УДК 658.562.012.7

А.А. Фролова

ПРИМЕНЕНИЕ АЛГОРИТМА ОБОБЩЕННОЙ ДИСПЕРСИИ ПРИ МНОГОМЕРНОМ СТАТИСТИЧЕСКОМ КОНТРОЛЕ ПРОЦЕССА ПО ИНДИВИДУАЛЬНЫМ НАБЛЮДЕНИЯМ

Фролова Анастасия Алексеевна, 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: anastasya19990916@mail.ru].

Аннотация

Часто на практике использование мгновенных выборок при статистическом контроле процесса невозможно по техническим условиям или нецелесообразно экономически. В этом случае ведется мониторинг процесса по индивидуальным наблюдениям. Для контроля среднего уровня многопараметрического процесса на базе алгоритма Хотеллинга разработаны соответствующие методы, однако при контроле многомерного рассеяния на основе алгоритма обобщенной дисперсии возникает ряд проблем. В статье предложены различные подходы к оценке обобщенной дисперсии, проводится их сравнительный анализ. Также рассматривается практическое применение предложенных алгоритмов.

Ключевые слова: обобщенная дисперсия, индивидуальные наблюдения, скользящий размах.

doi: 10.35752/1991-2927-2021-3-65-61-66

THE GENERALIZED VARIANCE ALGORITHM APPLICATION IN MULTIDIMENSIONAL STATISTICAL CONTROL OF THE PROCESS BASED ON INDIVIDUAL OBSERVATIONS

Frolova Anastasiia Alekseevna, graduated from the Faculty of Information Systems and Technologies of the Department of Applied Mathematics of Ulyanovsk State Technical University; an author of articles in the field of mathematical modeling in quality management problems. e-mail: anastasya19990916@mail.ru.

Abstract

In practice, the use of instantaneous samples for statistical control of the process is often impossible due to technical conditions or is not economically feasible. In this case, the process is monitored by individual observations. To control the average level of a multiparametric process based on the Hotelling algorithm, appropriate methods have been developed, but when controlling multidimensional scattering based on the generalized variance algorithm, a number of problems arise. Various approaches to the estimation of the generalized variance are proposed in the article, and their comparative analysis is carried out. The practical application of the proposed algorithms is also considered.

Keywords: generalized variance, individual observations, sliding range.

ВВЕДЕНИЕ

Для получения качественной продукции используются различные методы статистического контроля, в частности, статистическое управление процессами. Применение этих методов является одним из основных инструментов, обеспечивающих соответствие продукции техническим требованиям. При контроле множества коррелированных показателей используются многомерная карта Хотеллинга для мониторинга среднего уровня процесса и карта обобщенной дисперсии для

мониторинга рассеяния. Для нахождения обобщенной дисперсии необходимо найти элементы ковариационной матрицы, которые определяются по формуле:

$$s_{jkt} = \frac{1}{n-1} \sum (x_{ijt} - \bar{x}_j)(x_{ikt} - \bar{x}_k), \quad (1)$$

где x_{ijt} – результат i -го наблюдения по j -му показателю в t -й выборке ($i = 1, \dots, n$, n – объем t -й выборки, $j, k = 1, \dots, p$, p – количество контролируемых показателей, $t = 1, \dots, m$, m – количество выборок, взятых для

анализа процесса; $\bar{x}_j$ – среднее значение наблюдений по j -му показателю). Определитель $|S_t|$ матрицы (1) есть обобщенная дисперсия для t -й мгновенной выборки [1, 2].

Контрольные границы карты обобщенной дисперсии по Монтгомери [3]:

$$m_{|S|} \pm 3\sigma_{|S|}, \quad (2)$$

где математическое ожидание обобщенной дисперсии

$$m_{|S|} = b_1 |S|; \quad (3)$$

а ее стандартное отклонение

$$\sigma_{|S|} = \sqrt{b_2} |S|; \quad (4)$$

при этом для расчета коэффициентов

$$b_1 = \frac{1}{(n-1)^p} \prod_{j=1}^p (n-j); \quad (5)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \times \prod_{j=1}^p (n-j) \left[\prod_{k=1}^p (n-k+2) - \prod_{k=1}^p (n-k) \right]; \quad (6)$$

при этом необходимо выполнение условия

$$n \geq p+1. \quad (7)$$

В приведенных соотношениях предполагается, что для контроля многопараметрического процесса с p контролируемыми показателями используются мгновенные выборки (подгруппы) объемом n наблюдений.

