

## ПАРАМЕТРИЧЕСКИЙ БУТСТРЕП-МЕТОД ВЫЧИСЛЕНИЯ КОНТРОЛЬНЫХ ПРЕДЕЛОВ КАРТ ДЛЯ АСИММЕТРИЧНО РАСПРЕДЕЛЕННЫХ ДАННЫХ

В. Н. Лукин<sup>а, б</sup>, соискатель, аналитик бизнес-процессов

В. В. Яценко<sup>а</sup>, канд. техн. наук, доцент

<sup>а</sup>Санкт-Петербургский государственный электротехнический университет «ЛЭТИ», Санкт-Петербург, РФ

<sup>б</sup>ООО «Телесофт», Санкт-Петербург, РФ

**Постановка проблемы:** при диагностике состояния стабильности процесса с асимметричным распределением индивидуальных значений, описываемым теоретическим законом, аналитику зачастую важно, чтобы метод построения контрольной карты соответствовал следующим критериям: обладал низкой вероятностью совершения ошибок 1-го и 2-го рода, наличием вычислительной робастности контрольных пределов, отсутствием нелинейных преобразований исходных данных. Однако параметрические (наиболее точные) методы построения контрольных карт не удовлетворяют одновременно всем этим критериям. Целью работы является разработка параметрического метода построения контрольной карты, соответствующего перечисленным критериям, и сравнение чувствительности карт средних и стандартных отклонений, построенных по разработанному методу, с чувствительностью таких же карт, построенных по распространенным на практике методам, на стадии мониторинга процесса (фаза II). **Результаты:** разработанный параметрический бутстреп-метод преобразует ранее предложенный авторами подход к вычислению контрольных пределов, основанный на генерировании псевдослучайной выборки, за счет использования несмещенной оценки внутригрупповой изменчивости (среднего значения квадрата стандартного отклонения) на этапе расчета оценок параметров теоретического распределения, что приводит к снижению вероятности совершения ошибки 1-го рода на стадии мониторинга процесса. Использование средних статистик внутригрупповой изменчивости увеличивает вычислительную робастность контрольных пределов к присутствию особых причин вариации значений характеристики, что позволяет применять метод на стадии ретроспективного анализа процесса (фаза I). В методе отсутствует необходимость в нелинейных преобразованиях исходных данных, усложняющих интерпретацию и применение на практике результатов анализа. Предложенный метод позволяет построить контрольную карту любой статистики для любого теоретического закона распределения (с одним или двумя параметрами) индивидуальных значений характеристики процесса. Значения оценок вероятности совершения ошибок 1-го рода картами средних и стандартных отклонений, вычисленные по предложенному методу, находятся ближе к требуемым значениям, чем значения оценок вероятности совершения ошибок 1-го рода такими же картами, вычисленные по другим методам. Значения оценок вероятности совершения ошибок 2-го рода контрольными картами, вычисленные по предложенному методу, таковы, что эти карты способны оперативно выявлять сдвиги в положении и изменчивости значений характеристики процесса. **Практическая значимость:** разработанный метод в настоящее время успешно используется в телекоммуникационной отрасли для мониторинга характеристики средней длительности телефонного разговора.

**Ключевые слова** — статистическое управление процессами, контрольная карта, оценка контрольных пределов, параметрический бутстреп-метод, асимметричное распределение, вероятности ошибок 1-го и 2-го рода.

### Введение

Контрольные карты (КК) Шухарта являются мощным инструментом диагностики состояния стабильности процесса. Наиболее распространены карты статистик положения и изменчивости, особенно  $\bar{X}$ - и  $S$ -карты Шухарта [1]. В основе расчета их контрольных пределов лежит допущение о нормальном (или приблизительно нормальном) распределении индивидуальных значений характеристик процесса ( $X$ ). Однако известно много процессов, значениям характеристик которых свойственны асимметричные распределения, и тогда указанное ранее допущение будет увеличивать вероятность совершения ошибок 1-го и 2-го рода при принятии гипотезы о состоянии наблюдаемого процесса. В качестве примеров характеристик таких процессов можно привести длительность телефонного разговора, наработку устройства на отказ, отклонение формы поверх-

ности от круглости, предел прочности материала на разрыв и т. д.

### Обзор литературы

Можно выделить четыре основных подхода к построению карт статистик положения и изменчивости для случаев, когда распределению  $X$  свойственна асимметрия.

