

N-Step Sliding Recursion Formula of Variance and Its Implementation

Lang Yu*, Gang He**, and Ahmad Khwaja Mutahir**

Abstract

The degree of dispersion of a random variable can be described by the variance, which reflects the distance of the random variable from its mean. However, the time complexity of the traditional variance calculation algorithm is $O(n)$, which results from full calculation of all samples. When the number of samples increases or on the occasion of high speed signal processing, algorithms with $O(n)$ time complexity will cost huge amount of time and that may results in performance degradation of the whole system. A novel multi-step recursive algorithm for variance calculation of the time-varying data series with $O(1)$ time complexity (constant time) is proposed in this paper. Numerical simulation and experiments of the algorithm is presented and the results demonstrate that the proposed multi-step recursive algorithm can effectively decrease computing time and hence significantly improve the variance calculation efficiency for time-varying data, which demonstrates the potential value for time-consumption data analysis or high speed signal processing.

Keywords

Algorithm Complexity, Multi-Step Recursion Algorithm, Sliding Window, Variance

1. Introduction

1.1 Overview

Variance is one widely used parameter for describing the characteristics of a signal. With the development of information technology, the speed and frequency of communication is increasing and the needing for high speed signal analysis and processing is keeping up. Besides, various kinds of large-scale and fine-grained data in different fields such as medical, transportation, education, energy, aerospace, etc., have been generated and collected, and needing processing efficiently.

Some common basic statistics need to be calculated during data analysis to describe some basic related statistical characteristics of the data. Among them, the mean and variance of data [1] are two commonly used basic statistics. The former reflects the information of the overall center position, and the later reflects the overall dispersion. Furthermore, there are other commonly used statistics, such as sample standard deviation, sample skewness, sample kurtosis, and so on, which are also widely used in different aspects of society.

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received March 8, 2019; first revision June 4, 2019; accepted June 8, 2019.

Corresponding Author: Gang He (ganghe@swust.edu.cn)

* School of Science, Southwest University of Science and Technology, Mianyang, China (swust2018yl@163.com)

** School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China (ganghe@swust.edu.cn, khwajamutahir311@gmail.com)

1.2 Literature Review

According to the brief review of different kinds of literature, the variance and the mean value of a signal is commonly analyzed in practical problems. For example, sample variance is identified in nonlinear systems [2], automatic detection [3,4], financial and corporate decision-making, investment [5-11], analysis of variance [12-15], image and pattern recognition [16-19], linguistic analysis [20-22] and other fields.

It can be seen that mean and the variance value plays an extremely important role in data analysis. However, in most cases, the variance is calculated by theoretical definition. Not only does this approach require access to the entire data sequence for each computation, but it also requires more storage resources. However, when calculating the variance of time-varying data with large data capacity and real-time updates of data, it not only wastes the storage resources of the computer but also fails to output the calculated results in time. Therefore, many improved algorithms for variance calculation have been proposed.

Chen and Gao [23] took the derivative of each parameter in the k-order origin moment of the usual discrete distribution, and obtained the unified recursive formula for the k-order origin moment of binomial distribution, Poisson distribution and geometric distribution through identical deformation. Chen [24] also used the derivative method to derive the recursive formula of the k-order origin moment of binomial distribution, Poisson distribution, and geometric distribution with differential form, and gave the recursive formula of the k-order origin moment of the general form of these three distributions. Similarly, Chen and Ma [25], with the help of the derivation strategy, proposed a new formula for calculating k-order central moment in general form, according to the recursive formula of the k-order central moment in the differential form of binomial distribution, Poisson distribution, and geometric distribution. By constructing a special class of polynomials and series, Chen [26] proposed a new unified formula for calculating the k-order central moment of binomial distribution, Poisson distribution, and geometric distribution. In the study of Fan [27], he introduced in detail the recursive method to calculate the sample mean, variance, and covariance after the change of sample size.

Based on filtering and smoothing, Liu and Sun [28] proposed an efficient method for calculating the variance of non-zero mean data. Deng et al. [29] first proposed a one-step recursion method of variance by studying the classical variance algorithm. The algorithm can efficiently calculate the variance of the data sequence whose observed data capacity changes rapidly. However, some special data sequences have a fixed length, and the observed values change in real-time, such as satellite timing system and real-time data analysis [30-33]. At this time, the classical definition algorithm and the improved algorithm proposed in the above literature also have some shortcomings in dealing with this problem.

For the above problem, the existing literature has not proposed an effective calculation method. Therefore, the main work of this paper is as follows. Firstly, the recursion method proposed in [29] is briefly reviewed. Then, a one-step recursive formula is proposed to dynamically calculate the variance of this special sequence. Secondly, based on one-step recursion formula, it is extended to a more general case. Namely, the one-step recursion formula is extended to multi-step recursion. Then, through numerical experiments, the one-step recursive formula and multi-step recursive formula are compared with the algorithm proposed in [29] for calculation error and calculation time. At last, this paper is summarized.