Вместе с тем, часто на практике использование мгновенных выборок невозможно по техническим условиям или нецелесообразно экономически. В этом случае ведется мониторинг процесса по индивидуальным наблюдениям. В частности, при контроле процесса очистки питьевой воды на Санкт-Петербургском водоканале значения физико-химических показателей регистрируются лишь один раз в сутки [4], при контроле загрязненности сточных вод при производстве печатных плат в НПО «Марс» анализ стабильности химического состава элементов, попадающих в воду, проводится 2–3 раза в неделю [5].

В случае одномерного контроля независимых показателей широко применяются карты Шухарта по индивидуальным наблюдениям, при этом для оценки стандартного отклонения используется скользящий размах [6–8].

Цель настоящего исследования – разработка метода многомерного статистического контроля рассеяния процесса на основе алгоритма обобщенной дисперсии при использовании индивидуальных наблюдений.

Основные решаемые задачи для достижения цели:

- оценка влияния факторов на величину обобщенной дисперсии;
- разработка метода многомерного статистического контроля рассеяния процесса на основе алгоритма

обобщенной дисперсии при использовании индивидуальных наблюдений.

ПРОВЕДЕНИЕ ИСПЫТАНИЙ

Для проведения испытаний моделировались с применением генератора случайных чисел пакета «Анализ данных» электронных таблиц Excel [9] двумерные выборки по 100 наблюдений из нормально распределенных совокупностей с математическим ожиданием, равным нулю, и различными стандартными отклонениями ($\sigma = 1$ и $0,1$), а также с различными уровнями корреляции между показателями ($\rho = 0; 0,4; 0,6; 0,8$).

Для сравнения методов контроля по индивидуальным наблюдениям и наблюдениям с использованием мгновенных выборок вначале эти 100 наблюдений разбивались на 33 выборки по $n = 3$ наблюдения, 25 выборок по $n = 4$ наблюдения и 20 выборок по $n = 5$ наблюдений. Сравнивались соответствующие обобщенные дисперсии.

Результаты испытаний показаны в таблице 1, где $|S|$ – средняя обобщенная дисперсия, определяемая при контроле двух показателей по формуле:

$$|S| = \overline{S_{jkt}} * \overline{S_{kkt}} - \overline{S_{jkt}}^2, \quad (8)$$

где $\overline{S_{jkt}}, \overline{S_{kkt}}, \overline{S_{jkt}}$ – средние значения оценок элементов ковариационной матрицы S_t

$$s_{jkt} = \sigma_{jt}^2, s_{jkt} = \rho \sigma_{jt} \sigma_{kt}. \quad (9)$$

Видно, что при $\sigma = 1$ чем больше объем мгновенной выборки, тем больше обобщенная дисперсия $|S|$ (минимальное значение $|S|_3$ имеет место при $n = 3$), однако при $\sigma = 0,1$ это соотношение нарушается. Для $\sigma = 1$ вычислены Δ – абсолютное расхождение:

$$\Delta = |S|_3 - |S|_n,$$

и δ – относительное расхождение

$$\delta = \frac{\Delta}{|S|_3} 100\%,$$

по сравнению с $|S|_3$ при одинаковых корреляциях. Расхождение при $n = 3$ и $n = 5$ составило до 56%, что свидетельствует о существенном влиянии объема мгновенной выборки на величину обобщенной дисперсии при одинаковых обучающих выборках. Еще большее влияние на величину обобщенной дисперсии оказывает коррелированность показателей.

ОЦЕНКА ОБОБЩЕННОЙ ДИСПЕРСИИ ДЛЯ ИНДИВИДУАЛЬНЫХ НАБЛЮДЕНИЙ

Рассмотрим два варианта оценки. В качестве первого подхода, как и при контроле независимых показателей с применением карт Шухарта, воспользуемся скользящими размахами.