*Первый подход* заключается в увеличении объема выборки ( $n$ ) в подгруппах настолько, чтобы распределение средних значений характеристики процесса ( $\bar{X}$ ) стало приблизительно нормальным. Это позволит использовать  $\bar{X}$ -карту в качестве карты статистики положения и  $mR$ -карту (карту скользящих размахов для средних) в качестве карты статистики изменчивости. Однако увеличение объема выборки в подгруппах зачастую невозможно либо экономически нецелесообразно.

*Второй подход* предполагает использование непараметрических методов построения КК, не зависящих от теоретического закона распределения. Авторы работы [2] предложили эвристический метод взвешенной дисперсии (*weighted variance* — WV), основанный на аппроксимации полудисперсий, и представили формулы контрольных пределов  $\bar{X}$ - и R-карт для асимметричных распределений. Предложена [3] и другая версия WV-метода для расчета контрольных пределов  $\bar{X}$ - и S-карт. Метод взвешенных стандартных отклонений (*weighted standard deviations* — WSD), схожий с WV-методом [4], используется для построения  $\bar{X}$ -карт, карт кумулятивных сумм и карт экспоненциально взвешенных скользящих средних для асимметричных распределений. Для построения  $\bar{X}$ - и R-карт предложен метод поправки на асимметрию (*skewness correction* — SC) [5], в основе которого лежит учет значения коэффициента асимметрии распределения при вычислении контрольных пределов.

В работе [6] впервые предложено вычислять контрольные пределы  $\bar{X}$ -карты при помощи непараметрического бутстреп-метода<sup>1</sup>. Результативность КК, построенных по данному методу, особенно высока для распределений с сильной асимметрией [7]. Главным недостатком, ограничивающим использование непараметрических бутстреп-методов на практике, является отсутствие вычислительной робастности пределов КК: даже если контрольные пределы вычисляются по данным, содержащим в том числе и особые причины вариации, необходимо, чтобы полученные пределы можно было использовать для обнаружения отсутствия статистической управляемости [1]. Поэтому при использовании метода [6] необходимо предварительно формировать выборку, характеризующую стабильный процесс (что далеко не всегда возможно на практике), для последующей оценки квантилей.

*Третий подход* предполагает вычисление контрольных пределов карт при помощи кривых Джонсона (КД) или Пирсона. Его целесообразно использовать в тех случаях, когда изначально закон распределения  $X$  неизвестен. Например, в работе [8] предложено вычислять контрольные пределы  $\bar{X}$ -карты, основываясь на квантилях КД (при помощи подгонки первых четырех моментов наблюдаемых данных [9]). Недостатком указанного метода является отсутствие робастности КК (в случае нестабильного процесса), что проявляется в некорректных контрольных пределах. Некорректность заключается в подмене реального процесса, содержащего одновременно

<sup>1</sup> Бутстреп-метод (англ. bootstrap method) — практический компьютерный метод определения статистик распределений, основанный на многократной генерации выборок методом Монте-Карло на базе имеющейся выборки.

случайные и особые причины, стабильным процессом с асимметрией и эксцессом, присущими этому реальному процессу. Метод реализован в версиях программного продукта STATISTICA.

*Четвертый подход* связан с использованием параметрических методов. На практике распределения значений характеристик многих процессов описываются теоретическими законами (в некотором приближении): распределение длительностей телефонных разговоров описывается логарифмически нормальным законом, распределение значений наработки устройства на отказ — законом Вейбулла, распределение отклонений формы поверхности от круглости — законом Рэлея и т. д. Для контроля таких процессов применимы параметрические методы построения КК статистик положения и изменчивости.

В работе [10] предложен метод построения КК для логарифмически нормального закона распределения  $X$ . К недостаткам метода относят необходимость экспоненциального преобразования данных, что снижает корректность оценки контрольных пределов. В работе [11] предложен метод построения КК для случаев, когда индивидуальные значения характеристик процесса распределены в соответствии с законом Вейбулла. Практическое использование этого метода затруднено сложностью вычисления значений констант для определения пределов КК.

В последнее время для построения КК все более широко применяют параметрические бутстреп-методы (ПБ-методы). Идея ПБ-метода заключается в переходе от теоретического распределения значений  $X$  к распределению значений интересующей статистики при помощи генерирования псевдослучайных чисел для исследования характеристик (математического ожидания, медианы, дисперсии, квантилей и др.) распределения интересующей статистики. Без использования ПБ-метода определить аналитическую функцию распределения значений таких статистик, как  $\bar{X}$  и  $S$ , зачастую либо затруднительно, либо невозможно<sup>2</sup> [8].

