



УДК 658.562.012.7, 519.24

В. Н. Лукин
ООО «Телесофт»

В. В. Яценко
Санкт-Петербургский государственный электротехнический
университет "ЛЭТИ" им. В. И. Ульянова (Ленина)

Вычисление контрольных пределов \bar{X} -карты на основе генерирования псевдослучайных чисел

Предложен подход к оценке контрольных пределов \bar{X} -карты в условиях, когда изначально известен только общий аналитический вид распределения индивидуальных значений характеристики стабильного процесса (параметры модели неизвестны). Новизна подхода заключается в переходе от распределения индивидуальных значений к распределению средних при помощи генерирования псевдослучайной выборки и её последующего преобразования. Квантили преобразованной выборки используются в качестве оценок контрольных пределов, применяемых в процедуре статистического вывода о состоянии (стабильном или нестабильном) процесса.

Контрольная карта, оценка контрольных пределов, отличное от нормального распределение, генерирование псевдослучайных чисел

Контрольная \bar{X} -карта Шухарта (далее \bar{X} -карта Шухарта) является одним из наиболее распространённых методов диагностики настройки среднего значения характеристик процесса на номинальное, т. е. на такое значение, которое установлено для этого процесса в качестве требований. В основе расчёта контрольных пределов, используемых для построения \bar{X} -карт Шухарта, лежит допущение о нормальности распределения статистических характеристик процесса, в частности его индивидуальных и средних значений (X и \bar{X} соответственно). Однако на практике распределение может отличаться от нормального и указанное допущение может увеличивать вероятность ошибок принятия гипотезы о состоянии наблюдаемого процесса как первого, так и второго родов. В свою очередь, рост числа ошибок приводит к возрастанию потерь, вызванных неадекватными управленческими решениями менеджмента.

Дополнительно, отличие от нормального распределения возрастает из-за малых объемов под-

групп, так как технически не всегда есть возможность применять большие подгруппы с целью использования нормализующего свойства *центральной предельной теоремы*.

В литературе предложен ряд методов рассматриваемого контроля процесса для случаев, когда распределение X отличается от нормального.

Среди *параметрических методов* следует отметить работы Феррелла [1] и Нельсона [2], в которых используется предположение об исходных распределениях X , описываемых, соответственно, логарифмически нормальным законом и законом Вейбулла. К недостаткам метода Феррелла относят необходимость экспоненциального преобразования данных, что может приводить к некорректной оценке контрольных пределов [3]. Для тех случаев, когда изначально закон распределения X неизвестен, в работе Кастаглиола и Кху [4] предложено вычислять контрольные пределы \bar{X} -карты, основываясь на кривых Джонсона (при помощи подгонки первых четырех мо-

ментов наблюдаемых данных). Недостатком указанного метода является наличие проблемы робастности контрольных карт (КК), проявляющейся в некорректных контрольных пределах. Некорректность заключается в подмене реального процесса (содержащим одновременно случайные и особые причины) случайным процессом с асимметрией и эксцессом, присущими этому реальному процессу.

Среди *непараметрических методов* следует отметить работу Бэйджера [5], предложившего рассчитывать контрольные пределы для \bar{X} при помощи метода «бутстреп» («bootstrap»). В работе Джонса и Вудэлла [6] приводится вывод о том, что результативность КК, построенных по данному методу, особенно высока для распределений с сильной асимметрией. К недостаткам указанного метода относят необходимость предварительного формирования выборки, характеризующей стабильный процесс, для настройки контрольных пределов.

В ряде отраслей (телекоммуникации, информационные технологии, медицина, электротехника, электроника и др.) существуют процессы, имеющие схожие признаки, которые следует учитывать при настройке КК. К этим признакам относятся:

- характеристика X – непрерывная переменная (в отличие от атрибутивных);
- теоретическая модель распределения X постоянна, в то время как значения её параметров $\theta_1, \dots, \theta_m$ изменяются в зависимости от условий сбора данных;
- объём выборки в подгруппе n , количество подгрупп k изменяются в зависимости от условий сбора данных;
- распределение X характеризуется большим значением коэффициента асимметрии $\alpha_3 \geq 1.5$;
- объём выборки в подгруппе $n \leq 50$;
- во всех случаях формирование выборки осуществляется в условиях отсутствия информации о состоянии стабильности процесса.