Таблица 1
Оценка обобщенной дисперсии
при использовании мгновенных выборок

Стандартное отклонение	Коэффициент корреляции	Объем выборки n	Средняя обобщенная дисперсия $ S $
1	0	3	0,79
		4	1,14
		5	1,2
	0,4	3	0,13
		4	0,18
		5	0,19
	0,6	3	0,28
		4	0,41
		5	0,43
	0,8	3	0,51
		4	0,73
		5	0,77
0,1	0	3	0,000108
		4	0,000108
		5	0,000091
	0,4	3	0,0000173
		4	0,0000172
		5	0,0000145
	0,6	3	0,0000388
		4	0,0000387
		5	0,0000326
	0,8	3	0,0000690
		4	0,0000689
		5	0,000058

Был предложен алгоритм нахождения средней обобщенной дисперсии через скользящий размах MR по следующим зависимостям:

$$MR_{jt} = |x_{jt} - x_{j,t-1}|, \quad (10)$$

$$MR_{j,t-1} = |x_{j,t-1} - x_{j,t-2}|, \quad (11)$$

тогда размах по двум выборкам (использование двух выборок обусловлено необходимостью выполнения соотношения (7) при $p = 2$: в этом случае для оценки обобщенной дисперсии будут использованы 3 наблюдения)

$$R_{jt} = \frac{MR_{jt} + MR_{j,t-1}}{2}, \quad j = 1, 2, \quad t = 3 \dots m. \quad (12)$$

Оценки элементов ковариационной матрицы S_t находятся по формуле (9), приближенное значение стандартного отклонения вычисляется по формуле:

$$\sigma_{jt} = R_{jt} / d; \quad (13)$$

табличный коэффициент $d = 1,128$, поскольку скользящие размахи вычисляются по двум соседним наблюдениям, а коэффициент корреляции ρ оценивается по всем m наблюдениям.

Второй подход к оценке обобщенной дисперсии основан на разделении сформированных выборок на две: первый столбец будет включать в себя первые 50 значений, второй – все остальные. Разобьем смоделированные выборки по 3, 4 и 5 наблюдений. Для нахождения $|S|$ для многомерного контроля значения оценок элементов ковариационной матрицы вычисляются по формулам:

$$S'_{jkt} = \frac{S_{jjt} + S_{kkt}}{2}, \quad S'_{jkt} = \rho \sqrt{S'_{jt}} \sqrt{S'_{kt}}. \quad (14)$$

Сравним значения $|S|$, полученные двумя способами. В таблице 2 через $|S|'$ обозначена средняя обобщенная дисперсия, найденная путем разбиения исходной выборки на две части, $|S|''$ – средняя обобщенная дисперсия, вычисленная через скользящий размах. Соответствующие расхождения определяются по формулам:

Таблица 2

Сравнение различных подходов для оценки обобщенной дисперсии
по индивидуальным наблюдениям

Стандартное отклонение	Коэффициент корреляции	$ S '$	$ S ''$	Δ	δ	$ S $	δ_1	δ_2
1	0	1,18	0,99	0,19	16	1,19	1	19
	0,4	0,17	0,15	0,02	12	0,17	0	12
	0,6	0,38	0,37	0,01	3	0,33	15	12
	0,8	0,69	0,68	0,01	1	1	31	32
0,1	0	0,000107	0,0000988	0,0000082	8	0,0001025	4	3,6
	0,4	0,0000185	0,0000241	0,0000056	30	0,0000187	1	29
	0,6	0,0000417	0,0000526	0,0000109	26	0,0000419	0,5	26
	0,8	0,000074	0,0000888	0,0000148	20	0,000074	0	20

$$\Delta = \left| |S'| - |S''| \right|, \delta = \frac{\Delta}{|S'|} 100\%,$$

$$|S| = \frac{\Delta}{\delta} 100\%, \delta_1 = \frac{\left| |S| - |S'| \right|}{|S|} 100\%,$$

$$\delta_2 = \frac{\left| |S| - |S''| \right|}{|S|} 100\%.$$

Представляет интерес сравнить найденные через индивидуальные наблюдения значения обобщенных дисперсий с обобщенной дисперсией всей обучающей выборки. При отсутствии нарушений процесса эти значения должны быть достаточно близки между собой.

