

УДК 658.562.012.7

В.В. Овсянникова

АНАЛИЗ ЭФФЕКТИВНОСТИ АЛГОРИТМА ОБОБЩЁННОЙ ДИСПЕРСИИ ПРИ НАЛИЧИИ ТРЕНДА МНОГОМЕРНОГО РАССЕЙЯНИЯ ПРОЦЕССА

Овсянникова Виктория Владиславовна, окончила Ульяновский государственный технический университет по направлению «Прикладная математика» факультета информационных систем и технологий. Имеет статьи в области математического моделирования в задачах управления качеством. [e-mail: ovvika17@gmail.com].

Аннотация

Алгоритм обобщённой дисперсии используется для контроля многомерного рассеяния процесса. Под обобщённой дисперсией понимается определитель ковариационной матрицы процесса. Эффективность контроля оценивается по средней длине серий – количеству наблюдений от момента нарушения процесса до момента обнаружения этого нарушения. Для оценивания средней длины серий при различных вариантах контроля можно воспользоваться как аналитическими, так и численными методами. Один из наиболее распространенных типов нарушения процесса – это скачкообразное или постепенное увеличение его рассеяния. Цель работы – оценка средней длины серий карты обобщенной дисперсии при наличии тренда многомерного рассеяния.

Ключевые слова: многомерный статистический контроль, средняя длина серий, обобщённая дисперсия, тренд рассеяния процесса.

doi: 10.35752/1991-2927-2021-3-65-56-60

THE EFFICIENCY ANALYSIS OF GENERALIZED VARIANCE ALGORITHM UNDER THE TREND OF MULTIDIMENSIONAL SCATTERING

Viktoriya Vladislavovna Ovsiannikova, graduated from the Applied Mathematics Department of the Faculty of Information Systems and Technologies of Ulyanovsk State Technical University; an author of articles in the field of mathematical modeling in quality management problems. e-mail: ovvika17@gmail.com.

Abstract

The generalized variance algorithm is used to control the multidimensional scattering of the process. The generalized variance is understood as the determinant of the covariance matrix of the process. The control efficiency is estimated by the average length of the series, that is the number of observations from the moment of violation of the process to the moment of detection of this violation. Both analytical and numerical methods can be used to estimate the average length of the series for different control types. One of the most common types of the process violation is an abrupt or gradual increase in its dispersion. The aim of the article is to estimate the average length of the series of the generalized dispersion map under the trend of multidimensional scattering.

Keywords: multivariate statistical control, average series length, generalized variance, process scattering trend.

ПОСТАНОВКА ЗАДАЧИ

Многомерный статистический контроль, как и контроль независимых показателей процесса, осуществляется по среднему уровню процесса и по рассеянию [1–3]. Алгоритм обобщённой дисперсии используется для контроля многомерного рассеяния. Под обобщённой дисперсией понимается определитель ковариационной матрицы процесса.

Одним из наиболее распространенных нарушений процесса является скачкообразное увеличение рассеяния, при этом определитель ковариационной матрицы процесса $|\Sigma|$ увеличивается в d раз по сравнению с исходным значением:

$$|\Sigma| = d |\Sigma_0|. \quad (1)$$

Эффективность контроля оценивается по средней длине серий – количеству наблюдений от момента нарушения процесса до момента обнаружения этого нарушения. Для оценивания средней длины серий при различных вариантах контроля можно воспользоваться как аналитическими, так и численными методами.

Соответствующие расчеты и испытания при нарушении типа (1) приведены в статье [4].

Другой распространенный вариант нарушения стабильности процесса – постепенное увеличение рассеяния (тренд). В этом случае

$$|\Sigma| = [1 + d_0(t - t_0)], \quad (2)$$

где t – текущий номер выборки, t_0 – номер выборки, с которой начался тренд ($t - t_0$ – длительность тренда), d_0 – характеристика интенсивности тренда. Этот тип нарушения рассмотрим подробнее.

Цель настоящей работы – оценка средней длины серий для контрольной карты обобщенной дисперсии при наличии тренда многомерного рассеяния.

