УДК 658.562.012.7

В.Н. Клячкин, А.В. Алексеева

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ АЛГОРИТМА ОБОБЩЕННОЙ ДИСПЕРСИИ ПРИ СТАТИСТИЧЕСКОМ КОНТРОЛЕ ПРОЦЕССА

Клячкин Владимир Николаевич, доктор технических наук, профессор, окончил механический факультет Ульяновского политехнического института. Профессор кафедры «Прикладная математика и информатика» Ульяновского государственного технического университета. Имеет научные труды в области надежности и статистических методов. [e-mail: v_kl@mail.ru].

Алексеева Анастасия Валерьевна, окончила факультет информационных систем и технологий УлГТУ. Аспирант кафедры «Прикладная математика и информатика» УлГТУ, инженер по стандартизации АО «Ульяновское конструкторское бюро приборостроения». Имеет научные труды в области статистических методов в управлении качеством. [e-mail: age-89@mail.ru].

Аннотация

При контроле реального производственного процесса с применением статистических методов встает вопрос скорейшего обнаружения нарушений. В большинстве случаев в производственном процессе контролируется несколько показателей одновременно, причем изменение величин одних показателей приводит к изменению других. При наличии зависимости показателей для их контроля используют средства многомерного статистического контроля, в частности карты обобщенной дисперсии. Варьируя параметры карты, можно значительно повысить ее эффективность, таким образом минимизировав время нахождения процесса в нестабильном состоянии.

Применив подход А. Дункана, который он разработал для карт Шухарта, была получена формула математического ожидания длительности нестабильного состояния процесса и разработана программа на языке Python для ее минимизации.

Для апробации поставленной оптимизационной задачи приведен расчет данных двух показателей технологического процесса и получены оптимальные параметры карты обобщенной дисперсии, при которых длительность нахождения процесса в нестабильном состоянии минимальна.

Ключевые слова: стабильность рассеяния, обобщенная дисперсия, длительность нестабильного состояния процесса, задача оптимизации.

doi: 10.35752/1991-2927-2021-3-65-41-47

OPTIMIZATION OF PARAMETERS OF GENERALIZED DISPERSION ALGORITHM AT STATISTICAL PROCESS CONTROL

Vladimir Nikolaevich Kliachkin, Doctor of Sciences in Engineering; Professor; graduated from the Mechanical Faculty of Ulyanovsk Polytechnic Institute; Professor at the Department of Applied Mathematics and Informatics of Ulyanovsk State Technical; an author of scientific works in the field of reliability and statistical methods. e-mail: v_kl@mail.ru.

Anastasiia Valerevna Alekseeva, graduated from the Faculty of Information System and Technologies of Ulyanovsk State Technical University; Postgraduate Student at the Department of Applied Mathematics and Informatics of UISTU, Standardization Engineer at the Ulyanovsk Design Bureau of Instrumentation; an author of articles in the field of the statistical methods in quality control. e-mail: age-89@mail.ru.

Abstract

When monitoring a real production process using statistical methods, the question of early detection of violations arises. In most cases, several indicators are monitored simultaneously in the production process, and a change in the values of some indicators leads to a change in others. If there is a dependence of indicators for their monitoring, multivariate statistical control tools are used, in particular generalized variance chart. By varying the parameters of the chart, its efficiency can be significantly increased, this allows minimizing the time the process is in an unstable state.

Applying the approach of A. Duncan, which he developed for Shewhart charts, a formula for the expectation of the duration of an unstable state of a process was obtained and a Python program was developed to minimize it.

To test the set optimization problem, the calculation of the data of two process indicators is given and the optimal parameters of the generalized variance chart are obtained, at which the duration of the process in an unstable state is minimal.

Keywords: scattering stability, generalized variance, duration of unstable process state, optimization task.

Постановка задачи

При статистическом контроле процесса, как правило, проводится мониторинг стабильности среднего уровня процесса и его рассеяния. Для независимых показателей используются стандартные карты Шухарта (например, карты средних и стандартных отклонений), при наличии корреляций между показателями применяются многомерные методы: алгоритм Хотеллинга для мониторинга среднего уровня процесса и алгоритм обобщенной дисперсии для контроля многомерного рассеяния [1–3].

