Partition Optimization for a Random Process Realization to Estimate its Expected Value

Vladimir Marchuk¹, Igor Shrafel¹, Dmitry Chernyshov¹, Alexander Minaev¹, Stepan Buryakov¹

Abstract: The paper provides an analytical proof the optimal number of partitions of a non-stationary random process realization, which is necessary for estimating its expected value when using “the estimation reproduction” method for signal processing. This method allows to process signal with a limited volume of priori information about the desired signal function and statistical characteristics of the additive noise component.

Keywords: Desired signal, Noise, Signal-to-noise ratio, Expected value, Estimation, Random process.

1 Introduction

Processing of non-stationary random process requires efficiency preservation in conditions of limited amount of the priori information, both on the function of the useful signal and on the statistical characteristics of the additive noise component. The estimation reproduction (ER) method, was described in [1 – 4], it shows good results in such conditions of a priori indeterminacy. In works [5 – 11] this method was fully described, but we did not consider the question of optimal partition number of signal realization. In this paper the main attention will be payed to optimization of partitions number for a non-stationary random process in terms of estimating expected value.

2 The Mathematical Model of Method

Let \( x(t) = a(t) + \eta(t) \) be the sum of the deterministic signal \( a(t) \) and its noise component \( \eta(t) \) at instant \( t \). Suppose \( \eta(t) \) is centered random process with pairwise uncorrelated cross sections and a constant variance \( D_\eta(t) = D \). Consider original signal \( x(t) \) in equally distant moments of time \( t_k = t_i + (k-1)h \) and put \( x_k = x(t_k) \), \( a_k = a(t_k) \), \( k = 1,...,n \). Subsequently \( M_{x_k} = a_k \), \( D_{x_k} = D \)

¹Institute in Shakhtry of DSTU, Shevchenko str. 147, Shakhtry, Rostov region, 346500, Russia;
E-mails: marchuk@sssu.ru; shrafel17@mail.ru; dimchert@inbox.ru; ayminaev@gmail.com; buryakov@gmail.com
and random values $x_1, \ldots, x_n$ are uncorrelated pairwise. Hereinafter, the same notation will be used for a random process and its implementation.

Let assume that some realization $X = (x_1, \ldots, x_n)$ was received. We choose random integer numbers $1 \leq p \leq n$ and $(p - 1)$ that belong to elementary subset $P$ of $\{1, \ldots, n - 1\}$ subset. Let $n_1, n_2, \ldots, n_{p-1}$ be elements of $P$ (at $p \geq 2 \iff P \neq \emptyset$), that are arranged in ascending order $n_0 = 0, n_p = n$; Set $P$ generates a partition of the set $\{1, \ldots, n\}$ into $p$ pieces $T_i = \{n_{i-1} + 1, \ldots, n_i\}$, $1 \leq i \leq p$. Hereinafter, we shall call the set $P$ “a partition”.

Let us find OLS estimations with polynomials of degree zero $a_k$ for fragments $(x_k)_{k \in T_i}$ of realization $X$ at every $1 \leq i \leq p$ and with $k \in T_i$:

$$y_k = \frac{1}{n_i - n_{i-1}} \sum_{j=n_{i-1}+1}^{n_i} x_j, \quad k = n_{i-1} + 1, \ldots, n_i. \quad (1)$$

This method of estimating the expected value of a random process is used in the estimation reproduction method [1 – 4]. As it is known the criterion of the quality of estimation $Y = (y_k)_{k=1}^n$ is mean variance $D_m = \frac{1}{n} \sum_{k=1}^{n} M[ (y_k - a_k)^2 ]$.

Sometimes, to emphasize the dependence of the mean -variance on the partition $P$, we shall denote it by $D_m(P)$. Let us call partition $P^*$ its corresponding mean variance $D_m^* = D_m(P^*)$ optimal for any partition $P - D_m^* \leq D_m(P)$.