2. Variance Recursion Formula with the Variable Sequence Length

In this section, we prove the one-step recurrence formula proposed by Deng et al. [29] so as to introduce the variance recursion formula.

Definition 2.1. Let $F_n = (f_1, f_2, f_3, \dots, f_{n-1}, f_n)$ be a variable-length data sequence, \bar{F}_n is the mean of the sequence, and its formula is

$$\bar{F}_n = \frac{1}{n} \sum_{i=1}^n f_i. \tag{1}$$

Theorem 2.1. Let F_n and \bar{F}_n be as described in Definition 2.1, from which the recursive expression of \bar{F}_n and \bar{F}_{n-1} can be obtained as

$$\bar{F}_n = \bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n}. \tag{2}$$

Proof.

$$\begin{aligned} \bar{F}_n &= \frac{1}{n} \sum_{i=1}^n f_i = \frac{1}{n} \left(\sum_{i=1}^{n-1} f_i + f_n \right) = \frac{1}{n} \left[(n-1) \bar{F}_{n-1} + f_n \right] \\ \bar{F}_n &= \bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n}. \end{aligned} \tag{3}$$

Theorem 2.2. Let the variance definition formula be $V_n = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{F}_n)^2$, and then we can get the variance recursion formula with the variable sequence length

$$V_n = \frac{n-2}{n-1} V_{n-1} + \frac{(f_n - \bar{F}_{n-1})^2}{n}. \tag{4}$$

Proof. It can be obtained from the calculation formula of the definition of variance

$$V_{n-1} = \frac{1}{n-2} \sum_{i=1}^{n-1} (f_i - \bar{F}_{n-1})^2. \tag{5}$$

Formula (4) is expanded to obtain

$$\begin{aligned} V_n &= \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{F}_n)^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (f_i - \bar{F}_n)^2 + \frac{1}{n-1} (f_n - \bar{F}_n)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^{n-1} (f_i^2 + \bar{F}_n^2 - 2f_i \bar{F}_n) + \frac{1}{n-1} (f_n^2 + \bar{F}_n^2 - 2\bar{F}_n f_n). \end{aligned} \tag{6}$$

Combined with the mean recurrence formula (3), the above Eq. (6) can be written as

$$\begin{aligned}
V_n &= \frac{1}{n-1} \sum_{i=1}^{n-1} \left[f_i^2 + \left(\bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n} \right)^2 - 2f_i \left(\bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n} \right) \right] \\
&+ \frac{1}{n-1} \left[f_n^2 + \left(\bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n} \right)^2 - 2f_n \left(\bar{F}_{n-1} + \frac{f_n - \bar{F}_{n-1}}{n} \right) \right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n-1} \left[(f_i - \bar{F}_{n-1})^2 + \left(\frac{f_n - \bar{F}_{n-1}}{n} \right)^2 - 2(f_i - \bar{F}_{n-1}) \left(\frac{f_n - \bar{F}_{n-1}}{n} \right) \right] \\
&+ \frac{1}{n-1} \left[(f_n - \bar{F}_{n-1})^2 + \left(\frac{f_n - \bar{F}_{n-1}}{n} \right)^2 - 2(f_n - \bar{F}_{n-1}) \left(\frac{f_n - \bar{F}_{n-1}}{n} \right) \right] \tag{7} \\
&= \frac{1}{n-1} \sum_{i=1}^{n-1} (f_i - \bar{F}_{n-1})^2 - \frac{2}{n-1} \left(\frac{f_n - \bar{F}_{n-1}}{n} \right) \sum_{i=1}^{n-1} (f_i - \bar{F}_{n-1}) \\
&+ (n-1) \left(\frac{f_n - \bar{F}_{n-1}}{n} \right)^2 + \left(\frac{f_n - \bar{F}_{n-1}}{n} \right)^2 \\
&= \frac{1}{n-1} \sum_{i=1}^{n-1} (f_i - \bar{F}_{n-1})^2 + \frac{(f_n - \bar{F}_{n-1})^2}{n}.
\end{aligned}$$

Finally, by combining Eqs. (7) and (5), the recurrence formula of variance with variable sequence length can be obtained

$$V_n = \frac{n-2}{n-1} V_{n-1} + \frac{(f_n - \bar{F}_{n-1})^2}{n}. \tag{8}$$

According to the Eqs. (3) and (8), it is possible to quickly calculate the variance and mean of a numerical sequence whose length is constantly changing, which saves computing time and computer storage resources to a large extent.

3. Variance Recursion Formula Based on Sliding Window Mechanism

In some cases, there are data updates with invariant sequence length, but at that time, the variance recursion method with variable length cannot effectively calculate the variance of such data. For calculating the variance of such data, a one-step recursive formula for variance is derived using the idea of sliding windows. In addition, based on this, the formula is generalized to obtain an N-step recursive formula for calculating variance. Next, we first prove a one-step recursive formula.