В соответствии с данными условиями необходимо подобрать значения контрольных пределов \bar{X} -карты для принятия решения о состоянии процесса (стабильном или нестабильном).

Оценка стабильности процессов с помощью КК в основном сводится к проблеме перехода от распределения индивидуальных значений к распределению средних. Такая проблема отсутствует при нормальном распределении, но становится

непреодолимой в ряде практических случаев. Определить аналитическую функцию распределения средних значений зачастую либо невозможно, либо это сопряжено со значительными техническими трудностями, либо экономически нецелесообразно.

В связи с недостатками рассмотренных методов, а также при наличии указанных признаков, в статье предложен метод, основанный на генерировании псевдослучайной выборки, квантили которой используются в качестве оценки контрольных пределов, используемых в процедуре статистического вывода о состоянии процесса.

Метод содержит четыре последовательных этапа, в рамках которых осуществляется оценка контрольных пределов \bar{X} -карты для процесса, обладающего вышеперечисленными признаками.

На **первом этапе** метода происходит принятие решения об аналитическом виде распределения генеральной совокупности данных. Для этого на основе предварительно собранных данных больших объемов о процессах, с помощью критерия согласия (например, Пирсона или Колмогорова), эмпирические распределения описываются теоретическим законом (с параметрами $\theta_1, \dots, \theta_m$).

На **втором этапе** отбираются значения характеристики X_{ij} , взятые из реального, исследуемого на стабильность процесса, где $i = 1, \dots, k$ – номер подгруппы в выборке; $j = 1, \dots, n$ – номер значения в подгруппе.

При этом объём выборки равен $n \cdot k$.

Далее для каждой i -й подгруппы рассчитываются:

- среднее подгруппы \bar{X}_i :

$$\bar{X}_i = \sum_{j=1}^n X_{ij} / n, \quad (1)$$

– одна из возможных оценок внутригрупповой изменчивости, например, выборочная дисперсия:

$$D_i = \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 / (n-1). \quad (2)$$

По k подгруппам осуществляется оценка числовых характеристик – математического ожидания $M[X]$ и дисперсии $D[X]$ теоретического закона распределения (решение о виде закона было при-

нято на первом этапе). В качестве $M[X]$ и $D[X]$ рассчитываются соответственно:

– общее среднее:

$$\bar{\bar{X}} = \sum_{i=1}^k \bar{X}_i / k; \quad (3)$$

– среднее значение внутригрупповой дисперсии:

$$\bar{D} = \sum_{i=1}^k (D_i) / k. \quad (4)$$

В [7] показано, что даже если контрольные пределы вычисляются по данным, «находящимся» за контрольными пределами, полученные пределы всё равно можно использовать для обнаружения особых причин вариации, так как изначально не известно, находится ли процесс в состоянии статистической управляемости или нет.

Использование среднего значения внутригрупповой дисперсии может приводить к расширению контрольных пределов, однако применение более сложных критериев серий, тем не менее, позволяет обнаруживать особые причины вариации.

Связь числовых характеристик $M[X]$, $D[X]$ и параметров $\theta_1, \dots, \theta_m$ теоретического распределения, форма которого определена на первом этапе, имеет следующий вид (для распределений с одним и двумя параметрами):

$$\begin{cases} M[X] = g_1(\theta_1, \dots, \theta_m), \\ D[X] = g_2(\theta_1, \dots, \theta_m). \end{cases}$$

Тогда, приняв $\bar{\bar{X}}$ и \bar{D} , соответственно, за числовые характеристики $M[X]$ и $D[X]$, можно рассчитать параметры теоретического распределения $\theta_1, \dots, \theta_m$.

На **третьем этапе** с помощью одного из методов генерирования псевдослучайных чисел (например, прямого метода, метода обратной функции, метода режекции [8]) формируется выборка, значения которой X_{ij}^* (символом «*» обозначены характеристики, для расчёта которых используется псевдослучайная выборка) образуются в соответствии с плотностью распределения $f_x(\theta_1, \dots, \theta_m)$.