In this context, we derive the formula for the mean variance:

$$nD_m = \sum_{i=1}^{p} \sum_{j=n_{i-1}+1}^{n_i} M[ (y_j - a_j)^2 ] = \sum_{i=1}^{p} \sum_{j=n_{i-1}+1}^{n_i} M[ y_j^2 - 2y_j a_j + a_j^2 ] =$$

$$= \sum_{i=1}^{p} \sum_{j=n_{i-1}+1}^{n_i} \left[ M(y_j^2) - (My_j)^2 + (My_j)^2 - 2a_j My_j + a_j^2 \right] =$$

$$= \sum_{i=1}^{n} \left[ \sum_{j=n_{i-1}+1}^{n_i} Dy_j^2 + \sum_{j=n_{i-1}+1}^{n_i} (My_j - a_j)^2 \right]. \quad (2)$$

Note that due to the pairwise uncorrelated random variables $x_i$, $Dy_j = \frac{1}{(n_j - n_{j-1})^2} \sum_{k=n_{j-1}+1}^{n_j} Dx_k = \frac{D}{n_j - n_{j-1}}$, $n_{i-1} + 1 \leq j \leq n_i$, $1 \leq i \leq p$. Moreover, at the same $i, j$, $My_j = \bar{a}_i$, where $\frac{1}{n_j - n_{j-1}} \sum_{k=n_{j-1}+1}^{n_j} a_k$ is denoted as $\bar{a}_i$. Therefore
Partition Optimization for a Random Process Realization to Estimate its Expected Value

\[ nD_m = \sum_{i=1}^{p} \left[ \frac{D(n_i - n_{i-1})}{n_i - n_{i-1}} + \sum_{j=n_{i-1}+1}^{n_i} \left( a_j - \bar{a} \right)^2 \right] = \sum_{i=1}^{p} \left[ D + \sum_{j=n_{i-1}+1}^{n_i} a_j^2 - (n_i - n_{i-1})\bar{a}^2 \right] , \]

then

\[ nD_m = pD + \sum_{k=1}^{n} a_k^2 - \sum_{i=1}^{p} (n_i - n_{i-1})\bar{a}^2 . \]

In this paper we find optimal number of partitions and mean variance for the linear mathematical expectation \( a_k = \alpha + \beta t_k \). We obtain

\[ a_k = \alpha + \beta (t_i + (k - 1)h) = a + bk, \quad k = 1, \ldots, n , \]

where \( a = \alpha + \beta (t_i - h) \), \( b = \beta h \); \( \alpha, \beta \), that means that \( a, b \) are constant. Then

\[ \bar{a}_i = \frac{1}{n_i - n_{i-1}} \sum_{j=n_{i-1}+1}^{n_i} (a + bj) = \frac{1}{n_i - n_{i-1}} \left( a \sum_{j=n_{i-1}+1}^{n_i} 1 + b \sum_{j=n_{i-1}+1}^{n_i} j \right) = \frac{1}{n_i - n_{i-1}} \left( a(n_i - n_{i-1}) + \frac{b}{2}(n_{i-1} + 1 + n_i)(n_i - n_{i-1}) \right) =
\]

\[ = a + \frac{b}{2}(n_i + n_{i-1} + 1), \quad i = 1, \ldots, p. \]

We substitute this expression in the last sum of expression (4):

\[ \sum_{i=1}^{p} (n_i - n_{i-1})\bar{a}_i^2 = \sum_{i=1}^{p} (n_i - n_{i-1}) \left[ \left( a + \frac{b}{2} \right)^2 \left( \frac{2a + b}{2} \right) \left( n_i + n_{i-1} \right) + \frac{b^2}{4}(n_i + n_{i-1})^2 \right] = \]

\[ = \left( a + \frac{b}{2} \right)^2 \sum_{i=1}^{p} (n_i - n_{i-1}) + \frac{(2a + b)b}{2} \sum_{i=1}^{p} (n_i^2 - n_{i-1}^2) + \frac{b^2}{4} \sum_{i=1}^{p} (n_i^2 - n_{i-1}^2)(n_i + n_{i-1}) = \]

\[ = \left( a + \frac{b}{2} \right)^2 \left( n_p - n_0 \right) + \frac{(2a + b)b}{2} \left( n_p^2 - n_0^2 \right) + \frac{b^2}{3} \sum_{i=1}^{p} (n_i^3 - n_{i-1}^3) - \frac{b^2}{12} \sum_{i=1}^{p} (n_i - n_{i-1})^3 = \]

\[ = \left( a + \frac{b}{2} \right)^2 n + \frac{(2a + b)b}{2} n^2 + \frac{b^2}{3} n^3 - \frac{b^2}{12} \sum_{i=1}^{p} (n_i - n_{i-1})^3 . \]

