Research on Variance Change Point Detection and Subsampling Test

Jielin Yang

College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710054, China.

Abstract

This paper delves into the study of variance change point detection and Subsampling tests, proposing a method for variance change point detection based on sliding window mechanism. It also explores the principles of Subsampling and its operational steps in resampling for variance change point detection. The paper employs hypothesis testing to study cumulative sum statistics within a window, using the central limit theorem to prove that under the null hypothesis, the statistics converge in distribution to a functional of Brownian motion Under the alternative hypothesis, the test statistics are asymptotically consistent, and the divergence rate is provided. The sliding window mechanism is used to segment the data and the statistics, and then the Python software is used to simulate the numerical process of the variance change point detection model. This simulation yields critical value tables and empirical level under the null hypothesis, and empirical power tables at different change points under the alternative hypothesis. The analysis of the experimental results shows that the empirical power at the middle point is slightly higher than that of the time series before and after.

Keywords

Variance Change Point Detection, Subsampling Resampling, Cumulative Sum Statistics, Numerical Simulation.

1. Introduction

In the era of big data and artificial intelligence, time series data is recorded and stored in large volumes. A time series is a set of random variables ordered by time, referring to a sequence of numerical values of the same statistical indicator arranged in the order of occurrence. However, over time, the data may experience sudden jumps. For instance, at a certain position or moment, the numerical values or fluctuations of consecutive observations abruptly, meaning there is a significant statistical difference in the mean or variance of the data before and after this point. This is known as a mean change point or change point. If a change point appears in the data and the previous model is still used, it can lead to statistical inference errors. To avoid such errors, point analysis needs to be conducted on the data. The problem of change points was first introduced by Page[1] for detecting mean change points in independent distributions. Since then, change point problems have become a hot topic in many research fields such as statistics, medicine, engineering, and economics, with numerous research findings. An important problem in the detection of structural breaks in time series data is the detection of changes in a sequence of some parameter, numerical characteristic, or distribution in the model alters, such as a shift in mean [2], a change in variance [3] a change in tail exponent [4], a change in persistence [5-6] etc. A change point refers to a specific time point in a series of data points where the statistical properties or distribution of the data change significantly. This point is called a change point.

Research on change-point problems is an emerging field that spans multiple disciplines, including statistics, mathematics, computer science, and data mining. core of research in this field focuses on how to identify and predict change points in data, which are specific moments when the data distribution or statistical characteristics undergo significant changes Since the 1970s, the field of statistics has been paying attention to change-point problems, and its application range is very wide, covering industrial economic, financial, and earthquake prediction fields. The theoretical research on change-point problems has become relatively mature, especially in the application of maximum likelihood method, least squares, and CUSUM method, which play an important role in the estimation of change points and the study of their properties. In recent years, research on change detection problems has continued to make new progress. Researchers have reasonably classified change-point detection problems from different perspectives and have looked forward to the future development direction of the. This indicates that, despite the relatively rich theoretical foundation of change-point problems, exploration in both application and theory is still continuously deepening and expanding. Overall, on change-point problems not only has significant implications for the development of statistical theory but also has broad application prospects in many practical fields. With the advancement of technology and increase in data analysis needs, research in this field will continue to be valued and is expected to produce more innovative results. In the study of mean change-point problems, Chernoff and Zacks first proposed a test statistic for detecting a change point in the mean of independent distributions. Regarding the widespread use of least squares in linear processes, Bai used this method to study the estimation and testing of single change-point problems. For variance change models, Kilian and Zha developed a subsampling method to estimate the variance change-point model and introduced a subsampling test to determine the existence of a variance point. Loretan and Mollard proposed a method based on spectral analysis to detect variance change points. Quandt and Andrews, among others, proposed asymptotic testing methods to detect variance change points.

2. Theoretical foundation

In simple terms, a changepoint is a point or set of points in a continuous process where parameters change before and after the point, and the where the value changes are called the changepoint moment. In statistics, it is believed that these changing parameters are not only manifested in a single quantity but also a sudden change in mathematical characteristics or statistics. Therefore, changepoint problems can be divided into different types. Based on different application scenarios, researchers have constructed various models and test statistics to detect specific types of change points, and have studied the asymptotic distribution convergence rate, and convergence of sample data. The research on change point estimation focuses on estimating the actual location of the change point, as well as analyzing its properties distribution, and convergence. In the current academic literature on change point problems, common models include the mean change point model, the persistence change point model, and the change point model.