Далее необходимо разбить выборку из псевдослучайных чисел на подгруппы объёма n , и для каждой подгруппы найти среднее \bar{X}_i^* и выбо-

рочную дисперсию D_i^* в соответствии с формулами (1) и (2). Впоследствии \bar{X}_i^* и D_i^* будут использованы для определения квантилей распределений, необходимых для оценки контрольных пределов \bar{X} - и D -карт.

Если генерируется выборка, объём которой равен B , а размер подгруппы принять равным n , то число подгрупп определяется выражением

$$K = \frac{B}{n}.$$

На **четвёртом этапе** задаётся некоторая вероятность α – вероятность ошибки первого рода.

После ранжирования \bar{X}_i^* и D_i^* в порядке возрастания значения с порядковым номером $\alpha/2 \cdot K$ будут являться нижними квантилями $\bar{X}_{\text{нижн}}^*$ и $D_{\text{нижн}}^*$ (нижними контрольными пределами \bar{X} - и D -карт соответственно). Значения с порядковым номером $(1 - \alpha/2)K$ будут являться верхними квантилями $\bar{X}_{\text{верх}}^*$ и $D_{\text{верх}}^*$ (верхними контрольными пределами \bar{X} - и D -карт соответственно). Значения $\bar{\bar{X}}$ и \bar{D} , вычисленные на этапе 2, будут являться центральными линиями \bar{X} - и D -карт соответственно.

В завершение строятся контрольные \bar{X} - и D -карты, на которых откладываются значения, взятые из реального, исследуемого на стабильность, процесса (см. этап 2), а также контрольные пределы, определённые на этапе 4 рассматриваемого метода.

Пример. Исследуется стабильность телекоммуникационного процесса передачи голосового трафика, в котором в качестве характеристики X измеряется длительность телефонного разговора [9].

На **первом этапе** метода на основе предварительно собранного большого объёма данных о телефонных разговорах абонентов в различных ситуациях принимается решение об аналитическом виде распределения значений X , присущего данной отрасли.

Для обеспечения условий стабильности процессов качество связи телефонных разговоров было предварительно протестировано (при помощи прослушивания звонков группа экспертов выявляла случаи, когда качество связи, по их мнению, отличалось от номинального). Далее, на основании критерия согласия Колмогорова, была принята гипотеза (на уровне значимости $\alpha = 1\%$)