Now we can find:

\[ \sum_{k=1}^{n} a_k^2 = \sum_{k=1}^{n} (a + bk)^2 = \sum_{k=1}^{n} a^2 + 2ab \sum_{k=1}^{n} k + b^2 \sum_{k=1}^{n} k^2 = \]

\[ = a^2 n + ab(n+1)n + b^2 \frac{n(n+1)(2n+1)}{6} . \]
Substitute this expression in (4):

\[ nD_m = pD + n \left( a^2 + ab(n+1) + \frac{b^2}{6} (n+1)(2n+1) - \left( a + \frac{b}{2} \right)^2 - \frac{(2a+b)b}{2} n - \frac{b^2}{3} n^2 \right) + \]

\[ + \frac{b^2}{12} \sum_{i=1}^{p} (n_i - n_{i-1})^3 = pD + \frac{b^2}{12} \left( \sum_{i=1}^{p} m_i - n \right), \]

(8)

where \( m_i = n_i - n_{i-1} \).

From the obtained expression for the mean variance, in particular, it follows that any permutation of \( \left( m'_i \right)_{i=1}^{p} \) elements of \( \left( m_i \right)_{i=1}^{p} \) sequence, that corresponds to the optimal partition, it also generates an optimal partition

\[ P' = \left\{ n'_i = \sum_{j=1}^{i} m'_j \right\}_{i=1}^{p-1}. \]

Let us consider the case \( b = 0 \) separately. In this case the smallest value \( D_m = \frac{pD}{n} \) is reached when \( p = 1 \), and equals \( D^* = \frac{D}{n} \). Now let us set \( b \neq 0 \).

Let us show that if \( P \) is an optimal partition, then for all are \( 1 \leq i, j \leq p, |m_i - m_j| \leq 1 \). Assume that this is not so. That means that \( m_i - m_j \geq 2 \) for some pair of numbers \( 1 \leq i, j \leq p \). For definiteness, we assume that \( j > i \) (the case \( j < i \) is used in a similar way). Set \( N_k = n_k + 1 \) at \( i \leq k \leq j - 1 \), and \( N_k = n_k \) in other cases \( 0 \leq k \leq p \); \( Q = \{N_k \}_{k=1}^{p-1} \), \( M_k = N_k - N_{k-1}, k = 1, ..., p \). Then \( Q \) is partition \( M_k = m_k \) for all \( 1 \leq k \leq p, k \neq i, k \neq j, M_i = m_i + 1, M_j = m_j + 1 \).

Then:

\[ \frac{12n}{b^2} (D_m(Q) - D_m(P)) = \sum_{k=1}^{p} M_k^3 + \sum_{k=1}^{p} m_k^3 = \]

\[ = M_j^3 - m_i^3 + M_j^3 - m_i^3 = (m_i + 1)^3 - m_i^3 + \]

\[ + (m_j - 1)^3 - m_j^3 = 3(m_i^2 - m_j^2) + 3(m_i - m_j) = \]

\[ = 3(m_i - m_j)(1 - (m_i - m_j)) \leq 3(m_i - m_j)(1 - 2) < 0. \]

This contradicts the optimality of the partition \( P \). Hence, for an optimal partition \( P = P^* \) in a sequence of numbers \( m_1, ..., m_p \) there are, at most, two different values. If we denote smaller number as \( m \), than bigger one will be equal to \( m + 1 \). If number appears in sequence \( \ell \) times, then \( \ell_m + (p - \ell)(m + 1) = n \), it leads to \( \ell = (m + 1)p - n \); \( n/p = m + (p - \ell)/p \). Since
Partition Optimization for a Random Process Realization to Estimate its Expected Value

