

$$\lim_{n\to\infty} \left(2^{-n^3} \|\gamma_n\| \, \|\Gamma_n^{-1}\| \, \|\underline{e}^n\| \right). \tag{11}$$

Assumption (1) implies that $\gamma(k) \to 0$ as $k \to \infty$, thus

$$\|\gamma_n\| = O(\sqrt{n}). \tag{12}$$

Trivially

$$\|\underline{e}^n\| = \sqrt{n}. \tag{13}$$

To estimate $\|\Gamma_n^{-1}\|$ we should estimate the minimal eigenvalue of Γ_n. For this we use the results of Serra [29]. Since Γ_n is a Toeplitz matrix, see Doob [5] (page 476), by Theorem 3.1 of [29], we have that $ct^{n(n+1)/2}$ is an absolute lower bound for the minimal eigenvalue of Γ_n, where $c > 0$ and $0 < t < 1$ are constants. Thus we have that

$$\|\Gamma_n^{-1}\| \le O(t^{-n(n+1)/2}). \tag{14}$$

Putting together (12), (13), and (14) we get that

$$\limsup_{n\to\infty} \|\gamma_n\| \, \|\Gamma_n^{-1}\| \, \|\underline{e}^n\| \le O(nt^{-n(n+1)/2})$$

Combining this with (11) we have that

$$\lim_{n\to\infty} \left(2^{-n^3} \|\gamma_n\| \, \|\Gamma_n^{-1}\| \, \|\underline{e}^n\| \right) = 0.$$

■

References

[1] P. Algoet. Universal schemes for prediction, gambling and portfolio selection. *Annals of Probability*, 20:901–941, 1992.

[2] P. Algoet. Universal schemes for learning the best nonlinear predictor given the infinite past and side information. *IEEE Trans. Information Theory*, 45 no. 4:1165–1185, 1999.

[3] D.H. Bailey. Sequential Schemes for Classifying and Predicting Ergodic Processes. Technical report, Stanford University, 1976. PhD. Thesis.

[4] T. M. Cover. Open problems in information theory. *IEEE Joint Workshop on Information Theory "IEEE Press"*, pages 35–36, 1975.

[5] J. L. Doob *Stochastic Processes*. John Wiley & Sons, 1967.

[6] W. Feller *An Introduction to Probability Theory and Its Applications* John Wiley & Sons, 1971.

[7] L. Györfi and G. Lugosi. Strategies for sequential prediction of stationary time series. In M. Drop, P. L'Ecuyer, and F. Szidarovszky, editors, *Examination of Stochastic Theory, Methods and Applications*, pages 225–248. Kluwer Academic Publishers, 2002.

[8] L. Györfi, G. Lugosi, and G. Morvai. A simple randomized algorithm for sequential prediction of ergodic time series. *IEEE Trans. Information Theory*, 45 no. 7:2642–2650, 1999.

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

[10] T. Hida and M. Hitsuda. Gaussian Processes. Providence, RI:AMS 1993 Translation of *Mathematical Monographs*, vol 10., 1993.

[11] G. Morvai. Guessing the output of a stationary binary time series. In Y. Haitovsky, H. R. Lerche, and Y. Ritov, editors, *Foundations of Statistical Inference*, pages 205–213. Physika Verlag, 2003.

[12] G. Morvai and B. Weiss. Forecasting for stationary binary time series. *Acta Applicandae Mathematicae*, 79:25–34, 2003.

[13] G. Morvai and B. Weiss. Intermittent estimation of stationary time series. *Test*, 13:525–542, 2004.

[14] G. Morvai and B. Weiss. Forward estimation for ergodic time series. *Ann. I.H. Poincare*, PR 41:859–870, 2005.

[15] G. Morvai and B. Weiss. Inferring the conditional mean. *Theory of Stochastic Processes*, Vol. 11 (27) no. 1-2:112–120, 2005.

[16] G. Morvai and B. Weiss. Limitations on intermittent forecasting. *Statistics and Probability Letters*, 72:285–290, 2005.

[17] G. Morvai and B. Weiss. Prediction for discrete time series. *Probability Theory and Related Fields*, 132:1–12, 2005.

[18] G. Morvai and B. Weiss, "On estimating the memory for finitarily Markovian processes," *Ann. I.H.Poincaré-PR*, vol. 43, pp. 15-30, 2007.

[19] G. Morvai, B. Weiss, "On sequential estimation and prediction for discrete time series " *Stochastics and Dynamics*, vol. 7, no. 4, pp. 417-437, 2007.

[20] G. Morvai and B. Weiss, "Estimating the Lengths of Memory Words". *IEEE Transactions on Information Theory*, vol.54 no. 8 pp. 3804-3807, 2008.

[21] G. Morvai and B. Weiss, "On universal estimates for binary renewal processes," *The Annals of Applied Probability* vol. 18, no. 5, pp. 1970-1992, 2008.

[22] G. Morvai and B. Weiss, "Estimating the Residual Waiting Time for Binary Stationary Time Series," In *Proc. ITW2009, Volos, Greece, June 10-12* , pages 67–70, 2009.

[23] G. Morvai, S. Yakowitz, and P. Algoet. Weakly convergent nonparametric forecasting of stationary time series. *IEEE Trans. Inform. Theory*, 43 no. 2:483–498, 1997.

[24] G. Morvai, S. Yakowitz, and L. Györfi. Nonparametric inferences for ergodic, stationary time series. *Annals of Statistics*, 24 no. 1:370–379, 1996.

[25] D. S. Ornstein. Guessing the next output of a stationary process. *Israel J. Math*, 30:292–296, 1978.

[26] P. Révész. *The Law of Large Numbers*. Academic Press, 1968.

[27] B. Ya. Ryabko. Prediction of random sequences and universal coding. *Problems of Inform. Trans.* 24 no. 2:87–96, 1988. *(Problemy Peredachi Informatsii*, 24 no. 2:3–14, 1988.)

[28] D. Schäfer. Strongly Consistent Online Forecasting of Centered Gaussian Processes. *IEEE Trans. Information Theory*, Vol. 48 no. 3:791–799, 2002.

[29] S. Serra. On the extreme eigenvalues of Hermitian (block) Toeplitz matrices. *Linear Alg. Appl.*, 270:109–129, 1998.