Например, при $\sigma = 1$ и $\rho = 0$ имеем $|S| = 1,19$; расхождение $|S'|$ составляет $\delta = 1\%$, с $|S''|$ $\delta = 19\%$.

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ АЛГОРИТМОВ

Рассмотрим пример: контролировалось содержание алюминия и остаточного хлора в питьевой воде после водоочистки. Получено 100 наблюдений.

Построим карты обобщенных дисперсий по методу, основанному на разделении сформированных выборок

на две. На рисунках 1–2 представлены карты обобщенной дисперсии при объеме мгновенной выборки $n = 3$. Видно, что процесс находится в стабильности, так как нет выходов значений за границу.

Чтобы проконтролировать насколько правильными получатся значения средней обобщенной дисперсии для каждого наблюдения, необходимо найти оценки ковариационной матрицы для многомерного контроля по формуле (14). Находим среднюю обобщенную дисперсию совместно для алюминия и остаточного хлора по алгоритму Хотеллинга, через скользящий размах и по моему методу. На рисунке 3 представлены полученные карты при объеме мгновенной выборки $n = 3$.

Результаты испытаний представлены в таблице 3, где через $|S|$ обозначена средняя обобщенная дисперсия, найденная по алгоритму Хотеллинга, $|S'|$ – средняя обобщенная дисперсия, найденная путем разбиения исходной выборки на две части, $|S''|$ – средняя обобщенная дисперсия, вычисленная через скользящий размах. Соответствующие расхождения определяются по формулам:

$$\delta_1 = \frac{\left| |S| - |S'| \right|}{|S|} 100\%, \delta_2 = \frac{\left| |S| - |S''| \right|}{|S|} 100\%.$$

Рис. 1. Обобщенная дисперсия для алюминия

Рис. 2. Обобщенная дисперсия для остаточного хлора

Рис. 3а. Алгоритм Хотеллинга

Рис. 3б. Оценка обобщенной дисперсии через скользящий размах

Таблица 3

Результаты испытаний

N	$ S $	$ S' $	$ S'' $	δ_1	δ_2
3	0,0000012	0,00000125	0,00000063	4	48
4	0,0000012	0,00000136	-	13	-
5	0,0000018	0,00000181	-	1	-

Рис. 3в. Обобщенная дисперсия через разбиение – выборки на две

Видно, что максимальное расхождение для метода оценки обобщенной дисперсии через разбиение исходной выборки на две составило 13%. Для оценки обобщенной дисперсии через скользящий размах расхождение составило 48%, что указывает на необходимость доработки метода.

ЗАКЛЮЧЕНИЕ

Предложены подходы к оценке обобщенной дисперсии при контроле многомерного процесса по индивидуальным наблюдениям. Показано, что в зависимости от значений стандартных отклонений контролируемых показателей и корреляций между показателями точность оценки обобщенной дисперсии существенно различается, поэтому при исследовании процесса по обучающей выборке необходимо оценить соответствующую погрешность и выбрать наиболее точный подход. Было протестировано 2 метода оценки обобщенной дисперсии для индивидуальных наблюдений. Метод нахождения обобщенной дисперсии через скользящий размах требует доработки. Для метода оценки обобщенной дисперсии, который основан на разделении сформированных выборок на две, при многомерном контроле максимальное расхождение составило 13%. Наверное, можно сказать, что и для каждого наблюдения в отдельности оно будет примерно таким же.