3.1 The Recursion Formula of One-Step

Definition 3.1. Let the observation data sequence be $S = (s_1, s_2, \dots, s_{n-1}, s_n)$, then S is an n-dimensional vector. Hence, the mean of the sequence is $E(S) = \mu = \frac{1}{n} \sum_{i=1}^n s_i$, and the variance is $V(S) = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu)^2$.

Let $S = (s_1, s_2, \dots, s_{n-1}, s_n)$ be S_k and S_{k+1} respectively in the k and $k+1$ states, then the variance of

the sequence in the k and $k + 1$ states are V_k and V_{k+1} , respectively. And the calculation expression is as follows

$$V_k(S) = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu_k)^2, \tag{9}$$

$$V_{k+1}(S) = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu_{k+1})^2. \tag{10}$$

In the k state, let s_j^* replace s_j in the data sequence to get the data sequence $S_{k+1} = (s_1, s_2, \dots, s_{n-1}, s_n)$ in the state $k + 1$, where $s_j = s_j^*$.

Theorem 3.1. Let $\mu_{1,k}$ and $\mu_{1,k+1}$, $V_{1,k}(S)$ and $V_{1,k+1}(S)$ represent the mean and variance of the one-step recursive variance algorithm for the sequence in the k and $k + 1$ states, respectively. Then the following conclusions are established

$$\begin{aligned} V_{1,k+1}(S) &= V_{1,k}(S) - \frac{(s_j - s_j^*)}{n(n-1)} [(1-n)s_j^* - (1+n)s_j + 2n\mu_{1,k}] \\ &= V_{1,k}(S) + \frac{(s_j^* - s_j)}{n(n-1)} [(n-1)s_j^* + (n+1)s_j - 2n\mu_{1,k}]. \end{aligned} \tag{11}$$

Proof. It follows from formulas (9) and (10)

$$\begin{aligned} (n-1)V_{1,k}(S_k) &= \sum_{i=1}^n (s_i - \mu_{1,k})^2 \\ &= (s_1 - \mu_{1,k})^2 + (s_2 - \mu_{1,k})^2 + \dots + (s_j - \mu_{1,k})^2 + \dots + (s_n - \mu_{1,k})^2, \end{aligned} \tag{12}$$

$$\begin{aligned} (n-1)V_{1,k+1}(S_{k+1}) &= \sum_{i=1}^n (s_i - \mu_{1,k+1})^2 \\ &= (s_1 - \mu_{1,k+1})^2 + (s_2 - \mu_{1,k+1})^2 + \dots + (s_j^* - \mu_{1,k+1})^2 + \dots + (s_n - \mu_{1,k+1})^2. \end{aligned} \tag{13}$$

The difference between Eqs. (15) and (16) is obtained as follows:

$$\begin{aligned} &(n-1)V_{1,k}(S_k) - (n-1)V_{1,k+1}(S_{k+1}) \\ &= \left(n(\mu_{1,k})^2 - 2s_1\mu_{1,k} - 2s_2\mu_{1,k} - \dots - 2s_n\mu_{1,k} \right) + s_j^2 \\ &\quad - \left(n(\mu_{1,k+1})^2 - 2s_1\mu_{1,k+1} - 2s_2\mu_{1,k+1} - \dots - 2s_n\mu_{1,k+1} \right) - (s_j^*)^2 \\ &= \left[n(\mu_{1,k})^2 - 2\mu_{1,k}(s_1 + s_2 + \dots + s_{n-1} + s_n) \right] + s_j^2 \\ &\quad - \left[n(\mu_{1,k+1})^2 - 2\mu_{1,k+1}(s_1 + s_2 + \dots + s_{n-1} + s_n) \right] - (s_j^*)^2 \\ &= \left(n(\mu_{1,k})^2 - 2n(\mu_{1,k})^2 \right) + s_j^2 - \left(n(\mu_{1,k+1})^2 - 2n(\mu_{1,k+1})^2 \right) - (s_j^*)^2 \\ &= n(\mu_{1,k+1})^2 - n(\mu_{1,k})^2 + s_j^2 - (s_j^*)^2. \end{aligned} \tag{14}$$