В литературе предложен целый ряд ПБ-методов оценки контрольных пределов для мониторинга квантилей интересующей статистики ( $\xi_p$ ), когда значения  $X$  распределены в соответствии с некоторым теоретическим законом. Например, авторы работы [12] предложили метод оценки контрольных пределов для мониторинга квантилей  $W_p^w$ , когда  $X$  распределены в соответствии с законом Вейбулла, а авторы работы [13] — для мониторинга квантилей  $W_p^{b-s}$ , когда  $X$  распре-

<sup>2</sup> Это касается случаев, когда значения  $X$  распределены в соответствии с одним из следующих теоретических законов: логарифмически нормальным, Вейбулла, гамма и др.

делены в соответствии с законом Бирнбаума — Саундерса. Авторы работы [14] рассмотрели четыре теоретических распределения значений  $X$  (Вейбулла, Бэрра, Бирнбаума — Саундерса и Парето) и для каждого из них предложили ПБ-метод оценки контрольных пределов для мониторинга квантилей интересующей статистики с возможностью переменного объема выборки в подгруппах. Главным достоинством представленных выше ПБ-методов является высокая точность получения заданной  $\alpha$ -вероятности. Однако их использование возможно только в фазе II, так как изначально предполагается, что выборка, на основе которой вычисляются контрольные пределы, характеризует стабильный процесс.

Особенность ПБ-метода построения КК, предложенного нами [15], заключается в использовании средних (или медиан) из  $k$  ( $k$  — это число подгрупп) статистик положения и изменчивости для формирования теоретической модели распределения  $X$  (при помощи аналитической связи между математическим ожиданием  $M[X]$ , дисперсией  $D[X]$  и параметрами  $(\theta_1, \dots, \theta_m)$  теоретического распределения  $X$ ), в отличие от использования оценок параметров распределения, рассчитанных по  $n \times k$  значениям  $X_{ij}$  [12–14]. Как и для КК Шухарта, использование среднего статистик внутригрупповой изменчивости в методе [15] позволяет рассчитывать корректные оценки контрольных пределов даже на основе частично неоднородных данных, что является признаком вычислительной робастности контрольных пределов.

Важно отметить, что в ПБ-методе [15] отсутствует необходимость в нелинейных преобразованиях исходных значений  $X$  (для перехода к другой характеристике), усложняющих интерпретацию и применение на практике результатов анализа. При помощи предложенного метода можно построить КК любой статистики для любого теоретического закона распределения значений  $X$  (с одним или двумя параметрами).

Для преобразования метода [15] мы предлагаем на этапе расчета оценок параметров теоретического распределения  $X$  использовать несмещенную оценку внутригрупповой изменчивости  $S^2$  (в фазе II).

### Критерии выбора методов построения $\bar{X}$ - и $S$ -карт

Методы построения КК отбирались для исследования их чувствительности в фазе II в соответствии со следующими критериями:

- возможность описания теоретической моделью распределения значений  $X$  (если необходимо);
- наличие вычислительной робастности контрольных пределов;

— отсутствие нелинейных преобразований исходных значений  $X$ .

Комментируя эти критерии, необходимо отметить следующее.

Поскольку распределения значений характеристик многих процессов описываются теоретическими законами (в некотором приближении), то в исследовании будет использоваться допущение о том, что значения  $X$  рассматриваемых процессов распределены в соответствии с двумя наиболее распространенными на практике асимметричными теоретическими законами: логарифмически нормальным законом распределения и законом распределения Вейбулла. Это позволяет использовать параметрические методы построения КК.

На практике для большинства процессов изначально никогда не известно, находится данный процесс в состоянии статистической управляемости или нет. Поэтому при выборе метода построения КК важна вычислительная робастность контрольных пределов.

Нелинейные преобразования значений  $X$ , используемые в ряде методов построения КК, усложняют интерпретацию и применение на практике результатов анализа, что может привести к некорректной оценке состояния стабильности процесса. В связи с этим нелинейные преобразования данных являются нежелательными [1], и методы, использующие их для построения контрольных пределов, в исследовании чувствительности КК рассмотрены не будут.