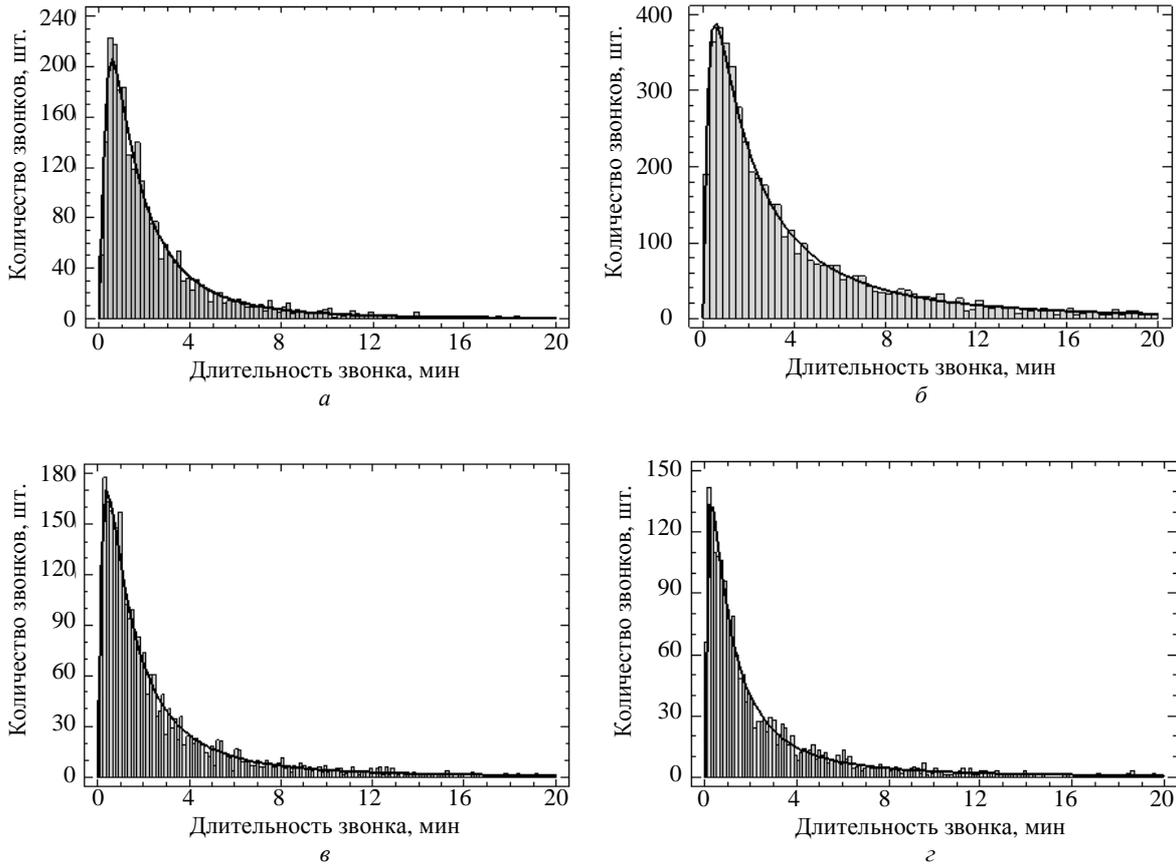


Рис. 1

о том, что эмпирическое распределение для данных случаев описывается логарифмически нормальным законом.

На рис. 1, а–г для четырех процессов представлены гистограммы распределений длительностей звонков и описывающие их кривые плотности логарифмически нормального распределения. Параметрами логарифмически нормального распределения являются μ и σ^2 .

На **втором этапе** метода для оценки стабильности отдельного телекоммуникационного процесса, состоящего из совокупности телефонных звонков, каждый из которых оценивается длительностью разговора X_{ij} , исследовалась выборка из 200 телефонных звонков, при этом $k = 10$; $n = 20$. Данные для подгруппы № 7 (см. таблицу) собирались в условиях прерывания связи во время разговора абонентов. Во всех остальных подгруппах выборки (кроме седьмой) проблемы с качеством связи отсутствовали.

В таблице приведены значения характеристик \bar{X}_i и D_i для каждой из подгрупп, вычисленные в соответствии с (1) и (2).

В соответствии с формулами (3) и (4) $\bar{X} = 5.61$, $\bar{D} = 131.68$.

Параметры логарифмически нормального распределения μ , σ^2 связаны с числовыми характеристиками $M[X]$, $D[X]$ следующими формулами:

$$M[X] = e^{\mu + \sigma^2/2},$$

$$D[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

Тогда, подставив вместо $M[X]$, $D[X]$ значения \bar{X} , \bar{D} , рассчитывают параметры логарифмически нормального распределения: $\mu = 0.9$, $\sigma^2 = 1.64$.

На **третьем этапе** при помощи программы STATGRAPHICS Centurion XV были сгенериро-

Параметр	Номер подгруппы (i)									
	1	2	3	4	5	6	7	8	9	10
\bar{X}_i	5.81	5.86	7.66	5.35	4.17	3.26	1.83	5.56	9.84	6.79
D_i	112.49	70.03	243.82	80.57	40.4	14.34	3.27	96.65	522.65	133.21

ваны $B = 20\,000$ псевдослучайных чисел X_{ij}^* , распределенных по логарифмически нормальному закону $\text{Log } N(0.9; 1.64)$.

После разбиения полученной выборки на $K = \frac{B}{n} = \frac{20\,000}{20} = 1000$ подгрупп, были рассчитаны значения \bar{X}_i^* и D_i^* для каждой из подгрупп, в соответствии с (1) и (2).

На рис. 2 представлены: *a* – график плотности $\text{Log } N(0.9; 1.64)$; *б* – гистограмма 20 000 псевдослучайных чисел X_{ij}^* , распределённых в соответствии с $\text{Log } N(0.9; 1.64)$; *в* – гистограмма распределения 1000 средних \bar{X}_i^* .

На **четвёртом этапе** была задана вероятность $\alpha = 1\%$.