When the length of the sequence is fixed, the recurrence formula of the mean is

$$\mu_{1,k+1} = \mu_{1,k} - \frac{s_j - s_j^*}{n}. \quad (15)$$

Bringing the mean recurrence formula (15) into Eq. (14), the intermediate process is as follows:

$$\begin{aligned} & (n-1)V_{1,k}(S_k) - (n-1)V_{1,k+1}(S_{k+1}) \\ &= n(\mu_{1,k+1})^2 - n(\mu_{1,k})^2 + s_j^2 - (s_j^*)^2 \\ &= n\left(\mu_{1,k} + \frac{s_j - s_j^*}{n}\right)^2 - n(\mu_{1,k})^2 + s_j^2 - (s_j^*)^2 \\ &= n\left[(\mu_{1,k})^2 + \left(\frac{s_j - s_j^*}{n}\right)^2 + \frac{2\mu_{1,k}(s_j - s_j^*)}{n}\right] - n(\mu_{1,k})^2 + s_j^2 - (s_j^*)^2 \\ &= \frac{n}{n-1}\left[(\mu_{1,k})^2 + \left(\frac{s_j - s_j^*}{n}\right)^2 + \frac{2\mu_{1,k}(s_j - s_j^*)}{n}\right] - \frac{n(\mu_{1,k})^2}{n-1} + \frac{(s_j^2 - (s_j^*)^2)}{n-1} \\ &= \frac{(s_j - s_j^*)}{n(n-1)}[s_j - s_j^* + 2n\mu_{1,k} - n(s_j + s_j^*)]. \end{aligned} \quad (16)$$

Simplification of the formula (16) to obtain a fast-recursive formula of the variance of fixed-length sequences

$$\begin{aligned} V_{1,k+1}(S_{k+1}) &= V_{1,k}(S_k) - \frac{(s_j - s_j^*)}{n(n-1)}[(1-n)s_j^* - (1+n)s_j + 2n\mu_{1,k}] \\ &= V_{1,k}(S_k) + \frac{(s_j^* - s_j)}{n(n-1)}[(n-1)s_j^* + (n+1)s_j - 2n\mu_{1,k}]. \end{aligned} \quad (17)$$

Thus Theorem 3.1 is proved.

3.2 The Recursion Formula of N-Step

Since the one-step recursive algorithm in the previous section has certain limitations in its use, it can only calculate the dynamic variance under a data change, so we generalize it to the N-step recursion.

Theorem 3.2. Let S_k and S_{k+1} be as described above, and the N-step variance recursion formula of the data sequence from k to $k+1$ is

$$V_{N,k+1}(S_{k+1}) = V_{N,k}(S_k) - \frac{\left(\sum_{i=1}^N (s_i^* - s_i)\right)^2}{n(n-1)} - \frac{2\mu_{N,k} \sum_{i=1}^N (s_i^* - s_i)}{n-1} - \frac{1}{n-1} \sum_{i=1}^N (s_i^2 - (s_i^*)^2). \quad (18)$$

where, $\mu_{N,k}$ and $\mu_{N,k+1}$, $V_{N,k}$ and $V_{N,k+1}$ represent the mean and variance of the N-step variance recursion algorithm for the data sequence in the k and $k+1$ states, respectively. Specifically, $\mu_{N,k+1}$ is defined as

$$\mu_{N,k+1} = \mu_{N,k} - \frac{\sum_{i=1}^N (s_i^* - s_i)}{n}. \tag{19}$$

Proof. Let the formula for calculating the mean and variance of the data sequence in the state k are as follows

$$V_{N,k}(S_k) = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu_{N,k})^2. \tag{20}$$

Suppose that the data sequence $(s_1^*, s_2^*, \dots, s_N^*)$ of length $N(1 \leq N \leq n)$ is used to replace the data sequence (s_1, s_2, \dots, s_N) of the same length in S_k , and the new data sequence S_{k+1} in the state of $k + 1$ is obtained, and the variance of the sequence is set as

$$V_{N,k+1}(S_{k+1}) = \frac{1}{n-1} \sum_{i=1}^n (s_i - \mu_{N,k+1})^2. \tag{21}$$

It can be obtained from Eqs. (20) and (21),

$$\begin{aligned} (n-1)V_{N,k}(S_k) &= \sum_{i=1}^n (s_i - \mu_{N,k})^2 \\ &= (s_1 - \mu_{N,k})^2 + \dots + (s_N - \mu_{N,k})^2 + \dots + (s_n - \mu_{N,k})^2, \end{aligned} \tag{22}$$

$$\begin{aligned} (n-1)V_{N,k+1}(S_{k+1}) &= \sum_{i=1}^n (s_i - \mu_{N,k+1})^2 \\ &= (s_1^* - \mu_{N,k+1})^2 + \dots + (s_N^* - \mu_{N,k+1})^2 + \dots + (s_n - \mu_{N,k+1})^2. \end{aligned} \tag{23}$$

