

УДК 681.518.5

ВЫБОР ПАРАМЕТРОВ АЛГОРИТМА ОБОБЩЕННОЙ ДИСПЕРСИИ ПРИ МНОГОМЕРНОМ СТАТИСТИЧЕСКОМ КОНТРОЛЕ РАССЕЙНИЯ ПРОЦЕССА

© 2020 А.В. Алексеева, В.Н. Клячкин

Ульяновский государственный технический университет, г. Ульяновск, Россия

Статья поступила в редакцию 10.12.2020

Для контроля стабильности функционирования агрегатов авиационной техники по результатам мониторинга группы показателей могут быть использованы методы статистического контроля процессов. При наличии значимых корреляций между показателями функционирования применяются многомерные методы. Контроль среднего уровня процесса при этом проводится на основе алгоритма Хотеллинга, контроль многомерного рассеяния – с помощью алгоритма обобщенной дисперсии. Если по условиям проведения процесса необходимо обеспечить наискорейшее обнаружение нарушения, то решается оптимизационная задача поиска таких значений объема выборки, частоты взятия выборок и положения контрольных границ, которые минимизируют среднее время нестабильного состояния процесса. Исходными данными являются количество контролируемых показателей процесса, целевое значение обобщенной дисперсии (оценивается по опытным данным), характеристика допустимого увеличения рассеяния, интенсивность нарушений процесса (параметр распределения Пуассона); время поиска нарушения после его обнаружения и время на расчет элемента выборки.

Ключевые слова: многомерный статистический контроль процесса, обобщенная дисперсия, время нестабильного состояния процесса

DOI: 10.37313/1990-5378-2021-23-1-79-83

Работа выполнена при поддержке гранта РФФИ и Правительства Ульяновской области, проект 18-48-730001

Для контроля стабильности функционирования агрегатов авиационной техники по результатам мониторинга группы показателей могут быть использованы методы статистического контроля процессов. При наличии значимых корреляций между показателями функционирования применяются многомерные методы. Контроль среднего уровня процесса при этом проводится на основе алгоритма Хотеллинга, контроль многомерного рассеяния – с помощью алгоритма обобщенной дисперсии [1-4]. Под обобщенной дисперсией понимают определитель ковариационной матрицы процесса [5-6].

В качестве примера на рис. 1 показаны соответствующие контрольные карты, используемые для оценки работоспособности агрегата по критерию стабильности вибраций, построенные в системе Statistica [7-8]. Контролировались три коррелированных показателя вибраций; измерения проводились с использованием мгновенных выборок по 5 наблюдений. Показаны результаты мониторинга по 20 выборкам. Про-

Алексеева Анастасия Валерьевна, аспирант.

E-mail: age-89@mail.ru

Клячкин Владимир Николаевич, доктор технических наук, профессор кафедры «Прикладная математика и информатика». E-mail: v_kl@mail.ru

цесс стабилен, поскольку ни на одной из карт нет выходов за контрольные границы (на обеих картах – только верхние границы, для карты Хотеллинга соответствующее значение составляет 16,75; для карты обобщенной дисперсии – 2624).

В общем случае предположим, что в процессе контролируется множество p показателей функционирования агрегата X_1, X_2, \dots, X_p . Через определенные (как правило, равные) промежутки времени берутся выборки из n наблюдений (n – объем мгновенной выборки, $n = 2 \dots 10$).

С целью оценки параметров контроля всего берется m выборок. По общей (обучающей) выборке mn наблюдений оценивается вектор средних $\mu_0 = (\mu_1, \mu_2, \dots, \mu_p)^T$ и ковариационная матрица Σ .