После ранжирования \bar{X}_i^* в порядке возрастания, значение с порядковым номером $\frac{\alpha}{2}K = \frac{0.01}{2}1000 = 5$ является нижним квантилем $\bar{X}_{\text{нижн}}^* = 1.94$ (нижним контрольным пределом \bar{X} -карты), а значение с порядковым номером $\left(1 - \frac{\alpha}{2}\right)K = \left(1 - \frac{0.01}{2}\right)1000 = 995$ – верхним квантилем $\bar{X}_{\text{верх}}^* = 15.92$ (верхним контрольным пределом \bar{X} -карты). Значение $\bar{\bar{X}} = 5.61$ является центральной линией \bar{X} -карты.

После ранжирования D_i^* в порядке возрастания значение с порядковым номером 5 является нижним квантилем $D_{\text{нижн}}^* = 2.55$ (нижним кон-

трольным пределом D -карты), а значение с порядковым номером 995 – верхним квантилем $D_{\text{верх}}^* = 2020.01$ (верхним контрольным пределом D -карты). Значение $\bar{D} = 131.68$ является центральной линией D -карты.

В результате использования предложенного метода были построены \bar{X} -карта (рис. 3, *a*) и D -карта (рис. 3, *б*) исследуемого процесса. Выход точки \bar{X}_7 за нижний контрольный предел \bar{X} -карты указывает на отсутствие управляемости процесса. Таким образом, контрольная карта адекватно идентифицировала неуправляемый процесс в точке № 7.

Далее тестовая выборка (см. этап 2) была использована для сравнения предложенного в статье метода с другими методами диагностики стабильности процесса:

- Шухарта (контрольные карты Шухарта);
- Феррелла;
- на основе кривых Джонсона.

В качестве КК Шухарта использовались \bar{X} -карта (рис. 3, *в*) и S -карта (рис. 3, *з*).

По методу Феррелла в качестве карты средних использовалась $e^{\ddot{Y}}$ -карта (рис. 3, *д*), где \ddot{Y}_i – полусумма минимального и максимального значений Y_{ij} в i -й подгруппе, а Y_{ij} – значение нормально распределённой характеристики Y (с параметрами распределения μ, σ^2), где $Y_{ij} = \ln X_{ij}$. В качестве карты внутригрупповой изменчивости использовалась e^R -карта (рис. 3, *е*), где R_i – размах значений Y_{ij} в i -й подгруппе.