The difference between Eqs. (22) and (23) is obtained

$$\begin{aligned} &(n-1)V_{N,k}(S_k) - (n-1)V_{N,k+1}(S_{k+1}) \\ &= \left(n(\mu_{N,k})^2 - 2s_1\mu_{N,k} - \dots - 2s_N\mu_{N,k} - \dots - 2s_n\mu_{N,k} \right) + \sum_{i=1}^N s_i^2 + \sum_{i=N+1}^n s_i^2 \\ &\quad - \left(n(\mu_{N,k+1})^2 - 2s_1^*\mu_{N,k+1} - \dots - 2s_N^*\mu_{N,k+1} - \dots - 2s_n\mu_{N,k+1} \right) - \sum_{i=1}^N (s_i^*)^2 - \sum_{i=N+1}^n s_i^2 \\ &= \left[n(\mu_{N,k})^2 - 2\mu_{N,k}(s_1 + s_2 + \dots + s_N + \dots + s_{n-1} + s_n) \right] + \sum_{i=1}^N s_i^2 \\ &\quad - \left[n(\mu_{N,k+1})^2 - 2\mu_{N,k+1}(s_1^* + s_2^* + \dots + s_N^* + \dots + s_{n-1} + s_n) \right] - \sum_{i=1}^N (s_i^*)^2 \\ &= \left(n(\mu_{N,k})^2 - 2n(\mu_{N,k})^2 \right) + \sum_{i=1}^N s_i^2 - \left(n(\mu_{N,k+1})^2 - 2n(\mu_{N,k+1})^2 \right) - \sum_{i=1}^N (s_i^*)^2 \\ &= n(\mu_{N,k+1})^2 - n(\mu_{N,k})^2 + \sum_{i=1}^N (s_i^2 - (s_i^*)^2). \end{aligned} \tag{24}$$

Simplify the multivariate variance recursion calculation formula by simplification Eq. (24), then

$$V_{N,k+1}(S_{k+1}) = V_{N,k}(S_k) - \frac{n}{n-1}(\mu_{N,k+1})^2 + \frac{n}{n-1}(\mu_{N,k})^2 - \frac{1}{n-1} \sum_{i=1}^N (s_i^2 - (s_i^*)^2). \quad (25)$$

When the length of the sequence is fixed, the mean in the k state is calculated as

$$\mu_{N,k} = \frac{s_1 + s_2 + \cdots + s_N + \cdots + s_n}{n}. \quad (26)$$

The mean calculation formula in the $k+1$ state is

$$\mu_{N,k+1} = \frac{s_1^* + s_1^* + \cdots + s_N^* + \cdots + s_n}{n}. \quad (27)$$

From the above Eqs. (26) and (27), the N-step mean recurrence formula can be obtained, which is calculated as follows

$$\mu_{N,k+1} = \mu_{N,k} - \frac{\sum_{i=1}^N (s_i^* - s_i)}{n}. \quad (28)$$

Bring Eq. (28) into Eq. (25) to get the N-step variance recursion calculation formula.

$$V_{N,k+1}(X_{k+1}) = V_{N,k}(S_k) - \frac{\left(\sum_{i=1}^N (s_i^* - s_i)\right)^2}{n(n-1)} - \frac{2\mu_{N,k} \sum_{i=1}^N (s_i^* - s_i)}{n-1} - \frac{1}{n-1} \sum_{i=1}^N (s_i^2 - (s_i^*)^2). \quad (29)$$

Thus Theorem 3.2 is proved.

4. Numerical Simulation

The main content of this section is to verify the accuracy and efficiency of the recursive formula proposed in this paper through numerical experiments. The sample data used for this numerical experiment is divided into two groups, each containing 20,000 random numbers between (-1,1).

4.1 Comparison of the One-Step Recursive Algorithm

For the random sequence used for the one-step recursive algorithm comparison, the method of random replacement is used. Randomly replace one data in the sequence each time, and then use the defined algorithm, the algorithm proposed by Deng et al. [29] and the recursive algorithm proposed in this paper to calculate the variance of the sequence. A total of 40 replacement operations are performed. Record the calculation time used each time and compare it with the definition algorithm and the Deng's algorithm. The calculation results of the one-step recursive algorithm compared with the other two algorithms are shown in Figs. 1 and 2. As can be seen from the Fig. 1, the algorithm proposed in this paper has higher time efficiency than the traditional definition algorithm and Deng's optimization algorithm. From Fig. 2, the calculation accuracy of the proposed algorithm is also higher than that of Deng's optimization algorithm.

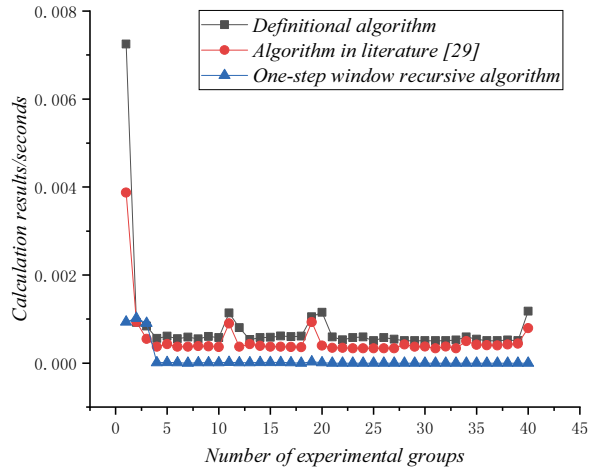


Fig. 1. Time efficiency comparison of three algorithms (one-step recursive algorithm).