Под обобщенной дисперсией понимают определитель ковариационной матрицы. По опытным данным для каждой мгновенной выборки объемом n наблюдений в момент времени t рассчитывается выборочная ковариационная матрица S_t . По аналогии с картами Шухарта определяется положение границ (это границы доверительного интервала для обобщенной дисперсии) и средней линии. По полученным данным строится контрольная карта.

В качестве примера на рисунке 1 показана карта обобщенной дисперсии, построенная в системе Statistica [4, 5] для вибраций гидроагрегата по коррелированным показаниям двух вибродатчиков (здесь положение верхней контрольной границы *Upper Control Limit* – $UCL = 0,01523$; нижней $LCL = 0$; центральной линии $0,00198$): нарушение стабильности процесса в 22-й выборке.

Эффективность статистического контроля определяется скоростью обнаружения нарушений процесса, а также затратами на контроль. Для повышения эффективности контроля можно подобрать соответствующие

параметры алгоритма: объем мгновенной выборки n , интервал между выборками h и положение контрольных границ UCL и LCL [6].

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

Интервал между взятием выборок h также зависит от возможностей производства и стоимости отбора выборки: при более стабильном процессе интервал между выборками может быть увеличен.

Положение контрольных границ определяется соответствующим доверительным интервалом: при слишком узком интервале увеличится количество ложных тревог с необоснованной остановкой для диагностики процесса, при слишком широком интервале увеличивается вероятность необоснованного пропуска нарушения процесса [7–9].

Цель настоящего исследования – разработка технологии оптимизации параметров алгоритма обобщенной дисперсии при статистическом контроле многомерного рассеяния процесса по критерию наискорейшего обнаружения нарушения.

ОСНОВНЫЕ ЗАВИСИМОСТИ

Для оптимизации параметров контрольных карт Шухарта А. Дункан предложил весь производственный цикл разделить на четыре интервала [10]:

- интервал, в течение которого процесс стабилен;
- интервал, в котором нарушение процесса произошло, но еще не обнаружено;

Рис. 1. Карта обобщенной дисперсии

- интервал времени на отбор выборки и обнаружение нарушения;

- время, необходимое на устранение нарушения.

Предполагается, что время возникновения нарушений подчиняется закону распределения Пуассона с параметром λ . Интервал времени, в течение которого процесс стабилен, определяется как математическое ожидание $1/\lambda$. Прогнозируемое время появления нарушения между выборками согласно Дункану определяется по формуле:

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

где h – интервал между выборками.

Аналогичный подход может быть использован и при оптимизации параметров алгоритма обобщенной дисперсии. Наиболее распространенным нарушением является скачкообразное увеличение рассеяния, при этом обобщенная дисперсия – определитель ковариационной матрицы процесса Σ увеличивается в d раз по сравнению с определителем исходной (целевой) матрицы процесса в стабильном состоянии Σ_0 :

$$|\Sigma| = d|\Sigma_0|. \quad (2)$$

Интервал времени от момента нарушения до устранения этого нарушения равен $h/(1 - \beta) - \tau$, где β – вероятность ошибки второго рода (процесс нестабилен, но карта не показывает его нарушения).

Предположим, что время на расчет элемента выборки является константой g , тогда интервал времени на отбор выборки и обнаружение нарушения равен ng . Время поиска нарушения после его обнаружения также можно считать постоянным и равным некоторому D [11, 12].

Тогда математическое ожидание длительности нестабильного состояния процесса

$$M(T_0) = \frac{h}{1 - \beta} - \frac{1 - (1 + \lambda h)e^{-\lambda h}}{\lambda(1 - e^{-\lambda h})} + ng + D \quad (3)$$

может рассматриваться как целевая функция, которую необходимо минимизировать.