СРЕДНЯЯ ДЛИНА СЕРИЙ КАРТЫ ОБОБЩЕННОЙ ДИСПЕРСИИ ПРИ НАЛИЧИИ ТРЕНДА

Средняя длина серий контрольной карты при наличии тренда может быть найдена по формуле [4]:

$$L(d_0) = \frac{1}{1 - P(d_0)} = \frac{1}{1 - P(LCL < |\Sigma| < UCL; |\Sigma| = [1 + d_0(t - t_0)]|\Sigma_0|)}, \quad (3)$$

где для алгоритма обобщенной дисперсии $P(d_0)$ – вероятность попадания обобщенной дисперсии $|\Sigma|$ в доверительный интервал при условии, что она увеличилась в $1 + d_0(t - t_0)$ раз по сравнению с заданной (нормативной) обобщенной дисперсией $|\Sigma_0|$; положение контрольных границ с учетом приближенной нормальности распределения определяется соотношением [2]

$$m_{|\Sigma_0|} \pm u_{1-\frac{\alpha}{2}} \sigma_{|\Sigma_0|}, \quad (4)$$

где $u_{1-\frac{\alpha}{2}}$ – квантиль нормального распределения порядка $1 - \frac{\alpha}{2}$;

α – уровень значимости;

математическое ожидание обобщенной дисперсии

$$m_{|\Sigma_0|} = b_1 |\Sigma_0|, \quad (5)$$

ее стандартное отклонение

$$\sigma_{|\Sigma_0|} = \sqrt{b_2} |\Sigma_0|, \quad (6)$$

где коэффициенты

$$b_1 = \frac{1}{(n-1)^p} \prod_{j=1}^p (n-j); \quad (7)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{j=1}^p (n-j) \left[\prod_{k=1}^p (n-k+2) - \prod_{k=1}^p (n-k) \right]. \quad (8)$$

Учитывая, что вероятность попадания нормально распределенной случайной величины в заданный интервал определяется по формуле [5]:

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - m}{\sigma}\right) - \Phi\left(\frac{x_1 - m}{\sigma}\right),$$

получим

$$P(LCL < |\Sigma| < UCL; |\Sigma| = [1 + d_0(t - t_0)]|\Sigma_0|) = \Phi\left(\frac{UCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right) - \Phi\left(\frac{LCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right),$$

где

$$UCL = m_{|\Sigma_0|} + u_{1-\frac{\alpha}{2}} \sigma_{|\Sigma_0|} = |\Sigma_0| \left[b_1 + u_{1-\frac{\alpha}{2}} \sqrt{b_2} \right];$$

$$LCL = m_{|\Sigma_0|} - u_{1-\frac{\alpha}{2}} \sigma_{|\Sigma_0|} = |\Sigma_0| \left[b_1 - u_{1-\frac{\alpha}{2}} \sqrt{b_2} \right];$$

$$m_{|\Sigma|} = b_1 |\Sigma| = b_1 (1 + d_0(t - t_0)) |\Sigma_0|;$$

$$\sigma_{|\Sigma|} = \sqrt{b_2} |\Sigma_0| = \sqrt{b_2} (1 + d_0(t - t_0)) |\Sigma_0|;$$

тогда

$$\begin{aligned}
 & P(LCL < |\Sigma| < UCL; |\Sigma| = 1 + d_0(t - t_0)|\Sigma_0|) = \\
 & = \Phi \left(\frac{|\Sigma_0| \left(b_1 + u_{1-\frac{\alpha}{2}} \sqrt{b_2} \right) - b_1(1 + d_0(t - t_0))|\Sigma_0|}{\sqrt{b_2}(1 + d_0(t - t_0))|\Sigma_0|} \right) - \Phi \left(\frac{|\Sigma_0| \left(b_1 - u_{1-\frac{\alpha}{2}} \sqrt{b_2} \right) - b_1(1 + d_0(t - t_0))|\Sigma_0|}{\sqrt{b_2}(1 + d_0(t - t_0))|\Sigma_0|} \right) = \\
 & = \Phi \left(\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} - \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right) - \Phi \left(-\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} - \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right).
 \end{aligned}$$

Отсюда

$$\begin{aligned}
 L(d_0) &= \frac{1}{1 - P(d)} = \\
 &= \frac{1}{\Phi \left(-\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} + \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right) + \Phi \left(-\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} - \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right)}
 \end{aligned}$$