0 ≤ p - ℓ ≤ p - ℓ, then \( m = \left\lceil \frac{n}{p} \right\rceil \), \( 1 - \ell = \left\{ \frac{n}{p} \right\} \Rightarrow \ell = p\left(1 - \left\{ \frac{n}{p} \right\} \right) \) (here and after, the symbols \([t]\) and \(\{t\}\) denote, respectively, the integer and fractional parts of the number \(t\)). With \( A = 3D/b^2 \), we get:

\[
\frac{12n}{b^2} D^*_m = \ell m^3 + (m + 1)^3(p - \ell) - n + 4Ap = \\
= (m^3 - (m + 1)^3)\ell + (m + 1)^3p + 4Ap - n = \\
= (m + 1)^3p - (3m^2 + 3m + 1)((m + 1)p - n) + \\
+4Ap - n = (m + 1)p(m^2 + 2m + 1 - 3m^2 - 3m - 1) + \\
+n(3m^2 + 3m) + 4Ap = \\
= 3nm(m + 1) - p(2m^2 + 3m^2 + m - 4A).
\]  

(10)

We emphasize that this expression is equal to \(\frac{12n}{b^2} D^*_m\) only at values 1 ≤ p ≤ n, that minimize it. Let’s find all \(p\). We will call it optimal values and denote as \(p^*\).

Let \(f(p)\) be the function with continuous argument \(p \in \left(\frac{n}{n + 1}, n\right]\), that is set by expression (2), where \(m = \left\lceil \frac{n}{p} \right\rceil\).

We show next that function \(f(p)\) is piecewise linear and continuous on the indicated interval. We have \(\left(\frac{n}{n + 1}, n\right] = \bigcup_{k=1}^{n} \left(\frac{n}{k + 1}, \frac{n}{k}\right]\). At all \(1 \leq k \leq n\), \(\frac{n}{k + 1} \leq p \leq \frac{n}{k}\) the conditions of inequality \(k \leq \frac{n}{p} \leq k + 1\) are fulfilled, whence \(m = k\); \(f(p) = 3nk(k + 1) - p(2k^3 + 3k^2 + k - 4A)\), i.e. function \(f(p)\) is linear on \(\left(\frac{n}{k + 1}, \frac{n}{k}\right]\). To prove its continuity on \(\left(\frac{n}{k + 1}, \frac{n}{k}\right]\) it is sufficient to establish continuity from the right \(f(p)\) at the points \(p = \frac{n}{k + 1}, k = 1, ..., n - 1\). For the indicated \(k\), we have:
which coincides with

\[
\lim_{p \to n+1} f(p) = 3nk(k+1) - \frac{n}{k+1}(2k^3 + 3k^2 + k - 4A) = \\
= n \left( 3k^2 + 3k - 2k^2 - k + \frac{4A}{k+1} \right) = n \left( k^2 + 2k + \frac{4A}{k+1} \right),
\]

Thus, the function \( f(p) \) is continuous on \( \left( \frac{n}{n+1}, n \right) \).

Let us consider the function \( g(\lambda) = 2\lambda^3 + 3\lambda^2 + \lambda - 4A \). It is continuous and increases on \([0, +\infty)\). Moreover:

\[
g(1) = 6 - 4A, \quad (13)
g(n-1) = 2(n-1)^3 + 3(n-1)^2 + n - 1 - 4A = \\
= 2n^3 - 3n^2 + n - 4A. \quad (14)
\]

With

\[ A < \frac{3}{2} - g(k) \geq g(1) > 0, \quad k = 1, \ldots, n - 1 \Rightarrow f(p) \]

decreases on \([1, n] \Rightarrow \min_{1 \leq p \leq n} f(p) = f(n) \), i.e.

\[
p^* = n \Rightarrow m^* = \left[ \frac{n}{p^*} \right] = 1,
\]

\[
D_m^* = \frac{b^2}{12n} f(n) = \frac{b^2}{12n} (3n \cdot 2 - n(2 + 3 + 1 - 4A)) = \frac{b^2 A}{3} = D.
\]