Вероятность ошибки второго рода (риск незамеченной разладки: обобщенная дисперсия увеличилась, однако при этом точки на карте оказались внутри контрольных границ) определится по формуле:

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

Здесь контрольные границы карты – это границы доверительного интервала для математического ожидания обобщенной дисперсии в предположении приближенной нормальности распределения [1]:

$$m_{|\Sigma_0|} \pm u_{1-\alpha/2} \sigma_{|\Sigma_0|},$$

где $u_{1-\alpha/2}$ – квантиль нормального распределения порядка $1 - \alpha/2$ (α – уровень значимости или вероятность ложной тревоги: вероятность того, что карта ошибочно покажет нарушение процесса), математическое ожидание обобщенной дисперсии

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

стандартное отклонение

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

коэффициенты

$$b_1 = \frac{1}{(n-1)^p} \prod_{j=1}^p (n-j); \quad (4)$$

$$\begin{aligned} b_2 &= \frac{1}{(n-1)^{2p}} \times \\ &\times \prod_{j=1}^p (n-j) \left[\prod_{k=1}^p (n-k+2) - \prod_{k=1}^p (n-k) \right], \end{aligned} \quad (5)$$

тогда контрольные границы

$$\left. \begin{aligned} UCL \\ LCL \end{aligned} \right\} = |\Sigma_0| \left(b_1 \pm u_{1-\alpha/2} \sqrt{b_2} \right). \quad (6)$$

При отрицательном результате расчета по формуле (6) для нижней контрольной границы принимается $LCL = 0$.

Если обобщенная дисперсия увеличилась в d раз, то

$$m_{|\Sigma|} = db_1 |\Sigma_0|,$$

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

После несложных преобразований получим

$$\begin{aligned} \beta &= \Phi\left[\frac{u_{1-\alpha/2}}{d} - \frac{b_1(d-1)}{d\sqrt{b_2}}\right] - \\ &- \Phi\left[-\frac{u_{1-\alpha/2}}{d} - \frac{b_1(d-1)}{d\sqrt{b_2}}\right]. \end{aligned} \quad (7)$$

Формулировка оптимизационной задачи:

- зная среднее время на расчет элемента выборки g и среднее время поиска нарушения после его обнаружения D ,

- определив по результатам предыдущих испытаний среднее время $1/\lambda$ нахождения конкретного процесса в стабильном состоянии, а также допустимое увеличение обобщенной дисперсии d ,

- вычислив по заданной обучающей выборке значение целевой обобщенной дисперсии $|\Sigma_0|$,

- найти значения параметров алгоритма обобщенной дисперсии: объема мгновенной выборки n ,

интервала между выборками h и положения контрольных границ UCL и LCL из условия минимума математического ожидания длительности нестабильного состояния процесса (3) с учетом соотношений (4)–(7).

При этом дополнительно могут быть 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 оцениваются специалистом по конкретному процессу. Также задается диапазон допустимых значений вероятности ложной тревоги $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$.

ПРОГРАММА РАСЧЕТА

Для решения оптимизационной задачи была написана программа на языке Python в среде PyCharm. Программа использует библиотеку с открытым исходным кодом `scipy`, предназначенную для научных и инженерных расчетов, из которой подключены модули `optimize` и `stats`. Модуль `optimize` обеспечивает нахождения как условного, так и безусловного экстремума функции.

Для поиска условного минимума используется функция:

$$\text{minimize}(f, x0, \text{method} = 'trust - \text{constr}', \text{bounds} = \text{bnds}), \quad (8)$$

где f – целевая функция, $x0$ – вектор начальных условий (исходные значения h , n и α), `method` – выбранный метод оптимизации, `bounds` – область поиска минимума функции. Границы поиска задаются вектором

$$\text{bnds} = \text{Bounds}((h_{\min}, n_{\min}, \alpha_{\min}), (h_{\max}, n_{\max}, \alpha_{\max})).$$

Для корректной работы программы при заданных параметрах было реализовано последовательное нахождение минимума функции тремя методами (в формуле (8) указан первый из них):

1. Поиск локального минимума в доверительной области (`trust-constr`). Доверительная область – это аналог доверительного интервала в случае многопараметрической целевой функции, которая аппроксимируется числовой функцией. При этом вначале выбирают ширину доверительной области, а потом направление. Чтобы метод обеспечивал сходимость с заданной точностью, необходимо чтобы начальное приближение было не слишком далеко от искомого оптимального решения. С этой целью на каждый шаг итерации накладываются ограничения, которые предотвращают «далекую» аппроксимацию.