или окончательно:

$$L(d_0) = \frac{1}{\Phi \left(-\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} + \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right) + \Phi \left(-\frac{u_{1-\frac{\alpha}{2}}}{1 + d_0(t - t_0)} - \frac{b_1(d_0(t - t_0))}{\sqrt{b_2}(1 + d_0(t - t_0))} \right)}. \quad (9)$$

Для проверки подставим значение $d_0 = 0$ (тренд отсутствует):

$$\begin{aligned}
 L(0) &= \frac{1}{2\Phi \left(-u_{1-\frac{\alpha}{2}} \right)} = \frac{1}{2 \left(1 - \Phi \left(u_{1-\frac{\alpha}{2}} \right) \right)} = \\
 &= \frac{1}{2 \left(1 - 1 + \frac{\alpha}{2} \right)} = \frac{1}{\alpha}.
 \end{aligned}$$

Например, при $\alpha = 0,0027$ (правило трех сигма), получим:

$$L(0) = \frac{1}{0,0027} \cong 370.$$

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ

Для проведения численного исследования рассчитаем коэффициенты b_1 и b_2 по формулам (7) и (8), подставим в формулу средней длины серий (9) и найдем значения $L(d)$ при некоторых значениях числа

контролируемых показателей p и различных объемах мгновенных выборок n . Пусть, например, количество контролируемых показателей $p = 2$, объем мгновенной выборки $n = 5$.

Коэффициенты:

$$b_1 = \frac{1}{(5-1)^2} (5-1)(5-2) = 0,75,$$

$$b_2 = \frac{1}{(5-1)^4} (5-1)(5-2) *$$

$$\begin{aligned}
 & * [(5-1+2)(5-2+2) - (5-1)(5-2)] = \\
 & = 0,84375.
 \end{aligned}$$

Пусть квантиль $u_{1-\frac{\alpha}{2}} = 3$. Теперь мы знаем значения

коэффициентов, квантиль известна, с помощью электронных таблиц Excel рассчитаем значения аргументов функции стандартного нормального распределения при $d = 0$ и $t - t_0 = 6$:

$$\begin{aligned}
 & -\frac{u_{1-\frac{\alpha}{2}}}{1+d_0(t-t_0)} + \frac{b_1(d_0(t-t_0))}{\sqrt{b_2(1+d_0(t-t_0))}} = \\
 & = -\frac{3}{1+0*6} + \frac{0,75*0*6}{\sqrt{0,84375(1+0*6)}} = -3; \\
 & -\frac{u_{1-\frac{\alpha}{2}}}{1+d_0(t-t_0)} - \frac{b_1(d_0(t-t_0))}{\sqrt{b_2(1+d_0(t-t_0))}} = \\
 & = -\frac{3}{1+0*6} - \frac{0,75*0*6}{\sqrt{0,84375(1+0*6)}} = -3.
 \end{aligned}$$

Рассмотрим значения d в диапазоне от 0 до 0,1 с шагом 0,01; с помощью встроенных функций Excel рассчитаем автоматически по аналогии остальные значения. Далее с помощью встроенной функции =НОРМСТРАСП(z ; интегральная) найдём значения функции стандартного нормального распределения:

$$\Phi_1 = \text{НОРМСТРАСП}(3; 1) = 0,00135;$$

$$\Phi_2 = \text{НОРМСТРАСП}(-3; 1) = 0,00135.$$

Конечный результат – средняя длина серий при $n = 5$ и $p = 2$:

$$L(d_0) = \frac{1}{0,00135 + 0,00135} = 370,4.$$

На рисунке показаны результаты расчета для выборки при различных значениях длительности тренда. По горизонтальной оси отложены значения d_0 , по вертикальной – средняя длина серий $L(d_0)$.

При слишком большом значении средней длины серий (когда нарушение тренда рассеяния обнаруживается недостаточно оперативно) возможны несколько раз-

личных методов повышения эффективности алгоритма обобщенной дисперсии [6–9]:

- анализ наличия неслучайных структур на карте обобщенной дисперсии,
- применение предупреждающей границы,
- использование алгоритма экспоненциально взвешенных скользящих средних для обобщенной дисперсии.

