Intermittent Estimation for Gaussian Processes
Appeared in:

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 \( E(X_{n+1}|X_0,\ldots,X_n) \) from a growing number of observations \( X_0,\ldots,X_n \) 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 \( E(X_{n+1}|X_0,\ldots,X_n) \) from the observations \( X_0,\ldots,X_n \) 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 estimation for all time instances \( n \) we estimate merely along a stopping time sequence. That is, looking at the data segment \( X_0,\ldots,X_n \) 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 auto-covariance function \( \gamma(k) = E(X_{n+k}X_n) \) and \( EX_n = m \). Define the following subclasses of stationary Gaussian processes: In \( \Phi_1 \) we have Gaussian processes satisfying the condition

\[
\sum_{j=0}^{\infty} |\gamma(j)| < \infty
\]

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

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 \( \epsilon_i \)'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 \( \Phi \), see Hida and Hitsuda [10].

Let

\[
\Gamma_n = (\gamma(|i-j|))_{i,j=1,\ldots,n} =
\begin{pmatrix}
\gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\
\gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0)
\end{pmatrix}
\]

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 Egy 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 \( \gamma_n = (\gamma(n), \ldots, \gamma(1)) \).

For Gaussian processes with nonsingular \( \Gamma_n \) (every process in \( \Phi_1 \) has this property) we have
\[
E(X_n | X_0, \ldots, X_{n-1}) = \gamma_n \Gamma_n^{-1}(X_0, \ldots, X_{n-1})^T + m = f_n^0(X_0 - m) + \ldots + f_n^i(X_{n-1} - m) + m. \tag{2}
\]
For Gaussian processes in \( \Phi \) with singular \( \Gamma_n \), which are all in fact in \( \Phi_2 \) let
\[
E(X_n | X_0, \ldots, X_{n-1}) = f_n^0(X_0 - m) + \ldots + f_n^i(X_{n-1} - m) + m, \tag{3}
\]
where for \( i > k \) (\( k \) is the order of the Markov process) \( f_n^i = 0 \). 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

\[
\frac{1}{\psi(z)} = \sum_{k=0}^{\infty} \varphi_k z^k \quad (|z| > 1)
\]
satisfy
\[
\sum_{k=d_{n+1}}^{\infty} |\varphi_k|^2 \leq \left( \frac{C}{\log n} \right)^r \tag{4}
\]
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 \( E(X_{n+1} | X_n^0) \) 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 \( \mathbb{R} \) be the set of all real numbers and put \( \mathbb{R}^* \) the set of all one-sided sequences of real numbers, that is,
\[
\mathbb{R}^* = \{ (\ldots, x_{-i}, x_0) : x_i \in \mathbb{R} \text{ for all } -\infty < i \leq 0 \}.
\]

Define the metric \( d^*(\cdot, \cdot) \) on \( \mathbb{R}^* \) as
\[
d^*((\ldots, x_{-1}, x_0), (\ldots, y_{-1}, y_0)) = \sum_{i=0}^{\infty} 2^{-i-1} |x_{-i} - y_{-i}| / |x_{-i} - y_{-i}|.
\]

**Definition 2.3:** The conditional expectation \( E(X_1 | X_0, \ldots, X_{-1}, X_0) \) is said to be almost surely continuous if for some set \( B \subseteq \mathbb{R}^* \) which has probability one the conditional expectation \( E(X_1 | X_0, \ldots, X_{-1}, X_0) \) 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 \( E(X_1 | X_0, \ldots, X_{-1}, X_0) \) is almost surely continuous. Unfortunately the conditional expectation \( E(X_1 | X_0, \ldots, X_{-1}, X_0) \) 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 \( X_n^{(m)} = (X_m, \ldots, X_n) \), where \( m \leq n \).

**Lemma 2.4:** Consider a stationary Gaussian process \( \{X_n\} \) and assume that it is not Markov of any order. Then for any \( K \geq 0 \) the conditional expectation \( E(X_1 | X_0^\infty) \) has a non degenerate Gaussian conditional distribution given \( X_0^\infty \) almost surely.

**Proof:** Observe that \( E(X_1 | X_0^\infty) \) is a normally distributed random variable. Let us fix a \( K \geq 0 \) and assume that the conditional expectation \( E(X_1 | X_0^\infty) \) has degenerate conditional distribution given \( X_0^\infty \) on a set with positive probability. We prove that this assumption leads to a contradiction.