В телекоммуникациях соответствие метода построения КК данным критериям особенно критично при анализе длительностей телефонных разговоров. Также отметим, что во многих процессах характеристика  $\bar{X}$  является отраслевой, например, в телекоммуникациях таковой является средняя длительность телефонного разговора (*average call duration* — ACD). Поэтому для контроля таких процессов наиболее удобно использовать  $\bar{X}$ -карту. В исследовании будет рассмотрен средний ( $n = 10$ ) объем внутригрупповой выборки, часто используемый при анализе телекоммуникационных процессов. В таких случаях для мониторинга изменчивости значений характеристики процесса предпочтительно использовать  $S$ -, а не  $R$ -карту [3].

В соответствии с вышеуказанными критериями для исследования были выбраны следующие методы построения  $\bar{X}$ - и  $S$ -карт:

- метод Шухарта [1];
- WV-метод [3];
- метод на основе КД [8, 9];
- ПБ-метод, предлагаемый авторами.

Алгоритмы построения КК по методу Шухарта, WV-методу и методу на основе КД, использованные в исследовании, можно получить от авторов статьи по запросу.

**Алгоритм параметрического бутстреп-метода**

$\bar{X}$ - и  $S$ -карты предложенного нами ПБ-метода строятся в соответствии со следующими шагами.

1. Принимается решение о теоретическом законе, описывающем распределение значений характеристики  $X$  текущего процесса. Например, на основе предварительно собранных данных больших объемов о процессе (стабильность которого должна быть проверена группой экспертов) решение об аналитическом виде распределения значений  $X$  может быть принято при помощи одного из критериев согласия [15].

2. Из текущего процесса отбираются  $n \times k$  значений характеристики  $X_{ij}$ , где  $i = 1, \dots, k$  — номер подгруппы в выборке;  $j = 1, \dots, n$  — номер значения в подгруппе.

Для каждой  $i$ -й подгруппы вычисляются следующие статистики:

— среднее подгруппы

$$\bar{X}_i = \sum_{j=1}^n X_{ij} / n; \tag{1}$$

— стандартное отклонение

$$S_i = \sqrt{\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 / (n-1)}; \tag{2}$$

— квадрат стандартного отклонения

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

3. По  $k$  подгруппам рассчитываются:

— общее среднее

$$\bar{\bar{X}} = \sum_{i=1}^k \bar{X}_i / k;$$

— среднее значение стандартного отклонения

$$\bar{S} = \sum_{i=1}^k (S_i) / k;$$

— среднее значение квадрата стандартного отклонения

$$\bar{S}^2 = \sum_{i=1}^k (S_i^2) / k.$$

4. Для теоретического распределения с двумя параметрами (с одним параметром — аналогично), определенного на шаге 1, математическое ожидание  $M[X]$ , дисперсия  $D[X]$  и параметры распределения  $(\theta_1, \theta_2)$  имеют аналитическую связь следующего вида:

$$\begin{cases} M[X] = g_1(\theta_1, \theta_2) \\ D[X] = g_2(\theta_1, \theta_2) \end{cases}.$$

Статистики  $\bar{\bar{X}}$  и  $\bar{S}^2$  (или  $(\bar{S})^2$ ) принимаются, соответственно, за  $M[X]$  и  $D[X]$ , и рассчитываются оценки параметров теоретического распределения  $(\hat{\theta}_1, \hat{\theta}_2)$ .

В фазе I контрольные пределы рассчитываются для того, чтобы определить, находился ли процесс в стабильном состоянии в том промежутке времени, когда  $n \times k$  значений  $X_{ij}$  отбирались для построения КК. Если процесс находился в нестабильном состоянии, то необходимо идентифицировать и устранить особые причины вариации значений характеристики  $X$ . Если состояние процесса было стабильным, то выборка из  $n \times k$  значений  $X_{ij}$ , на основе которой строилась КК, может быть использована в фазе II [16]. В фазе I целесообразно использовать статистику  $(\bar{S})^2$ , так как она более робастна к присутствию неуправляемой вариабельности значений  $X$ , чем статистика  $\bar{S}^2$ .

В фазе II контрольные пределы рассчитываются на основе выборки, характеризующей стабильный процесс (взятой из фазы I). В интерактивном режиме новые значения исследуемой статистики сравниваются со значениями контрольных пределов для оперативного выявления неуправляемой вариабельности процесса [16].

В отличие от  $(\bar{S})^2$  статистика  $\bar{S}^2$  является несмещенной оценкой дисперсии (т. е.  $\bar{S}^2$  является более точной оценкой  $D[X]$ , чем  $(\bar{S})^2$ ), поэтому в фазе II использование статистики  $\bar{S}^2$  более целесообразно.