Для t -й выборки вектор средних

$$\mu_{jt} = \frac{1}{mn} \sum_{t=1}^m \sum_{i=1}^n x_{ijt}, \quad j = 1 \dots p, \quad t = 1 \dots m,$$

тогда оценка компонент общего вектора средних:

$$\mu_j = \frac{1}{m} \sum_{t=1}^m \mu_{jt}, \quad j = 1 \dots p, \quad (1)$$

(x_{ijt} – результат i -го наблюдения в t -й выборке по j -му параметру; $i = 1 \dots n$; $t = 1 \dots m$);

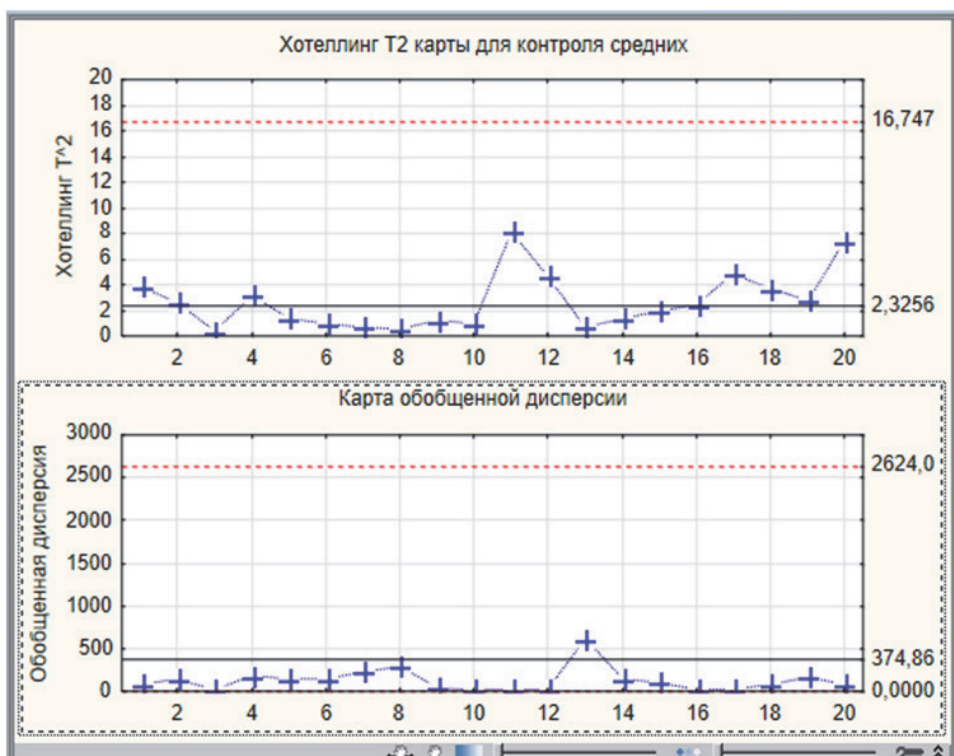


Рис. 1. Многомерные контрольные карты

оценки элементов ковариационной матрицы S:

$$s_{jkt} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijt} - \mu_{jt})(x_{ikt} - \mu_{kt}),$$

$$j = 1 \dots p, \quad k = 1 \dots p, \quad (2)$$

по которым строится ковариационная матрица S_t для t-й мгновенной выборки размерности $p \times p$; ее определитель (обобщенная дисперсия t-й мгновенной выборки) откладывается на карте. Также вычисляется оценка средней ковариации по j-му показателю

$$\bar{s}_{jk} = \frac{1}{m} \sum_{t=1}^m s_{jkt}, \quad (3)$$

по этим значениям строится ковариационная матрица размерности $p \times p$ и ее определитель $|S|$, который является оценкой целевой обобщенной дисперсии $|\Sigma_0|$.

По существу, алгоритм обобщенной дисперсии сводится к проверке нулевой гипотезы H_0 о том, что ковариационная матрица процесса соответствует заданному целевому значению: $H_0: |S| = |\Sigma_0|$.

Контрольные границы карты обобщенной дисперсии в предположении приближенной нормальности распределения [4] – это границы доверительного интервала для математического ожидания обобщенной дисперсии:

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

где $u_{1-\alpha/2}$ – квантиль нормального распределения порядка $1 - \alpha/2$ (α – уровень значимости), математическое ожидание обобщенной дисперсии

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

стандартное отклонение

$$s_{|s|} = \sqrt{b_2} |\Sigma_0|; \quad (6)$$

коэффициенты

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

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

тогда контрольные границы

$$\left. \begin{matrix} UCL \\ LCL \end{matrix} \right\} = |\Sigma_0| (b_1 \pm u_{1-\alpha/2} \sqrt{b_2}). \quad (9)$$

При отрицательном результате расчета по формуле (9) для нижней контрольной границы принимается $LCL = 0$.