2. Метод `Broyden-Fletcher-Goldfarb-Shanno` (`l-bfgs-b`). Этот метод относится к итерационным квази-ньютоновским методам численной оптимизации. Применение данного метода не предполагает вычисления частных производных второго порядка, т. к. нет необходимости в определении гессиана функции напрямую. Гессиан определяется приближенно, исходя из

предыдущих итераций, чем и отличается данный метод от ньютоновских. В разработанной программе используется модификация данного метода с ограниченным использованием памяти в многомерном кубе. Данный метод устойчив и обладает сверхлинейной сходимостью.

3. Метод Нелдера-Мида. Метод является надежным и эффективным методом прямого поиска, т. е. использующим только значения функции. Он относится к симплексным методам. Суть метода заключается в вычислении значений функции в вершинах многогранника и последовательном итерационном перемещении симплекса в оптимальном направлении. Первоначально регулярный симплекс применялся на каждой итерации. Нелдер и Мид модифицировали метод путем использования неправильных симплексов. При этом перемещение симплекса осуществляется с помощью процедур отражения, растяжения и сжатия.

Упрощенная блок-схема алгоритма программы представлена на рисунке 2.

Рис. 2. Блок-схема алгоритма программы

Более подробно алгоритм оптимизации, например, по методу Нелдера-Мида показан на рисунке 3.

ЧИСЛЕННЫЙ ПРИМЕР

Предположим, что контролируется процесс, исходные данные которого следующие: $p = 2$ (контролируется два коррелированных показателя процесса); $d = 2$ (характеристика нарушения процесса: обобщенная

Рис. 3. Блок-схема алгоритма метода Нелдера-Мида

дисперсия в два раза превышает целевое значение); $\lambda = 0,125$ (нарушение происходит примерно один раз за 8 часов); $g = 0,0167$ (время, необходимое на вычисление одного элемента выборки); $D = 0,083$ (время, затрачиваемое на выявление и устранение нарушения).

Специалист, исходя из технических возможностей процесса, задал ограничения объема выборки (n) от 3 до 5 наблюдений, частоты отбора выборок (h) от 1 до 8 часов и вероятности ложной тревоги (α) от 0,0001 до 0,01.

Требуется найти такие значения n , h и α , при которых время нахождения процесса в нерегулируемом состоянии было бы минимальным.

Таким образом, решается оптимизационная задача с ограничениями:

$$M(T_0) \rightarrow \min$$

при $3 < n < 5$; $1 < h < 8$; $0,0001 < \alpha < 0,01$.

На рисунке 4 представлены результаты расчета в разработанной программе.

Как видно из результатов расчета, два метода (l-bfgs-b и Нелдера-Мида) показали одинаковый результат: оптимальные значения параметров карты обобщенной дисперсии $n = 3$, $h = 1$, $\alpha = 0,001$, при которых целевая функция $M(T_0)$ принимает минимальное значение (0,76 часа). Положение контрольных границ карты при найденном уровне значимости α определяется по формуле (6).

Метод «trust-constr» дал несколько иные результаты: $n = 5$, $h = 1$, $\alpha = 0,0001$. Длительность нестабильного состояния процесса при полученных значениях параметров карты равна 0,74 часа.

```

Интерпретатор Python

1. Поиск локального минимума в доверительной области (trust-constr)
[1.e+00 5.e+00 1.e-04]
h= 1.0000000000
n= 5.0000000000
alfa = 0.0001000000
F(x) = 0.7424901188541844

2. Последовательное квадратичное программирование с ограничениями, ньютоновский метод решения системы Лагранжа (SLSQP)
[1.e+00 3.e+00 1.e-03]
h= 1.0000000000
n= 3.0000000000
alfa = 0.0010000000
F(x) = 0.7597749505926538

3. Метод от четверки Broyden-Fletcher-Goldfarb-Shanno (l-bfgs-b)
[1.e+00 3.e+00 1.e-03]
h= 1.0000000000
n= 3.0000000000
alfa = 0.0010000000
F(x) = 0.7597749505393139

Стек вызовов Переменные Выражения Точки останова Вывод Сообщения Интерпретатор Python

```

Рис. 4. Результаты решения оптимизационной задачи

Метод «trust-constr» дает чуть лучший результат, чем два других, следовательно, в качестве оптимальных принимаются именно его значения.

ЗАКЛЮЧЕНИЕ

Предложена модель оптимизации параметров карты обобщенной дисперсии с целью минимизации длительности нестабильного состояния процесса. На примере данных технологического процесса были найдены объем выборки, периодичность взятия выборок и значение положения контрольных границ, при которых время нахождения процесса в нестабильном состоянии минимально.