Now let us consider a case where \( A > \frac{1}{4}(2n^3 - 3n^2 + n) \). We have \( g(k) \leq g(n-1) < 0, \quad k = 1, \ldots, n - 1 \) and function \( f(p) \) increases on

\[
[1, n] \Rightarrow \min_{1 \leq p \leq n} f(p) = f(1) \Rightarrow p^* = 1, m^* = \left[ \frac{n}{p^*} \right] = n.
\]
Partition Optimization for a Random Process Realization to Estimate its Expected Value

\[ D_m^* = \frac{b^2}{12n} f(1) = \frac{b^2}{12n} (3n^2(n+1) - 2n^3 - 3n^2 - n + 4A) = \frac{b^2A}{12} (n^2 - 1) + \frac{D}{n}. \]

Finally with \( \frac{3}{2} \leq A \leq \frac{1}{4} (2n^3 - 3n^2 + n) \) function \( g(\lambda) \) on \([1, n - 1]\) has one and only one root \( \lambda_0 \), that is found by means of Kardano formula:

\[ \lambda_0 = \sqrt[3]{A + \sqrt{A^2 - \frac{1}{1728}}} + \sqrt[3]{A - \sqrt{A^2 - \frac{1}{1728}}}. \] (15)

Put \( k_0 = [\lambda_0] \). At \( \lambda \geq k_0 + 1 \) \( g(\lambda) > 0 \), and at \( 0 < \lambda < k_0 \) \( g(\lambda) < 0 \). Moreover, if \( k_0 = \lambda_0 \) (i.e. \( \lambda_0 \) is integer) \( g(k_0) = 0 \), and in case \( k_0 \neq \lambda_0 \) \( g(k_0) < 0 \). From the research results of the function \( f(p) \), the conclusion follows that it decreases on \( \left[1, \frac{n}{k_0 + 1}\right] \) and increases on \( \left[\frac{n}{k_0 + 1}, n\right] \) (when \( \lambda_0 \) is integer it increases on \( \left[1, \frac{n}{k_0 + 1}\right] \), but remains constant on \( \left[\frac{n}{k_0 + 1}, \frac{n}{k_0}\right] \)).

Consider the case when \( \lambda_0 \) is non-integer. Then \( k_0 \leq n - 2 \) and function \( f(p) \) reaches its lowest value on \([1, n]\) in a single point \( p = p_0 = \frac{n}{k_0 + 1} \in \left[\frac{n}{n - 1 + \frac{1}{2}}\right] \). Now two cases are possible. If \( p_0 \) is integer and \( p^* = p_0 \),

\[ D_m = \frac{b^2}{12n} f\left(\frac{n}{k_0 + 1}\right) = \frac{b^2}{12n} \left(\frac{n}{k_0 + 1} + 2k_0 + \frac{4A}{k_0 + 1}\right) = \]
\[ = \frac{b^2}{12} \left(k_0^2 + 2k_0\right) + \frac{D}{k_0 + 1}. \] (16)

If \( p_0 \) is non-integer, then for finding \( p^* \) it is necessary to compare \( f\left([p_0]\right) \) and \( f\left([p_0]+1\right) \).

If \( f\left([p_0]\right) < f\left([p_0]+1\right) \) then \( p^* = p_0 \), if \( f\left([p_0]\right) > f\left([p_0]+1\right) \) then \( p^* = [p_0]+1 \); and if \( f\left([p_0]\right) = f\left([p_0]+1\right) \), the both values \([p_0]\) and \([p_0]+1\) are optimal.
If $\lambda_0$ is integer then at $\left[ \frac{n}{n_0} \right] + \left[ -\frac{n}{k_0 + 1} \right] \geq 0$ optimal values will be all integers $p \in \left[ -\left[ -\frac{n}{k_0 + 1} \right], \left[ \frac{n}{k_0} \right] \right]$. If then we choose $p^*$ as one of the numbers $p_1 = \left[ \frac{n}{k_0 + 1} \right]$, $p_2 = \left[ \frac{n}{k_0} \right] + 1$ for which the value $f(p_i)$ is less $(1 \leq k \leq 2)$. In case $f(p_1) = f(p_2)$ both values $p_1$ and $p_2$ are optimal.

Let us summarize the results.