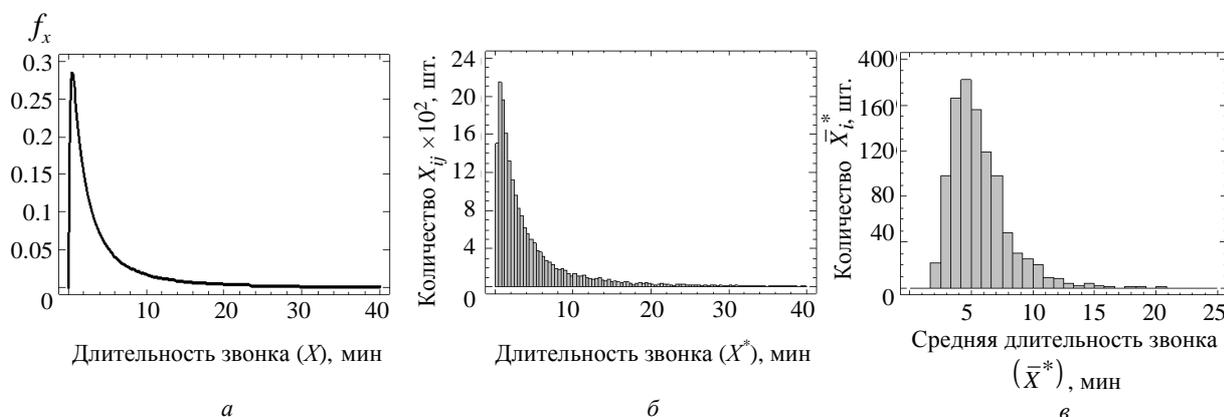


Рис. 2

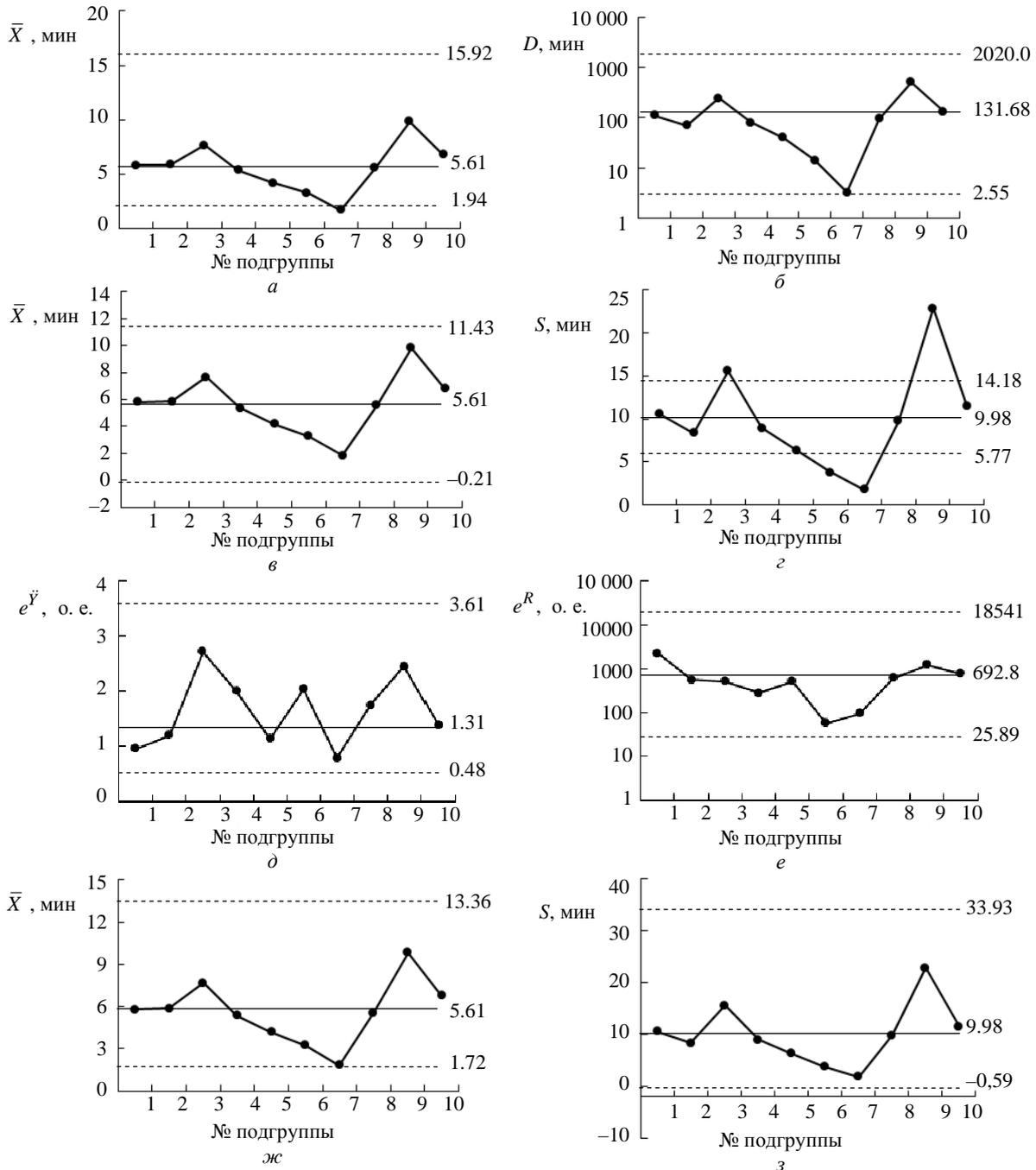


Рис. 3

Моменты кривой Джонсона, используемые для построения \bar{X} -карты (рис. 3, ж) и S -карты (рис. 3, з), были получены на основе распределений значений, соответственно, X_{ij} и S_i , где $S_i = \sqrt{D_i}$.