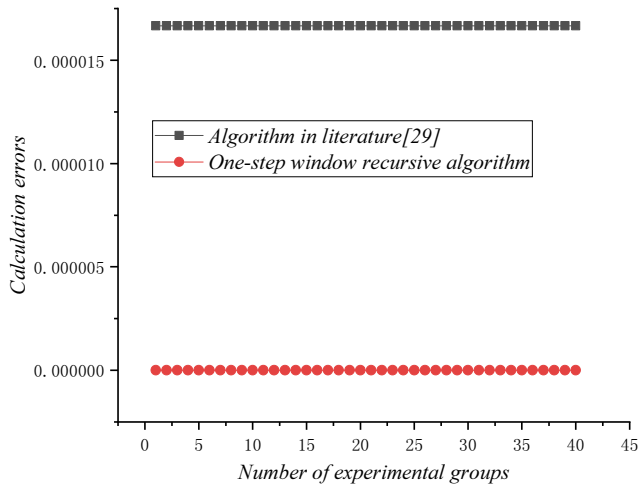


Fig. 2. Calculation accuracy comparison (one-step recursive algorithm).

4.2 Comparison of the N-Step Recursive Algorithm

Similar to the previous section, this section mainly studies the comparison of three different algorithms when the N-step of the data sequence is updated. For the random sequence used to compare the N-step recursive algorithm, the number of data replaced at each time in the sequence satisfies an arithmetic sequence with 500 as the first term and 500 as the tolerance. Then, the classical definition formula, the recursive formula proposed by Deng, and the newly proposed N-step recursive formula are used to calculate the variance of this dynamically changing random sequence. And the time consumption and calculation error of the three algorithms are recorded in each calculation. The calculation results of the N-step recursive algorithm and comparison with the other two algorithms are given in Figs. 3 and 4. As can be seen from the figures, when the sequence length is fixed, and the data is dynamically updated and replaced multiple times, the time efficiency and calculation accuracy of the N-step recursive algorithm are better than the definition algorithm and the improved algorithm of Deng et al. [29].

This numerical experiment clearly shows that the recursive formula proposed in this paper has advantages that other algorithms do not have when calculating the variance of the time-varying data series with a constant sample size. By introducing the idea of sliding windows, the efficiency of calculating the variance of such data series is greatly improved.

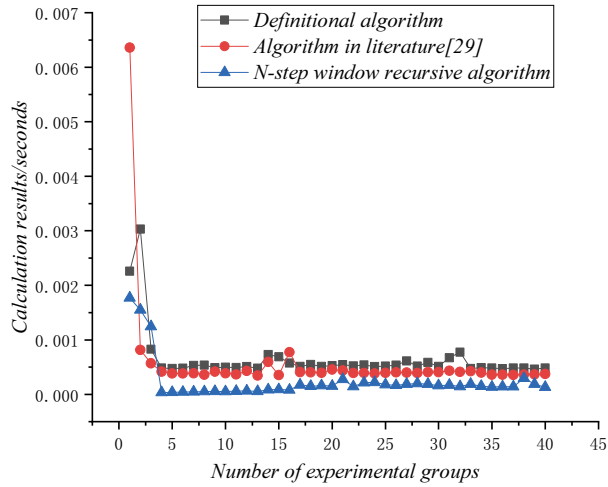


Fig. 3. Time efficiency comparison of three algorithms (N-step recursive algorithm).

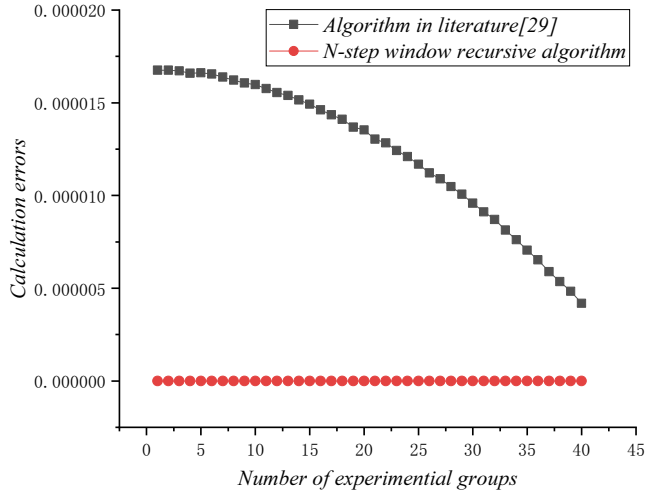


Fig. 4. Calculation accuracy comparison (N-step recursive algorithm).