ЗАКЛЮЧЕНИЕ

Проведенное исследование нарушения процесса при наличии постепенного увеличения рассеяния (тренда) позволило установить зависимость для расчета средней длины серий алгоритма обобщенной дисперсии в зависимости от интенсивности и длительности тренда.

СПИСОК ЛИТЕРАТУРЫ

1. Клячкин В.Н., Крашенинников В.Р., Кувайскова Ю.Е. Прогнозирование и диагностика стабильности функционирования технических объектов : монография. М.: РУСАЙНС, 2020. 200 с.
2. Montgomery D.C. Introduction to statistical quality control. NewYork: John Wiley and Sons, 2009. 754 p.
3. Алексеева А.В., Клячкин В.Н. Методика проведения статистических испытаний при мониторинге многомерного рассеяния // Информатика, моделирование, автоматизация проектирования : сб. науч. тр. X Всерос. школы-семинара аспирантов, студентов и молодых ученых / под ред. А.Н. Афанасьева. Ульяновск, 2018. С. 49–51.
4. Овсянникова В.В. Анализ эффективности алгоритма обобщенной дисперсии // Информатика, моделирование, автоматизация проектирования. Международная научно-практическая конференция студентов, аспирантов и молодых ученых : сб. науч. тр. Ульяновск, 2021. С. 156–160.

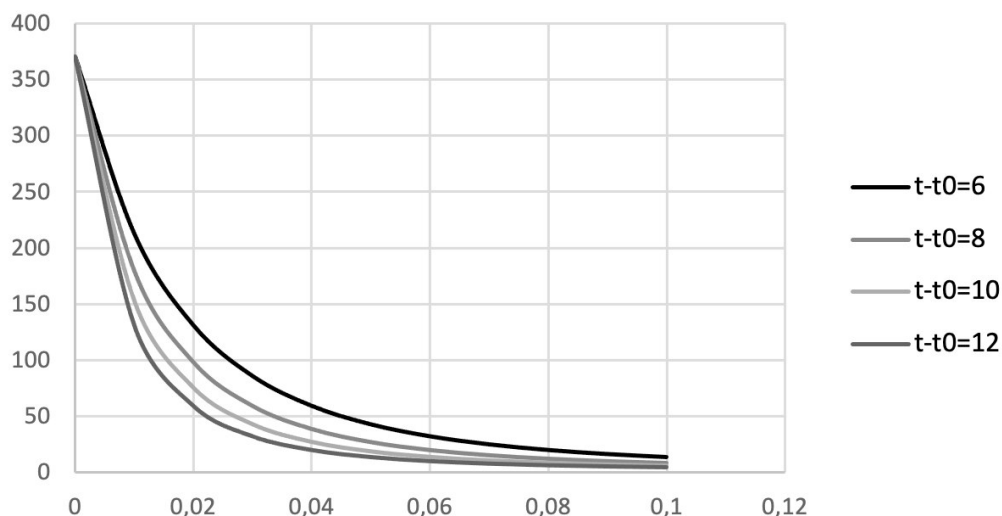


Рис. Средняя длина серий при наличии тренда

5. Клячкин В.Н., Кувайскова Ю.Е., Алексеева В.А. Статистические методы анализа данных. М. : Финансы и статистика, 2016. 240 с.

6. Клячкин В.Н., Алексеева А.В. Оценка стабильности вибрации на основе алгоритма обобщенной дисперсии // Известия Самарского научного центра Российской академии наук. 2018. Т. 20, № 4 (3). С. 491–493.

7. Алексеева А.В. Поиск структур специального вида на карте обобщенной дисперсии как метод повышения эффективности контроля // Вестник Ульяновского государственного технического университета. 2020. № 2–3 (90–91). С. 20–23.

8. Алексеева А.В. Повышение эффективности статистического контроля многомерного рассеяния процесса // Автоматизация процессов управления. 2020. № 3 (61). С. 101–107.

9. Клячкин В.Н., Алексеева А.В. Методы повышения эффективности алгоритма обобщенной дисперсии при контроле многомерного рассеяния // Метрологическое обеспечение инновационных технологий : матер. III Междунар. форума в рамках празднования 80-летия Санкт-Петербургского государственного университета аэрокосмического приборостроения, 300-летия Российской академии наук / под ред. В.В. Окрепилова. СПб., 2021. С. 135–136.