**Theorem.** Let $x_1, \ldots, x_n$ be pairwise uncorrelated random values with expected value $M{x_i} = a + b \cdot k$ and variance $D{x_i} = D$, where $k = 1, \ldots, n, D > 0$, and $a,b$ are constant. We introduce two more constants: $A = \frac{3D}{b^2}$ (if we consider $A = +\infty$) and $\lambda_0 = \sqrt{A + \sqrt{A^2 - \frac{1}{1728}}} + \sqrt{A - \sqrt{A^2 - \frac{1}{1728}}}$ (if $A > \frac{1}{24\sqrt{3}}$). Also we introduce function:

$$f(p) = 3n\left[ \frac{n}{p} \right]\left( \left[ \frac{n}{p} \right] + 1 \right) - p\left( 2\left[ \frac{n}{p} \right]^3 + 3\left[ \frac{n}{p} \right]^2 + \left[ \frac{n}{p} \right] - 4A \right). \tag{17}$$

Let us define the set $S$ as set of integers as follows:

1. if $A < \frac{3}{2}$ put $S = \{n\}$;
2. if $A > \frac{1}{4}(2n^3 - 3n^2 + n)$, then $S = \{1\}$;

3. In case $\frac{3}{2} \leq A \leq \frac{1}{4}(2n^3 - 3n^2 + n)$ it creates branching;

3.1 if $\lambda_0$ -is integer, we denote $q_1 = -\left[ -\frac{n}{\lambda_0 + 1} \right]$, $q_2 = \left[ \frac{n}{\lambda_0} \right]$. Then:

3.1.1. if $q_2 \geq q_1$ then $S = \{q_1, q_1 + 1, \ldots, q_2\}$;

3.1.2. if $q_2 < q_1$ put $p_1 = \left[ \frac{n}{\lambda_0 + 1} \right]$, $p_2 = q_2 + 1$. 

340
Partition Optimization for a Random Process Realization to Estimate its Expected Value

3.1.2.1. if $f(p_i) < f(p_2)$ then $S = \{p_i\}$;
3.1.2.2. if $f(p_i) > f(p_2)$ then $S = \{p_2\}$;
3.1.2.3. if $f(p_i) = f(p_2)$, we assume $S = \{p_i, p_1\}$;

3.2. Let $\lambda_0$ be non-integer. Put $k_0 = \lfloor \lambda_0 \rfloor$, $p_0 = \frac{n}{k_0 + 1}$.

Then:

3.2.1. if $p_0$ is integer, then $S = \{p_0\}$;
3.2.2. if $p_0$ is non-integer, we denote $p' = \lfloor p_0 \rfloor$;
3.2.2.1. if $f(p') < f(p' + 1)$ we assume $S = \{p'\}$;
3.2.2.2. if $f(p') > f(p' + 1)$, then $S = \{p' + 1\}$;
3.2.2.3. finally if $f(p') = f(p' + 1)$ then $S = \{p', p' + 1\}$.

For partition $P = \{n_i\}_{i=1}^{p-1}$ to be optimal it is necessary and sufficient so that for $p \in S$ in sequence $(m_i = n_i - n_{i-1})_{i=1}^p$ a number $\left\lfloor \frac{n}{p} \right\rfloor$ was present $p \left(1 - \left\{ \frac{n}{p} \right\} \right)$ times. And if $\frac{n}{p}$ is non-integer the number $\left\lfloor \frac{n}{p} \right\rfloor + 1$ has to be present $p \left\{ \frac{n}{p} \right\}$ times. Optimal mean dispersion for all mentioned cases can be found with formula

$$D_m^* = \frac{b^2}{12n} f(p).$$

In subparagraphs 1, 2 and 3.2.1 $D_m^*$ is calculated using simplified equations

$$D_m^* = D, \quad D_m^* = \frac{b^2}{12n} (n^2 - 1) + \frac{D}{n} \quad \text{and} \quad D_m^* = \frac{b^2}{12} \left( k_0^2 + 2k_0 \right) + \frac{D}{k_0 + 1} \text{ respectively.}$$

3 Conclusion

In this paper, the optimal value of the partition and the value of the mean variance were found for the linear expected value. Mean variance characterizes the quality of the obtained estimation.

4 Acknowledgment

The reported study was supported by the Russian Foundation for Basic research (RFBR), research project № 15-07-99685.
5 References