Mean change-point model and persistent change point model:

$$y_t = \begin{cases} \mu_1 + \varepsilon_t, 1 \le t \le [N\tau], \\ \mu_2 + \varepsilon_t, [N\tau] \le t \le N. \end{cases}$$

$$y_t = \begin{cases} c_1 + m_t, 1 \le t \le [N\tau], \\ c_2 + n_t, [N\tau] \le t \le N. \end{cases}$$

In the model, c_1 and c_2 are both constants, where $\{m_t\}$ is a stationary process and $\{n_t\}$ is unit root process, where N is the sample size and τ is the changepoint.

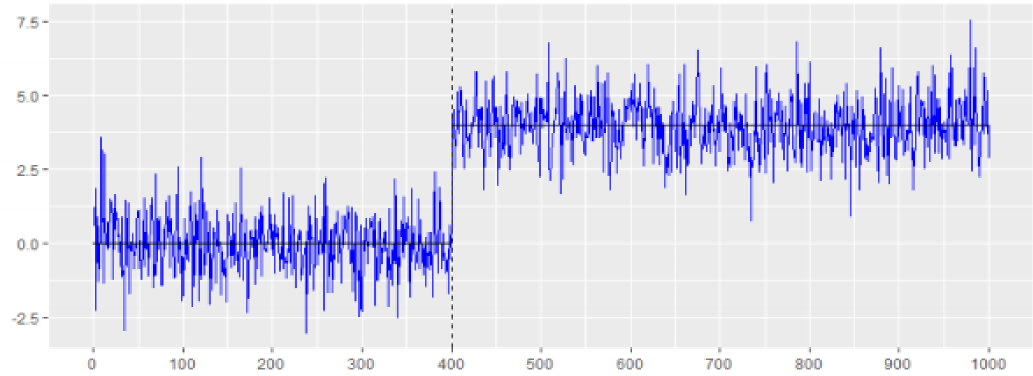

Fig. 1 Mean Change Point

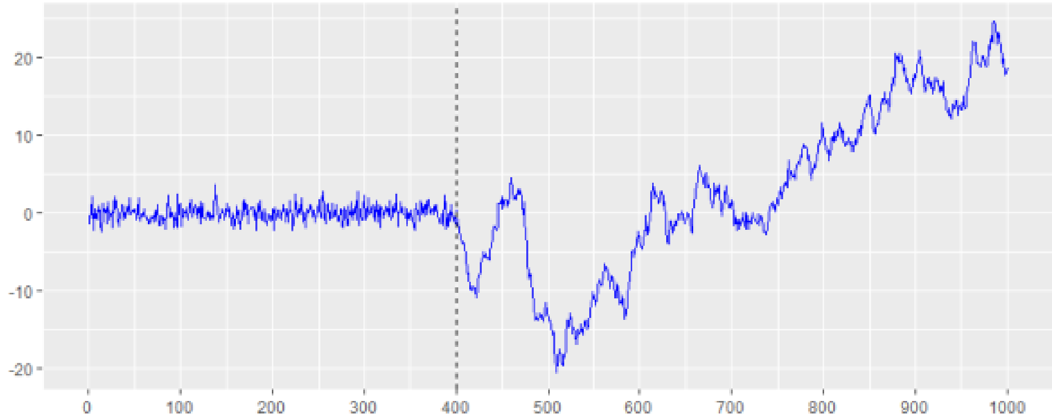

Fig. 2 Persistent changepoint

As shown in Figure 1, when the sample size is 1000, the change point is set at τ*=04, which means the change point is at the 400th sample point with a mean of 0 before the change point and a mean of after the change point. In the figure, the change point is easily observable due to the large magnitude of the change. However, when the magnitude of the mean is small, it is impossible to directly observe the magnitude of the mean change and the location of the change point. Therefore, it becomes necessary to use statistical methods change point detection and location estimation.

As shown in Figure 2, the time series changes from a stationary process to a unit root process, with a sudden change in persistence τ = 0.4, which occurs at sample point 400.