The above assumption is equivalent with \( P(D) > 0 \), where
\[
D = \{ x_0^\infty : E(X_1 | X_0^\infty = x_0^\infty) = C(x_{-K}, \ldots, x_0) \},
\]
and \( C(x_{-K}, \ldots, x_0) \) is a function depending merely on \( x_{-K}, \ldots, x_0 \).

Consider an arbitrary \( k \geq K \). On one hand on \( D \)
\[
E(X_1 | X_{-k}^\infty) = E(E(X_1 | X_0^\infty) | X_0^\infty) = C(X_{-K}, \ldots, X_0),
\]
on the other hand
\[
E(X_1 | X_{-k}^\infty) = f_1^{k+1}(X_0 - m) + \ldots + f_{k+1}^{k+1}(X_{-k} - m) + m.
\]
Thus we have that for all \( k \geq K \) on \( D \)
\[
f_1^{k+1}(X_0 - m) + \ldots + f_{k+1}^{k+1}(X_{-k} - m) = f_{K+1}^{K+1}(X_0 - m) + \ldots + f_{K+1}^{K+1}(X_{-K} - m).
\]
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 \( X_0, \ldots, X_k \) is a non-degenerate \( k + 1 \) dimensional normal distribution.) This implies that the coefficients are equal. Hence we get that
\[
f_i^{K+1} = f_i^{k+1} \text{ if } i \leq K
\]
and \( f_i^k = 0 \) if \( i > K \) for all \( k \geq K \). Thus we have
\[
E(X_1 | X_{-k}^\infty) = f_1^{K+1}(X_0 - m) + \ldots + f_{K+1}^{K+1}(X_{-K} - m) + m
\]
for all \( k \geq K \). Considering the limit of the left hand side we get
\[
\lim_{k \to \infty} E(X_1 | X_{-k}^\infty) = E(X_1 | X_0^\infty) = f_1^{K+1}(X_0 - m) + \ldots + f_{K+1}^{K+1}(X_{-K} - m) + m.
\]
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 $\mathbb{P}_2$. 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 $E(X_1|X_0^n)$ 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 $E(X_1|X_0^n)$ is almost surely continuous. The convergence

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

is fulfilled almost surely. Let $\tilde{\Omega} \subset \Omega$ be such that on $\tilde{\Omega}$ the limit (5) is fulfilled, the conditional expectation $E(X_1|X_0^n)$ is continuous and the conditional expectation $E(X_1|X_0^n)$ has non-degenerate conditional distribution given $X_0^n$ 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 $d^*(\tilde{\omega}, \omega^{'}) < \delta$, $\omega^{' \prime} \in \tilde{\Omega}$ then

$$|E(X_1|X_0^n)(\tilde{\omega}) - E(X_1|X_0^n)(\omega^{'})| < \varepsilon. \tag{6}$$

Identify $\tilde{\omega} = (\ldots, \tilde{x}_{-1}, \tilde{x}_0)$ and $\omega^{' \prime} = (\ldots, x_{-1}', x_0')$. Due to the definition of the metric $d^*$ if $K = [\log_2 \frac{1}{\delta}]$ and $\tilde{x}_K = x_{-K}', \ldots, \tilde{x}_0 = x_0'$, then $d^*(\tilde{\omega}, \omega^{' \prime}) < \delta/2$. Since $\tilde{\omega} \in \tilde{\Omega}$ by Lemma 2.4 we have $E(X_1|X_0^n)$ has a non-degenerate conditional distribution given $X_0^n = \tilde{x}_{-K}$. Thus there exists an $\omega^{' \prime} \in \tilde{\Omega}$ with $\omega^{' \prime} = (\ldots, \tilde{x}_{-K}, \tilde{x}_{-K + 1}, \ldots, \tilde{x}_0)$ such that

$$|E(X_1|X_0^n)(\tilde{\omega}) - E(X_1|X_0^n)(\omega^{' \prime})| > \varepsilon,$n

which contradicts (6).

Remark 2.7: We note that for Gauss-Markov processes the conditional expectation $E(X_1|X_0^n)$ is continuous.