СПИСОК ЛИТЕРАТУРЫ

1. Клячкин В.Н. Многомерный статистический контроль технологического процесса. М. : Финансы и статистика, 2003. 192 с.
2. Клячкин В.Н., Крашенинников В.Р., Кувайскова Ю.Е. Прогнозирование и диагностика стабильности функционирования технических объектов : монография. М. : РУСАЙНС, 2020. 200 с.
3. Montgomery D.C. Introduction to statistical quality control. New York: John Wiley and Sons, 2009. 754 p.
4. Клячкин В.Н., Барт А.Д. Оценка стабильности показателей качества при очистке питьевой воды // Экологические системы и приборы. 2017. № 8. С. 14–20.
5. Клячкин В.Н., Ширкунова К.С., Барт А.Д. Анализ стабильности химического состава сточных вод при производстве печатных плат // Экология и промышленность России. 2019. Т. 23 (5). С. 47–51.
6. Адлер Ю.П., Шпер В.Л. Практическое руководство по статистическому управлению процессами. М. : Альпина Паблишер, 2019. 234 с.
7. Уилер Д., Чамберс Д. Статистическое управление процессами. Оптимизация бизнеса с использованием контрольных карт Шухарта. М. : Альпина Бизнес Букс, 2009. 409 с.
8. Фролова А.А. Анализ погрешностей при оценке рассеяния процесса по индивидуальным наблюдениям // Информатика, моделирование, автоматизация проектирования. Международная научно-практическая конференция студентов, аспирантов и молодых ученых : сб. науч. тр. Ульяновск, 2021. С. 185–189.
9. Клячкин В.Н., Кувайскова Ю.Е., Алексеева В.А. Статистические методы анализа данных. М. : Финансы и статистика, 2016. 240 с.

REFERENCES

1. Kliachkin V.N. *Mnogomernyi statisticheskii kontrol tekhnologicheskogo protsessa* [Multivariate Statistical Process Control Charts]. Moscow, Finansy i statistika Publ., 2003. 192 p.
2. Kliachkin V.N., Krashennnikov V.R., Kuvaiskova Iu.E. *Prognozirovanie i diagnostika stabilnosti funktsionirovaniia tekhnicheskikh obektov. Monografiia* [Forecasting and Diagnostics of Consistency of Technical Object Operation. Monography]. Moscow, Rusayns Publ., 2020. 200 p.
3. Montgomery D.C. *Introduction to Statistical Quality Control*. New York, John Wiley and Sons Publ., 2009. 754 p.
4. Kliachkin V.N., Bart A.D. Otsenka stabilnosti pokazatelei kachestva pri oчитке pitevoi vody [Evaluation of Stability of Quality Parameters in Drinking Water Treatment]. *Ekologicheskie sistemy i pribory* [Ecological Systems and Devices], 2017, iss. 8, p. 14–20.
5. Kliachkin V.N., Shirkunova K.S., Bart A.D. Analiz stabilnosti khimicheskogo sostava stochnykh vod pri proizvodstve pechatnykh plat [Analysis of the Stability of the Chemical Composition of Wastewater in the Production of Printed Circuit Boards]. *Ekologiya i promyshlennost Rossii* [Ecology and Industry of Russia], 2019, vol. 23, iss. 5, pp. 47–51.
6. Adler Iu.P., Shper V.L. *Prakticheskoe rukovodstvo po statisticheskomu upravleniiu protsessami* [Practical Guidance on Statistical Process Control]. Moscow, Alpina Publ., 2019. 234 p.
7. Wheeler D., Chambers D. *Statisticheskoe upravlenie protsessami. Optimizatsiia biznesa s ispolzovaniem kontrolnykh kart Shukharta* [Understanding Statistical Process Control. Business Optimization with Shewhart Charts]. Moscow, Alpina Business Books Publ., 2009. 409 p.
8. Frolova A.A. Analiz pogreshnostei pri otsenke rasseianiia protsessa po individualnym nabludeniiam [Analysis of Errors when Evaluating Process Scattering based on Individual Observations]. *Informatika, modelirovanie, avtomatizatsiia proektirovaniia. Mezhdunarodnaia nauchno-prakticheskaiia konferentsiia studentov, aspirantov i molodykh uchenykh. Sb. nauch. tr.* [Proc. of Int. Sci. and Pract. Conference for Students, Postgraduates, and Young Scientists on Informatics, Modeling, Computer-Aided Design]. Ulyanovsk, 2021, pp. 185–189.
9. Kliachkin V.N., Kuvaiskova Iu.E., Alekseeva V.A. *Statisticheskie metody analiza dannykh* [Statistical Data Analysis Methods]. Moscow, Finansy i Statistika Publ., 2016. 240 p.