3. CUSUM test for variance change-point detection

In the study of variance change point detection, different test statistics are affected by sample size *N*, change point location τ, and the magnitude of variance and other variable parameters. In this paper, we will study the problem of variance change point detection by constructing cumulative sum statistics. The Sliding Window Method is a simple and intuitive approach in variance change point detection. It involves sliding a fixed-sized window *L* over time series data and calculating the variance of the data within each window. This allows for the observation of how the variance changes over time. When the difference in variance two consecutive windows exceeds a certain threshold, it can be inferred that a variance change point exists at the boundary of the windows. The specific steps of the Sliding Method are as follows: (1) Window Size Selection: First, an appropriate window size is determined. The window size affects the sensitivity of the detection; a window can smooth out random fluctuations but may delay the detection of a change point. A smaller window can increase the sensitivity of the change point

ISSN: 1813-4890

detection but may be affected random fluctuations. (2) Variance and Critical Value Calculation: The window is slid along the time series, and the variance of the data within each window and critical value are calculated. (3) Variance Comparison: The variances of adjacent windows are compared. If the difference in variance between two adjacent windows is large exceeds a pre-set threshold, it can be inferred that a variance change point exists at the boundary of the windows. (4) Change Point Determination: Based the comparison results, the position of the variance change point is determined. Sometimes, to reduce false positives, statistical tests can be used to determine the significance of the point, such as the Chow test or the F test.

The cumulative sum and statistics are defined as follows:

$$\sum(i,j) = \max_{0\le\tau\le1} \frac{\sqrt{L}\hat{\psi}}{\hat{V}} \left| \frac{\sum_{t=i}^{i+[L\tau]-1} \widehat{u_t}^2}{\sum_{t=i}^{j} \widehat{u_t}^2} - \frac{[L\tau]}{L} \right|$$

$$H_0\colon\ y_t = (\mu + u_t) \quad u_t = (\sigma\varepsilon_t)$$

$$H_1\colon\ y_t \neq (\mu + u_t) \quad u_t = (\sigma\varepsilon_t)$$

$$\sigma = \begin{cases} \sigma_1, t \le k^* \\ \sigma_2, t > k^* \end{cases}$$

Under the null hypothesis, as L approaches infinity, the cumulative sum statistic for variance change points converges in distribution to a functional of Brownian motion. Under the null hypothesis, the cumulative sum statistic diverges, and the divergence rate is fastest at the k^* moment.

4. Numerical simulation

In statistics, there are two methods of resampling data: one is the Bootstrap resampling method, and the other is the Subsampling resampling statistical method. This is the hypothesis test of our variance change-point model. What our variance change-point monitoring needs to do is to calculate the critical by sliding window mechanism on the sample, then evaluate the power under the null hypothesis and simulate the size value (i.e., the rejection under the null hypothesis); then evaluate the power under the alternative hypothesis and simulate the empirical power.

Table 1 Critical value

Numble	$\sigma_1 = 1$	$\sigma_1 = 1.5$	$\sigma_1 = 2$
N=200	1.2263	1.2286	1.2376
N=300	1.2555	1.2452	1.2524
N=500	1.2726	1.2789	1.2796
N=800	1.2986	1.2877	1.2935
N=1000	1.2974	1.2943	1.3042

Table 2 Empirical potential table with τ^*=0.3

Numble	□=0.5			□=1			□=2		
	$\sigma_1 = 1$	$\sigma_1 = 1.5$	$\sigma_1 = 2$	$\sigma_1 = 1$	$\sigma_1 = 1.5$	$\sigma_1 = 2$	$\sigma_1 = 1$	$\sigma_1 = 1.5$	$\sigma_1 = 2$
N=200	0.6684	0.6138	0.5804	0.7554	0.7022	0.6406	0.7946	0.7666	0.7311
N=300	0.7558	0.6926	0.6726	0.8112	0.7706	0.7467	0.8193	0.8102	0.8025
N=500	0.8246	0.7794	0.7421	0.8397	0.8285	0.8129	0.8429	0.8353	0.8318
N=800	0.8118	0.7931	0.7562	0.8257	0.8287	0.8166	0.8312	0.8340	0.8330
N=1000	0.8271	0.8059	0.7756	0.8322	0.8323	0.8251	0.8344	0.8376	0.8310