Карта обобщенной дисперсии не всегда достаточно эффективна для обнаружения нарушения процесса. Возможны различные модификации алгоритма с целью повышения его чувствительности. Один из способов – использование алгоритма экспоненциально взвешенных скользящих средних для обобщенной дисперсии: исследования показывают, что во многих случаях такой подход эффективнее обычной карты обобщенной дисперсии.

Другой вариант – применение предупреждающей границы на карте обобщенной дисперсии: в этом случае процесс считается нарушенным не только, если точка на карте выйдет за контрольную границу, но и в том случае, когда

несколько точек подряд попали между предупреждающей и контрольной границами.

Возможен и поиск структур специального вида на карте обобщенной дисперсии (вероятность появления таких структур на карте при нормальном ходе процесса мала – соизмерима с вероятностью ложной тревоги); например, приближение определенного количества точек к контрольной границе свидетельствует о нарушении процесса, наличие нескольких точек подряд на убывание или возрастание показывает наличие тренда процесса, и т.п. [7].

Поскольку обобщенная дисперсия не может полностью характеризовать многомерное рассеяние (при существенно различных ковариационных матрицах их определители могут оказаться одинаковыми), иногда целесообразно и одновременное использование карт для одномерного рассеяния по всем контролируемым показателям.

Эффективность метода контроля определяется как скоростью обнаружения нарушений процесса, так и затратами на контроль. Для повышения эффективности контроля можно подобрать соответствующий объем мгновенной выборки, интервал между выборками и положение контрольных границ.

При больших затратах на взятие выборки берется минимальный ее объем. Если же технологический процесс допускает проведение большого количества измерений, то объем выборки может быть увеличен, при этом повышается точность оценки настройки процесса.

Интервал между взятием выборок также зависит от возможностей производства и стоимости отбора выборки, но он должен быть гораздо меньше, чем интервал между нарушениями. Таким образом, у более стабильного процесса интервал между выборками можно увеличить.

Эффективность контроля также зависит от положения контрольных границ. Если доверительный интервал установить слишком узким, то увеличится количество ложных тревог с необоснованной остановкой для диагностики процесса, что увеличивает стоимость контроля. Если интервал слишком широк, увеличивается вероятность необоснованного пропуска нарушений процесса, что может привести к выпуску дефектной продукции и увеличению стоимости

на исправление брака (или к пропуску неработоспособного состояния объекта).

В рассмотренном испытании (рис. 1) приняты значения интервала между выборками $h = 30$ секунд, объем мгновенной выборки $n = 5$, в соответствии с формулами (1)-(9) верхняя контрольная граница $UCL = 2624$, нижняя – $LCL = 0$ (уровень значимости в соответствии с правилом три сигма принят равным $\alpha = 0,0027$; при этом квантиль $u_{1-\alpha/2} = 3$).

Для расчета оптимальных параметров карт Шухарта А.Дункан предложил весь производственный цикл (рис. 2) разделить на четыре интервала [9]:

- а) интервал, в течении которого процесс стабилен;
- б) интервал, в котором нарушение процесса произошло, но еще не обнаружено;
- в) интервал времени на отбор выборки и обнаружение нарушения;
- г) время, необходимое на устранение нарушения.

Этот же подход может быть использован и при оптимизации параметров алгоритма обобщенной дисперсии. Предполагается, что время возникновения нарушений подчиняется закону распределения Пуассона с параметром λ . Прогнозируемое время появления нарушения между выборками согласно Дункану определяется по формуле:

$$\tau = \frac{1 - (1 + \lambda h)e^{-\lambda h}}{\lambda(1 - e^{-\lambda h})}, \quad (10)$$

где h интервал между выборками.

Интервал времени, в течение которого процесс стабилен, определяется как математическое ожидание $1/\lambda$. Интервал времени от момента нарушения до устранения этого нарушения равно $h/(1 - \beta) - \tau$, где β – вероятность ошибки второго рода (процесс стабилен, но карта показывает его нарушение). Время на расчет элемента выборки предполагается константой g , тогда интервал времени на отбор выборки и обнаружение нарушения равен ng . Время поиска нарушения после его обнаружения также является константой D [10].