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

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

$$P_k = \{i2^{-(k+1)} \leq x < (i+1)2^{-(k+1)} : \text{for } i = 0, 1, -1, \ldots\}.$$

The choice of $\{P_k\}_{k=0}^\infty$ 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 $x \in \mathbb{R}$ the unique interval in $P_k$ that contains $x$. Let

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

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

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

Note that the quantizer used at step $n$ depends on the data and $\lambda_n \geq n$. Let the $n$th 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 $\Phi$. Then

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

Proof: Consider the following decomposition

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

First we prove that

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

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

$$P(X_{\lambda_n+1} \in C, X_0^n) = \lambda_n = 1 = \sum_{l \geq 2} \sum_{t \geq l} P(X_{\lambda_n+1} \in C, X_0^n = A_0^{\lambda_n+1}, B_n(t, l)) = \sum_{l \geq 2} \sum_{t \geq l} P(T^t \in C, X_0^n = A_0^{\lambda_n+1}, B_n(t, l)) = P(X_{\lambda_n+1} \in C, X_0^n = A_0^{\lambda_n+1}, \lambda_n-1 = 1)$$

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 $X_{\lambda_n+1}$ has the same distribution as $X_1$. Now observe that $X_{\lambda_n+1} - E(X_{\lambda_n+1}|X_0^n)$ is a sequence of orthogonal random variables with zero mean and the variance is less than or equal to $E((X_{\lambda_n+1})^2) = E((X_1)^2)$. 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

$$\frac{1}{n} \sum_{j=0}^{n-1} \sum_{l \geq 2} \sum_{t \geq l} \sum_{k \geq 0} \sum_{j \geq k} P(X_{\lambda_j+1}^k|X_0^n) = E(X_{\lambda_n+1}|X_0^n)$$

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

$$E(X_{\lambda_n+1}|X_0^n) - E(X_{\lambda_n+1}|X_0^n) \to 0 \text{ almost surely.}$$
Indeed, using the linearity of the autoregression function we have that
\[
\lim_{n \to \infty} |E(X_n | [X_0^{n-1}]^{n-1}) - E(X_n | X_0^{n-1})| \leq \\
\lim_{n \to \infty} \sup_{y_0 \in [X_0^{n-1}]} \left| \sum_{i=0}^{n-1} f_i^n (X_i - y_i) \right| \leq \\
\lim_{n \to \infty} \sup_{y_0 \in [X_0^{n-1}]} \left| \sum_{i=0}^{n-1} f_i^n (X_i - y_i) \right| \leq \\
\lim_{n \to \infty} 2^{-n^3} \sum_{i=1}^{n} |f_i^n| = \\
\lim_{n \to \infty} 2^{-n^3} \sup_{E_n \in \mathcal{E}_n} \sum_{i=1}^{n} f_i^n e_i^n,
\]
where \( \mathcal{E}_n = \{ e^n : e^n = (e_1^n, \ldots, e_n^n), e_i^n \in \{-1, +1\} \} \). If \( \{X_n\} \) belongs to the class \( \Phi_2 \), i.e. the process is Markov of some order \( k \), then for \( n, i > k \) \( f_i^n = 0 \), and therefore the limit
\[
\lim_{n \to \infty} 2^{-n^3} \sup_{E_n \in \mathcal{E}_n} \sum_{i=1}^{n} f_i^n e_i^n = 0.
\]

We have to deal with the case when the process belongs to \( \Phi_1 \). Since
\[
\sum_{i=1}^{n} f_i^n e_i^n = \gamma_n \Gamma_n^{-1} e^n,
\]
(cf. (2)) we get
\[
\lim_{n \to \infty} |E(X_n | [X_0^{n-1}]^{n-1}) - E(X_n | X_0^{n-1})| \leq \\
\lim_{n \to \infty} \left( 2^{-n^3} \sup_{E_n \in \mathcal{E}_n} \gamma_n \Gamma_n^{-1} e^n \right) \leq \\
\lim_{n \to \infty} \left( 2^{-n^3} \| \gamma_n \| \| \Gamma_n^{-1} \| \| e^n \| \right).
\]

Assumption (1) implies that \( \gamma(k) \to 0 \) as \( k \to \infty \), thus
\[
\| \gamma_n \| = O(\sqrt{n}).
\]

Trivially
\[
\| e^n \| = \sqrt{n}.
\]

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

Putting together (12), (13), and (14) we get that
\[
\lim_{n \to \infty} \sup \| \gamma_n \| \| \Gamma_n^{-1} \| \| e^n \| \leq O(n t^{-n(n+1)/2})
\]

Combining this with (11) we have that
\[
\lim_{n \to \infty} \left( 2^{-n^3} \| \gamma_n \| \| \Gamma_n^{-1} \| \| e^n \| \right) = 0.
\]