5. Conclusion

For a sequence data with fixed length and fast updating of data, the traditional algorithm using classical definition will cost large time and resource consumption due to $O(n)$ time complexity and frequent access to the computer disk, which may cause performance degradation of the whole system. However, the N-

step sliding recursion formula proposed in this paper does not need to read the entire data sequence from the computer disk each time after the data is updated. Instead, the variance of the updated sequence can be directly calculated based on the results of the previous calculation and the current updated data, which shortened the calculation time from $O(n)$ to $O(1)$ time complexity. It can not only improve the time efficiency but also greatly save the storage resources of the computer, making the data analysis real-time and low-cost. In addition, this type of recursion can be extended to the calculation of kurtosis and skews, or the calculation of higher-order origin moments and central moments, which demonstrates the potential value for time-consumption data analysis or high speed signal processing.

Acknowledgement

This paper has been supported by Sichuan Science and Technology Program (No. 2020YFS0318, 2019YFS0155, 2019YFS0146, 2020YFG0430, 2020YFS0307), the Educational Reform Research Project of Southwest University of Science and Technology (No. 15xnzd05), Undergraduate Innovation Fund Project Accurate Funding Special Project by Southwest University of Science and Technology (No. JZ19-057).

References

- [1] R. A. Fisher, "XV.—The correlation between relatives on the supposition of Mendelian inheritance," *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, vol. 52, no. 2, pp. 399-433, 1919.
- [2] S. Orcioni, S. Cecchi, and A. Carini, "Multivariate nonlinear system identification using wiener basis functions and perfect sequences," in *Proceedings of 2017 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, 2017, pp. 2679-2683.
- [3] S. Baek and D. Y. Kim, "Abrupt variance and discernibility analyses of multi-sensor signals for fault pattern extraction," *Computers & Industrial Engineering*, vol. 128, pp. 999-1007, 2019.
- [4] S. Durant, I. Sulykos, and I. Czigler, "Automatic detection of orientation variance," *Neuroscience Letters*, vol. 658, pp. 43-47, 2017.
- [5] J. Y. Gotoh, M. J. Kim, and A. E. Lim, "Robust empirical optimization is almost the same as mean-variance optimization," *Operations Research Letters*, vol. 46, no. 4, pp. 448-452, 2018.
- [6] N. Gatzert, "An analysis of transaction costs in participating life insurance under mean-variance preferences," *Insurance: Mathematics and Economics*, vol. 85, pp. 185-197, 2019.
- [7] C. S. Pun, "Time-consistent mean-variance portfolio selection with only risky assets," *Economic Modelling*, vol. 75, pp. 281-292, 2018.
- [8] M. Zhang and P. Chen, "Mean-variance asset-liability management under constant elasticity of variance process," *Insurance: Mathematics and Economics*, vol. 70, pp. 11-18, 2016.
- [9] T. Bodnar, N. Parolya, and W. Schmid, "Estimation of the global minimum variance portfolio in high dimensions," *European Journal of Operational Research*, vol. 266, no. 1, pp. 371-390, 2018.
- [10] A. Clements and Y. Liao, "Forecasting the variance of stock index returns using jumps and cojumps," *International Journal of Forecasting*, vol. 33, no. 3, pp. 729-742, 2017.
- [11] A. Kaeck, P. Rodrigues, and N. J. Seeger, "Equity index variance: evidence from flexible parametric jump-diffusion models," *Journal of Banking & Finance*, vol. 83, pp. 85-103, 2017.