Таким образом, средняя продолжительность цикла регулировки, согласно Дункану, может быть представлена в виде



Рис. 2. Цикл обнаружения и устранения нарушения процесса

$$M(T) = 1/\lambda + h/(1 - \beta) - \tau + ng + D.$$

Тогда средняя длительность нестабильного состояния процесса

$$M(T_0) = h/(1 - \beta) - \tau + ng + D. \quad (11)$$

Если по условиям проведения процесса необходимо обеспечить наискорейшее обнаружение нарушения (нарушение состоит в увеличении обобщенной дисперсии в d раз), то решается оптимизационная задача поиска таких значений объема выборки n , частоты взятия выборок h и положения контрольных границ UCL и LCL, которые минимизируют значение функции (11).

Исходные данные: p – количество контролируемых показателей процесса, $|\Sigma_0|$ – целевое значение обобщенной дисперсии (оценивается по опытным данным), d – характеристика увеличения рассеяния: фактическое значение обобщенной дисперсии в d раз превышает целевое: $|\Sigma| = d|\Sigma_0|$, α_0 – максимально допустимая вероятность ошибки первого рода (процесс стабилен, но карта показывает нарушение), β_0 – максимально допустимая вероятность ошибки второго рода (процесс нестабилен, но карта не показывает нарушение: риск незамеченной разладки), λ – параметр распределения Пуассона – интенсивность нарушений процесса; D – время поиска нарушения после его обнаружения, g – время на расчет элемента выборки.

Требуется найти значения: n – объем мгновенной выборки, h – промежуток времени между выборками, $h > ng$; положение контрольных границ UCL и LCL, а и β – вероятности ошибок 1-го и 2-го рода (характеризующие положение контрольных границ; $\alpha \leq \alpha_0$, $\beta \leq \beta_0$), минимизирующие среднее время поиска и устранения нарушения процесса.

Вероятность ошибки второго рода, или риск незамеченной разладки (обобщенная дисперсия увеличилась в d раз, при этом точки на карте оказались внутри контрольных границ, т.е. карта не обнаружила увеличение рассеяния) определяется по формуле

$$\begin{aligned} \beta &= P(LCL < |\Sigma| < UCL; |\Sigma| = d|\Sigma_0|) = \\ &= \Phi\left(\frac{UCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right) - \Phi\left(\frac{LCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right) \\ &= \Phi\left(\frac{UCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right) - \Phi\left(\frac{LCL - m_{|\Sigma|}}{\sigma_{|\Sigma|}}\right), \end{aligned}$$

здесь

$$UCL, LCL = |\Sigma_0|(b_1 \pm u_{1-\alpha/2}\sqrt{b_2});$$

$$m_{|\Sigma|} = b_1|\Sigma| = b_1d|\Sigma_0|;$$

$$\sigma_{|\Sigma|} = \sqrt{b_2}|\Sigma| = \sqrt{b_2}d|\Sigma_0|;$$

тогда

$$\begin{aligned} \beta &= \Phi\left(\frac{-b_1 d - 1}{\sqrt{b_2} d} + \frac{u_{1-\alpha/2}}{d}\right) - \\ &- \Phi\left(\frac{-b_1 d - 1}{\sqrt{b_2} d} - \frac{u_{1-\alpha/2}}{d}\right). \quad (12) \end{aligned}$$

При этом могут быть дополнительно заданы ограничения $n_{min} \leq n \leq n_{max}$ и $h_{min} \leq h \leq h_{max}$, где минимальные и максимальные значения объема выборки и интервала между выборками при необходимости оцениваются специалистом по конкретному процессу.