ISSN: 1813-4890

To make the experimental results more convincing, we loop T=2000 times in each numerical simulation. When setting the sliding window, we choose the window L to be 1/5 of the sample size N, the step M=4, so the number of windows is $\frac{N-L}{M}$. And when calculating the critical value, we choose the empirical 95% to approximate the critical value under the null hypothesis. We choose the sample size N=200,300,500,800,1000 for simulation calculations. Since there is no change-point under the null hypothesis, we only need to change the sample size N and the variance before the change-point when simulating the critical value and size. However, when simulating the empirical power, since there is a change-point under the alternative hypothesis we need to change the sample size N, the variance before the change-point, the change-point time τ, and the variance change amplitude to how these parameter changes affect the empirical power. Next, I will present the simulation results of the critical value C, size, and power obtained from the algorithm running Python, and analyze the simulation results.

The empirical level is an important parameter in the change-point problem. In this regard, Li Yuanyuan from Xi'an University of Science and Technology found that under the null hypothesis, the empirical level increases with the sample size and the magnitude of the variance jump, and decreases with the increase of long memory index, in the study of mean change-point and unit root tests in long memory sequences.

As can be seen from this process, we need to repeatedly use the variance change point cumulative sum statistics in the calculation, and the number of obtained will be very large, while the final critical value fluctuates around 1.2. It can be seen that the critical value obtained by our numerical simulation very convincing. The determination of the critical value not only completes the calculation of a large number of statistics under the null hypothesis, but also provides a basis for comparison we calculate the size and empirical power later, which makes all the statistics in the later work comparable.

5. Conclusion

In contrast to the traditional offline variance change-point tests, we constructed a cumulative sum statistic for monitoring variance change-points using the central theorem and sliding window mechanism. We then used the theoretical methods of hypothesis testing to prove that under the null hypothesis, the limit of this statistic converges in distribution a functional of Brownian motion. We also proved that under the alternative hypothesis, the statistic diverges consistently, and provided the corresponding divergence rate. This process that the cumulative sum statistic is related to the sample size N and the magnitude of the variance change, proving the feasibility of the statistic in solving the variance change-point. We introduced the principle and steps of the subsampling resampling method, and then further verified the scientific validity of the cumulative sum statistic using Python software. We the critical values under the null hypothesis, the empirical power under the null hypothesis, and the empirical power under the alternative hypothesis. Through the analysis of experimental results, found that under the null hypothesis, the empirical level tends to the significance level α as the sample size increases; under the alternative hypothesis, the empirical power is positively with the sample size N and the magnitude of the variance change; the empirical power is slightly higher when the change-point is in the middle of the time series than the first half or the second half of the series. These results also indirectly indicate that changes in these parameters have a significant impact on the performance of the statistic.

Acknowledgements

Natural Science Foundation (71473194).

References

[1] Page E. S. Continuous inspection schemes, Biometrika, vol. 41 (1954) No.2, p.100-115.

[2] James B., James K. L., & Siegmund D. Asymptotic approximations for likelihood ratio tests and confidence regions for a change-point in the mean of a multivariate normal distribution, Statistica Sinica, vol. 2 (1992) No.1, p. 69-90.

[3] Hawkins D. M., & Zamba K. D. A change-point model for a shift in variance, Journal of Quality Technology, vol. 37 (1994) No.1, p.21-31.

[4] Quintos C., Fan Z., & Phillips P. Structural Change Tests in Tail Behaviour and the Asian Crisis, Social Science Electronic Publishing, vol. 68 (2001) No.3, p. 633-663.

[5] Sohrabi M., & Zarepour M. Asymptotic theory for M-estimates in unstable AR (p) processes with infinite variance innovations, Journal of Statistical Planning and Inference, vol. 198 (2019) No.3, p. 105-118.

[6] Xiaofeng Zhang, Hao Jin, & Yunfeng Yang. Quasi-autocorrelation coefficient change test of heavy-tailed sequences based on M-estimation, AIMS Mathematics, vol. 9 (2024) No.7, p. 19569-19596.