СПИСОК ЛИТЕРАТУРЫ

1. Montgomery D.C. *Introduction to Statistical Quality Control*. New York: John Wiley and Sons, 2009. 754 p.
2. Уиллер Д., Чамберс Д. Статистическое управление процессами / Оптимизация бизнеса с использованием контрольных карт Шухарта ; пер. с англ. М. : Альпина Бизнес Букс, 2009. 409 с.
3. Клячкин В.Н., Крашенинников В.Р., Кувайскова Ю.Е. Прогнозирование и диагностика стабильности функционирования технических объектов : монография. М. : РУСАЙНС, 2020. 200 с.
4. Боровиков В. STATISTICA: искусство анализа данных. Для профессионалов. СПб. : Питер, 2001. 656 с.
5. Алексеева А.В. Повышение эффективности статистического контроля многомерного рассеяния процесса // Автоматизация процессов управления. 2020. № 3 (61). С. 101–107.
6. Алексеева А.В., Клячкин В.Н. Выбор параметров алгоритма обобщенной дисперсии при многомерном статистическом контроле рассеяния процесса // Из-

вестия Самарского научного центра РАН. 2021. Т. 23, № 1 (99). С. 79–83.

7. Bersimis S., Psarakis S., Panaretos J. Multivariate Statistical Process Control Charts: An Overview // *Quality and Reliability Engineering International*. 2007. V. 23. pp. 517–543.

8. 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.

9. Ryan T.P. *Statistical Methods for Quality Improvement*. New York: John Wiley and Sons, 2011. 687 p.

10. 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. pp.228–242.

11. 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. pp. 07–26.

12. An economic-statistical design of control chart with multiple assignable causes / Y. Fong-Jung, T. Ching-Shih, H. Kai-I, W. Zhang // *Journal of Quality*, 2010. Vol. 17, № 4. pp. 327–338.

REFERENCES

1. Montgomery D.C. *Introduction to Statistical Quality Control*. New York, John Wiley and Sons Publ., 2009. 754 p.
2. Wheller D., Chambers D. *Statisticheskoe Upravlenie Protseessami. Optimizatsiia biznesa s ispolizovaniem kontrolnykh kart Shukharta*. Per. s angl. [Understanding Statistical Process. Transl. from Engl.]. Moscow, Alpina Businnes Books Press, 2009. 409 p.
3. Kliachkin V.N., Krasheninnikov V.R., Kuvaiskova Iu.E. *Prognozirovanie i diagnostika stabilnosti funktsionirovaniia tekhnicheskikh obektov. Monografiia* [Forecasting and

Diagnosis of Operation Consistency of Technical Objects. Monograph]. Moscow, Rusayns Publ., 2020. 200 p.

4. Borovikov V. *Statistika: iskusstvo analiza dannykh. Dlia professionalov* [Statistics: Data Analytics]. St. Petersburg, Piter Publ., 2001. 656 p.

5. 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.

6. Alekseeva A.V., Kliachkin V.N. Vybor parametrov algoritma obobshchennoi dispersii pri mnogomernom statisticheskom kontrole rasseianiia protsessa [Selection of the Parameters of the Generalized Dispersion Algorithm for Multivariate Statistical Control of the Process Scattering]. *Izvestiia Samarskogo nauchnogo tsentra RAN* [Proc. of Samara Scientific Center of RAS], 2021, vol. 23, no. 1 (99), pp. 79–83.

7. Bersimis S., Psarakis S., Panaretos J. Multivariate Statistical Process Control Charts: An Overview. *Quality*

and Reliability Engineering International, 2007, vol. 23, pp. 517–543.

8. García-Díaz J. Carlos. The ‘effective variance’ Control Chart for Monitoring the Dispersion Process with Missing Data. *Industrial Engineering*, 2007, vol. 1, no. 1, pp. 40–45.

9. Ryan T.P. *Statistical Methods for Quality Improvement*. New York, John Wiley and Sons Publ., 2011. 687 p.

10. 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, pp. 228–242.

11. 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, no. 2, pp. 7–26.

12. 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, no. 4, pp. 327–338.