- [12] N. Al Aamery, J. F. Fox, M. Snyder, and C. V. Chandramouli, "Variance analysis of forecasted streamflow maxima in a wet temperate climate," *Journal of Hydrology*, vol. 560, pp. 364-381, 2018.
- [13] A. B. Naik and A. C. Reddy, "Optimization of tensile strength in TIG welding using the Taguchi method and analysis of variance (ANOVA)," *Thermal Science and Engineering Progress*, vol. 8, pp. 327-339, 2018.
- [14] I. S. Sobolev, A. N. Orekhov, T. Bratec, L. P. Rikhvanov, and N. P. Soboleva, "Variance-Correlation analysis in the exploration of hydrothermal (fluidogenous) deposits using surface gamma-ray spectrometry," *Journal of Applied Geophysics*, vol. 159, pp. 597-604, 2018.
- [15] G. Jiang and W. Wang, "Error estimation based on variance analysis of k-fold cross-validation," *Pattern Recognition*, vol. 69, pp. 94-106, 2017.
- [16] Y. Zhao, Y. Ming, X. Liu, E. Zhu, K. Zhao, and J. Yin, "Large-scale k-means clustering via variance reduction," *Neurocomputing*, vol. 307, pp. 184-194, 2018.
- [17] A. M. Deylami and B. M. Asl, "Iterative minimum variance beamformer with low complexity for medical ultrasound imaging," *Ultrasound in Medicine & Biology*, vol. 44, no. 8, pp. 1882-1890, 2018.
- [18] Z. Jin, L. Min, M. K. Ng, and M. Zheng, "Image colorization by fusion of color transfers based on DFT and variance features," *Computers & Mathematics with Applications*, vol. 77, no. 9, pp. 2553-2567, 2019.
- [19] G. Simarro, K. R. Bryan, R. M. Guedes, A. Sancho, J. Guillen, and G. Coco, "On the use of variance images for runoff and shoreline detection," *Coastal Engineering*, vol. 99, pp. 136-147, 2015.
- [20] X. Chen, A. Molina-Cristobal, M. D. Guenov, and A. Riaz, "Efficient method for variance-based sensitivity analysis," *Reliability Engineering & System Safety*, vol. 181, pp. 97-115, 2019.
- [21] M. I. Radaideh, S. Surani, D. O'Grady, and T. Kozlowski, "Shapley effect application for variance-based sensitivity analysis of the few-group cross-sections," *Annals of Nuclear Energy*, vol. 129, pp. 264-279, 2019.
- [22] W. Yun, Z. Lu, K. Zhang, and X. Jiang, "An efficient sampling method for variance-based sensitivity analysis," *Structural Safety*, vol. 65, pp. 74-83, 2017.
- [23] H. Chen and H. Gao, "Unified recursion formula of k step original matrix in common discrete distribution," *Journal of Human Business College*, vol. 2002, no. 2, pp. 110-111, 2002.
- [24] Q. Chen, "Discrete distribution's recursion formula of k-taped origin matrix," *Statistics & Information Tribune*, vol. 2000, no. 1, pp. 26-28, 2000.
- [25] Q. Chen and Y. Ma, "Discrete distribution's recursion formula of k-taped central moment," *Statistics & Information Tribune*, vol. 1999, no. 2, pp. 36-38, 1999.
- [26] C. G. Chen, "The counting method of k-order central moment of three kinds of discrete random variable," *Journal of Jiangnan University (Social Science Edition)*, vol. 2000, no. 6, pp. 66-68, 2000.
- [27] S. Fan, "Recurrence formula calculating the statistics for changed sample size," *Journal of Inner Mongolia Agricultural University (Natural Science Edition)*, vol. 11, no. 1, pp. 153-162, 1990.
- [28] J. Liu and X. Sun, "Real-time calculation method of variance of non-zero mean measurement and control data," *Mathematics Learning and Research*, vol. 2016, no. 21, pp. 148-150, 2016.
- [29] H. B. Deng, J. F. Liu, and Y. N. Wang, "Recursive algorithm for mean variance and its application," *Computer and Modernization*, vol. 1996, no. 4, pp. 9-11, 1996.
- [30] A. S. Chen, H. C. Chang, and L. Y. Cheng, "Time-varying variance scaling: application of the fractionally integrated ARMA model," *The North American Journal of Economics and Finance*, vol. 47, pp. 1-12, 2019.
- [31] A. M. Chu, R. W. Li, and M. K. So, "Bayesian spatial-temporal modeling of air pollution data with dynamic variance and leptokurtosis," *Spatial Statistics*, vol. 26, pp. 1-20, 2018.
- [32] Y. Yang and S. Wang, "Two simple tests of the trend hypothesis under time-varying variance," *Economics Letters*, vol. 156, pp. 123-128, 2017.
- [33] J. Wang, L. Wang, and X. He, "The study of high accuracy time keeping based on FPGA when navigation satellite losing connection," *Chinese Journal of Electron Devices*, vol. 39, no. 1, pp. 140-143, 2016.



Lang Yu <https://orcid.org/0000-0002-0777-8750>

He studied Information and Computing Science at the School of Science from Southwest University of Science and Technology, Mianyang, China, Since 2016. His main research interests include algorithm optimization, grey systems, and computational mathematics.



Gang He <https://orcid.org/0000-0002-4029-7596>

He received the B.S. degree in Engineering Mechanics from Sichuan University, Chengdu, China, in 2006. He studied Biomedical Engineering for 7 years when he received the M.S. degree in 2009 and received the Ph.D. degree in Professor GuangFu Yin's Laboratory of Sichuan University in 2013. He worked as post-doctorate at the Sichuan University for 2 years in the field of First principle calculations on Titanium materials. He worked as a lecturer at Southwest University of Science and Technology, Mianyang, Sichuan, China, since 2015.



Ahmad Khwaja Mutahir <https://orcid.org/0000-0002-1115-7690>

He received the B.S. degree in the School of Computer Science and Technology from Islamia College University, Peshawar, Pakistan, in 2015. Now, he studied M.S. Computing Science and Technology at the School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China. His main research interests include digital neural network (machine learning).