REFERENCES

1. Kliachkin V.N., Krashennnikov V.R., Kuvaiskova Iu.E. *Prognozirovanie i diagnostika stabilnosti funkcionirovaniia tekhnicheskikh obektov. Monografiia* [Forecasting and Diagnostics of Steady Operation of Technical Objects]. Moscow, Rusayns Publ., 2020. 200 p.

2. Montgomery D.C. *Introduction to Statistical Quality Control*. New York, John Wiley and Sons Press, 2009. 754 p.

3. Alekseeva A.V., Kliachkin V.N. Metodika provedeniia statisticheskikh ispytaniy pri monitoringe mnogomernogo rasseianiia [Statistical Test Procedures for Multidimensional Dissipation]. *Informatika, modelirovanie, avtomatizatsiia proektirovaniia. Sb. nauch. tr. 10 Vseros. shkoly-seminara aspirantov, studentov i molodykh uchenykh. Pod red. A.N. Afanaseva* [Proc. of the 10th Russian Workshop for Postgraduates, Students and Young Scientists on Informatics, Modeling, Computer-Aided Design. A.N. Afanasev (ed.)]. Ulyanovsk, 2018, pp. 49–51.

4. Ovsianikova V.V. Analiz effektivnosti algoritma obobshchennoi dispersii [Performance Analysis of Generalized Variance Algorithm]. *Informatika,*

modelirovanie, avtomatizatsiia proektirovaniia. Mezhdunarodnaia nauchno-prakticheskaia konferentsiia studentov, aspirantov i molodykh uchenykh. Sb. nauch. tr. [Proc. of the Int. Workshop for Postgraduates, Students and Young Scientists on Informatics, Modeling, Computer-Aided Design.] Ulyanovsk, 2021, pp. 156–160.

5. Kliachkin V.N., Kuvaiskova Iu.E., Alekseeva V.A. *Statisticheskie metody analiza dannykh* [Statistical Methods for Data Analysis]. Moscow, Finansy i Statistika Publ., 2016. 240 p.

6. Kliachkin V.N., Alekseeva A.V. Otsenka stabilnosti vibratsii na osnove algoritma obobshchennoi dispersii [Evaluation of Stability of Vibration Based on the Algorithm of Generalized Variance]. *Izvestiia Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk* [Proc. of Samara Federal Research Scientific Centre of the Russian Academy of Sciences], 2018, vol. 20, no. 4 (3), pp. 491–493.

7. Alekseeva A.V. Poisk struktur spetsialnogo vida na karte obobshchennoi dispersii kak metod povysheniia effektivnosti kontrolya [Search for Structures of a Special Type on the Map of Generalized Variance as a Method of Increasing Control Efficiency]. *Vestnik Ulyanovskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of Ulyanovsk State Technical University], 2020, no. 2–3 (90–91), pp. 20–23.

8. Alekseeva A.V. Povyshenie effektivnosti statisticheskogo kontrolya mnogomernogo rasseianiia protsessa [Improving the Efficiency of Statistical Control of Hydro-Unit Vibrations]. *Avtomatizatsiia protsessov upravleniia* [Automation of Control Processes], 2020, no. 3 (61), pp. 101–107.

9. Kliachkin V.N., Alekseeva A.V. Metody povysheniia effektivnosti algoritma obobshchennoi dispersii pri kontrole mnogomernogo rasseianiia [Methods for Increasing the Efficiency of the Generalized Variance Algorithm when Multidimensional Statistical Control]. *Metrologicheskoe obespechenie innovatsionnykh tekhnologii. Mater. III Mezhdunar. foruma v ramkakh prazdnovaniia 80-letii Sankt-Peterburgskogo gosudarstvennogo universiteta aerokosmicheskogo priboroostroeniia, 300-letii Rossiiskoi akademii nauk. Pod red. V.V. Okrepilova* [Proc. of the 3d Int. Forum Devoted to the SUAI's 80th Anniversary (St. Petersburg State University of Aerospace Instrumentation), the RAS' 300th Anniversary on Metrological Support of Innovative Technologies. V.V. Okrepilov (ed.)]. St. Petersburg, 2021, pp. 135–136.