Результаты сравнительного эксперимента (при заданной вероятности $\alpha = 1\%$) подтвердили преимущество предложенного метода, с помощью которого была обнаружена точка неуправляемости процесса, в то время как известные методы (Шу-

харта, Феррелла, на основе кривых Джонсона) эту точку не обнаружили.

В статье не приводится сравнение предложенного метода с методом, рассмотренным в работе Бэйджера [5], так как последний предполагает предварительное формирование выборки, характеризующей стабильный процесс, для предварительной настройки контрольных пределов.

На основании изложенного можно сделать следующие выводы:

1. На практике существуют условия, в которых известные методы диагностики процессов не позволяют оценить их стабильность экономически целесообразным способом.

2. Предложенный метод оценки контрольных пределов позволяет решить проблему перехода от распределения индивидуальных значений к распределению средних для получения искомым квантилей. Такая проблема отсутствует при нормальном распределении, однако возникает в ряде практических случаев, связанных с отсутствием нормальности распределения данных.

3. Для сравнения чувствительности предложенного метода к выявлению особых причин вариаций с чувствительностью известных методов был реализован эксперимент. Для этого выборка с

известной точкой неуправляемости процесса была подвергнута предложенной процедуре построения контрольных карт. Результат эксперимента подтвердил преимущество предложенного метода, с помощью которого была обнаружена точка неуправляемости процесса, в то время как известные методы (Шухарта, Феррелла, на основе кривых Джонсона) эту точку не обнаружили.

4. Предложенный метод не содержит сложных аналитических вычислительных процедур, которые могли бы ограничить его применение в управлении качеством.

Предложенный метод в настоящее время успешно используется на практике в телекоммуникационной отрасли для улучшения качества сервиса.

СПИСОК ЛИТЕРАТУРЫ

1. Ferrell E. B. Control Charts for Lognormal Universe // Industrial Quality Control. 1958. Vol. 15, № 2. P. 4–6.

2. Nelson R. P. Control Charts for Weibull Processes with Standards Given // IEEE Transactions on Reliability. 1979. Vol. 28, № 4. P. 283–298.

3. Cheng S. W., Xie H. Control Charts for Lognormal Data. Tamkang // J. of Science and Engineering. 2000. Vol. 3, № 3. P. 131–137.

4. Castagliola P., Khoo Michael B. C. A Synthetic Scaled Weighted Variance Control Chart for Monitoring the Process Mean of Skewed Populations. // Communications in Statistics – Simulation and Computation. 2009. Vol. 38, № 8. P. 1659–1674.

5. Bajgier S. M. The use of Bootstrapping to Constructs Limits on Control Charts / Proceedings of the Decision Science Institute. San Diego, CA, 1992. P. 1611–1613.

6. Jones L. A., Woodall W. H. The Performance of Bootstrap Control Charts // J. of Quality Technology. 1998. Vol. 30, № 4. P. 362–375.

7. Уиллер Д., Чамберс Д. Статистическое управление процессами: оптимизация бизнеса с использованием контрольных карт Шухарта / пер. с англ. М.: Альпина Бизнес Букс, 2009. 409 с.

8. Генерация псевдослучайных чисел по заданному закону распределения / MATLAB. Exponenta. URL: http://matlab.exponenta.ru/statist/book3/1_2.php/.

9. Лукин В. Н., Яценко В. В. Оценка стабильности циклических процессов с использованием контрольных карт Шухарта // Изв. СПбГЭТУ «ЛЭТИ». 2013. № 5. С. 116–122.

V. N. Lukin
«Telesoft» Ltd

V. V. Yashenko
Saint-Petersburg state electrotechnical university «LETI»

COMPUTATION OF CONTROL LIMITS OF \bar{X} -CHART BASED ON PSEUDORANDOM NUMBERS GENERATION

This article is proposing a new approach to evaluation of control limits of \bar{X} -chart when initially only a general distribution function of individual values of a characteristic of a stable process is known (parameters of a model are unknown). The switch from the distribution of individual values to the distribution of means by pseudorandom numbers generation and its further transformation determine the novelty of the approach. The quantiles of the transformed sample are being used as estimators of control limits, which are being applied to identify the state (stable or instable) of the process.

Control chart, evaluation of control limits, non-normal distribution, pseudorandom numbers generation