В исследовании чувствительности  $\bar{X}$ - и  $S$ -карт будет рассматриваться фаза II, поэтому в качестве оценки  $D[X]$  принимается статистика  $\bar{S}^2$ .

Для логарифмически нормального распределения  $\text{Log}N(\mu; \sigma^2)$

$$\begin{cases} \hat{\mu} = \ln(\bar{\bar{X}}) - \frac{\hat{\sigma}^2}{2} \\ \hat{\sigma} = \sqrt{\ln\left(\frac{\bar{S}^2}{e^{2 \ln(\bar{\bar{X}})} + 1}\right)}. \end{cases}$$

Для распределения Вейбулла  $W(\delta; \psi)$

$$\begin{cases} \hat{\psi} = \frac{\bar{\bar{X}}}{\Gamma\left(1 + \frac{1}{\hat{\delta}}\right)} \\ \frac{\Gamma\left(1 + \frac{2}{\hat{\delta}}\right)}{\Gamma^2\left(1 + \frac{1}{\hat{\delta}}\right)} = 1 + \frac{\bar{S}^2}{(\bar{\bar{X}})^2} \end{cases}, \tag{3}$$

где гамма-функция  $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ .

Оценка параметра формы  $\hat{\delta}$  находится из нижнего уравнения (3) при помощи одного из численных методов, доступных в большинстве математических программных продуктов (точность оценок  $\hat{\delta}$  в приведенном далее исследовании составляет 0,001).

5. С помощью одного из методов генерирования псевдослучайных чисел формируется выборка объема  $N$ , значения которой  $X_{ij}^*$  (символом «\*» в алгоритме ПБ-метода обозначены характеристики, для расчета которых используется псевдослучайная выборка) образуются в соответствии с плотностью теоретического распределения  $f_x(\hat{\theta}_1, \hat{\theta}_2)$ . Псевдослучайные числа, распределенные в соответствии с  $\text{Log}N(\hat{\mu}; \hat{\sigma}^2)$  и  $W(\hat{\delta}; \hat{\psi})$ , в приведенном далее исследовании были сгенерированы при помощи программного продукта STATGRAPHICS Centurion XV.

6. Полученная выборка разбивается на  $B = N/n$  подгрупп, и для каждой подгруппы объема  $n$  рассчитываются значения статистик  $\bar{X}_i^*$  и  $S_i^*$  в соответствии с (1) и (2).

7. В порядке возрастания ранжируются  $B$  значений  $\bar{X}_i^*$ . Нижним пределом  $LCL_{\bar{X}}$  является значение с порядковым номером  $(\alpha/2)B$  (в приведенном далее исследовании задавалось значение  $\alpha = 0,0027$ ). Верхним пределом  $UCL_{\bar{X}}$  является значение с порядковым номером  $[1 - (\alpha/2)]B$ .

8. В порядке возрастания ранжируются  $B$  значений  $S_i^*$ . Нижним пределом  $LCL_S$  является значение с порядковым номером  $(\alpha/2)B$ . Верхним пределом  $UCL_S$  является значение с порядковым номером  $[1 - (\alpha/2)]B$ .

Необходимо отметить, что на шаге 6 вместо значений  $\bar{X}_i^*$  и  $S_i^*$  можно рассчитывать значения любой другой интересующей статистики, чтобы на шагах 7, 8 вычислять контрольные пределы для этой статистики.

### Исследование чувствительности $\bar{X}$ - и $S$ -карт

Для исследования чувствительности  $\bar{X}$ - и  $S$ -карт, построенных по представленным методам, в фазе II было проведено компьютерное моделирование по методу Монте-Карло. В качестве показателей чувствительности КК использованы оценки вероятностей совершения ошибок 1-го и 2-го рода (оценки  $\alpha$ - и  $\beta$ -вероятности соответственно). Данные, использованные в исследовании, получены генерированием выборки двух асимметричных теоретических распределений: логарифмически нормального распределения  $\text{Log}N(\mu; \sigma^2)$  и распределения Вейбулла  $W(\delta; \psi)$ . Значения  $X$ , на основе которых вычислялись контрольные пределы  $\bar{X}$ - и  $S$ -карт, характеризуют выборку стабильного процесса. Для карт, построенных по методу Шухарта и WV-методу, в ка-

честве контрольных пределов были приняты границы  $\pm 3\sigma$ , а для карт, построенных по методу на основе КД и ПБ-методу, — квантили  $UCL_{99,865\%}$  и  $LCL_{0,135\%}$ .

Значения оценок  $\alpha$ -вероятности  $\bar{X}$ - и  $S$ -карт вычисляются по следующим шагам.