Предполагая, например, опасным увеличение обобщенной дисперсии в 1,75 раз ($d = 1,75$), интервал между выборками $h = 1$, интенсивность нарушений $\lambda = 0,1$, а значения $g = 0,12$ и $D = 0,08$, найдем среднее время поиска и устранения нарушения процесса $M(T_0) = 4,48$. Варьируя параметры n , h , UCL, можно минимизировать эту величину. Эта задача решается методами оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. Клячкин В.Н., Крашенинников В.Р., Кувайскова Ю.Е. Прогнозирование и диагностика стабильности функционирования технических объектов: монография. М.: РУСАЙНС, 2020 г. 200 с.
2. Уилер Д., Чамберс Д. Статистическое управление процессами. Оптимизация бизнеса с использованием контрольных карт Шухарта. М.: Альпина Бизнес Букс, 2009. 409 с.
3. Ryan T.P. Statistical methods for quality improvement. New York: John Wiley and Sons, 2011. 687 p.
4. Montgomery D.C. Introduction to statistical quality control. New York: John Wiley and Sons, 2009. 754 p.
5. Клячкин В.Н., Алексеева А.В. Оценка стабильности вибраций на основе алгоритма обобщенной дисперсии // Известия Самарского научного центра РАН. 2018. Т. 20. № 4-3 (84). С. 491-493.
6. Клячкин В.Н., Алексеева А.В. Оценка стабильности функционирования гидроагрегата по результатам мониторинга вибраций // Приборы и системы. Управление, контроль, диагностика. 2019. № 3. С. 11-18.
7. Алексеева А.В. Повышение эффективности статистического контроля многомерного рассеяния процесса // Автоматизация процессов управления. 2020. №3(61). С. 101-107.
8. Боровиков В. STATISTICA: искусство анализа данных. Для профессионалов. СПб.: Питер, 2001. 656 с.
9. Duncan, A.J. The economic design of x-chart used to maintain current control of the process // Journal of the American Statistical Association, 1956. Vol.51. P.228-242.
10. Costa F.B., Claro A.E. Economic design of control charts for monitoring a first order autoregressive process // Brazilian journal of operations & production management, 2009. Vol.6. №2. P.07-26.

11. *Fong-Jung Y., Ching-Shih T., Kai-I H., Zhang W.* An economic-statistical design of control chart with multiple assignable causes // *Journal of quality*, 2010. Vol.17. №4. P.327-338.
12. *Алексеева А.В.* Программа моделирования выборок для оценки эффективности контроля вибраций гидроагрегата // *Информатика, моделирование, автоматизация проектирования* : сб. тр. X Всерос. школы-семинара аспирантов, студентов и молодых ученых. Ульяновск : УлГТУ, 2018. С. 46–48.
13. *Bersimis S., Psarakis S., Panaretos J.* Multivariate Statistical Process Control Charts: An Overview // *Quality and reliability Engineering International*. 2007. V. 23. pp. 517–543.
14. *García-Díaz, J. Carlos.* The ‘effective variance’ control chart for monitoring the dispersion process with missing data // *Industrial Engineering*. 2007. Vol. 1. № 1. Pp. 40–45.

SELECTION OF THE PARAMETERS OF THE GENERALIZED DISPERSION ALGORITHM FOR MULTIVARIATE STATISTICAL CONTROL OF THE PROCESS SCATTERING

© 2021 A.V. Alekseeva, V.N. Klyachkin

Ulyanovsk State Technical University, Ulyanovsk, Russia

To control the stability of the functioning of aviation equipment units based on the results of monitoring a group of indicators, methods of statistical processes control can be used. In the presence of significant correlations between performance indicators, multivariate methods are used. In this case, the control of the average level of the process is carried out on the basis of the Hotelling algorithm, the control of multivariate scattering is carried out using the generalized variance algorithm. If, according to the conditions of the process, it is necessary to ensure the fastest detection of a violation, then the optimization problem of finding such values of the sample size, sampling frequency and position of the control boundaries is solved that minimizes the average time of the unstable state of the process. The initial data are the number of process indicators monitored, the target value of the generalized variance (estimated from experimental data), the characteristic of the permissible increase in scattering, the intensity of process disturbances (parameter of the Poisson distribution); time to search for a violation after its detection and time to calculate the sample element.

Keywords: multivariate statistical process control, generalized variance, time of unstable process state.

DOI: 10.37313/1990-5378-2021-23-1-79-83

Anastasia Alekseeva, Postgraduate Student.

E-mail: age-89@mail.ru.

Vladimir Klyachkin, Doctor of Technics, Professor at the Applied mathematics and Informatics Department.

E-mail: v_kl@mail.ru