1. Вычисляются контрольные пределы карт.

1.1. Генерируется  $n = 10$  значений характеристики  $X$ , распределенных в соответствии с  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$ , при помощи программного продукта STATGRAPHICS Centurion XV.

1.2. Шаг 1.1 повторяется  $k = 10$  раз для формирования  $k$  подгрупп (по  $n$  значений в каждой подгруппе).

1.3. Контрольные пределы  $\bar{X}$ - и  $S$ -карт рассчитываются в соответствии с алгоритмами метода Шухарта, WV-метода, метода на основе КД и ПБ-метода.

1.4. Шаги 1.1–1.3 повторяются  $10^2$  раз для каждого метода.

1.5. Шаги 1.1–1.4 повторяются для  $\text{Log}N(0,44; 1,32)$ ,  $\text{Log}N(1,53; 0,52)$ ,  $\text{Log}N(1,74; 0,1)$ ,  $W(0,75; 5)$ ,  $W(1,24; 3)$ ,  $W(2,6; 3)$ .

2. Вычисляются значения оценок  $\alpha$ -вероятности.

2.1. Генерируется  $n = 10$  значений характеристики  $X$ , распределенных в соответствии с  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$  (аналогично шагу 1.1).

2.2. По  $n$  значениям, полученным на шаге 2.1, рассчитываются статистики  $\bar{X}_i$  и  $S_i$ .

2.3. В соответствии с  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$  из шага 2.1 выбирается одна из  $\bar{X}$ -карт, построенных по некоторому методу на шагах 1.1–1.5. Значение  $\bar{X}_i$ , рассчитанное на шаге 2.2, сравнивается со значениями  $UCL_{\bar{X}}$  и  $LCL_{\bar{X}}$  этой  $\bar{X}$ -карты по следующим формулам:

$$I_{UCL}(\bar{X}_i) = \begin{cases} 1, & \text{если } \bar{X}_i > UCL_{\bar{X}}; \\ 0, & \text{если } \bar{X}_i \leq UCL_{\bar{X}}; \end{cases}$$

$$I_{LCL}(\bar{X}_i) = \begin{cases} 1, & \text{если } \bar{X}_i < LCL_{\bar{X}}; \\ 0, & \text{если } \bar{X}_i \geq LCL_{\bar{X}}. \end{cases}$$

Значения  $I_{UCL}(\bar{X}_i)$  и  $I_{LCL}(\bar{X}_i)$  рассчитываются для каждой из  $\bar{X}$ -карт, построенных по выбранному методу на шагах 1.1–1.5 (при  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$  из шага 2.1).

2.4. В соответствии с  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$  из шага 2.1 выбирается одна из  $S$ -карт, построенных по некоторому методу на шагах 1.1–1.5. Значение  $S_i$ , рассчитанное на шаге 2.2, сравнивается со значениями  $UCL_S$  и  $LCL_S$  этой  $S$ -карты по следующим формулам:

$$I_{UCL}(S_i) = \begin{cases} 1, & \text{если } S_i > UCL_S; \\ 0, & \text{если } S_i \leq UCL_S; \end{cases}$$

$$I_{LCL}(S_i) = \begin{cases} 1, & \text{если } S_i < LCL_S; \\ 0, & \text{если } S_i \geq LCL_S. \end{cases}$$

Значения  $I_{UCL}(S_i)$  и  $I_{LCL}(S_i)$  рассчитываются для каждой из  $S$ -карт, построенных по выбранному методу на шагах 1.1–1.5 (при  $\text{Log}N(\mu; \sigma^2)$  или  $W(\delta; \psi)$  из шага 2.1).

2.5. Шаги 2.1–2.4 повторяются  $10^4$  раз.

2.6. Для выбранного метода вычисляются средние из значений  $I_{UCL}(\bar{X}_i)$  и  $I_{LCL}(\bar{X}_i)$ , которые являются оценкой  $\alpha_{\bar{X}}$ -вероятности выхода за  $UCL_{\bar{X}}$  и  $LCL_{\bar{X}}$  соответственно.

2.7. Для выбранного метода вычисляются средние из значений  $I_{UCL}(S_i)$  и  $I_{LCL}(S_i)$ , которые являются оценкой  $\alpha_S$ -вероятности выхода за  $UCL_S$  и  $LCL_S$  соответственно.

2.8. Шаги 2.3–2.7 повторяются для каждого из методов построения КК.

2.9. Шаги 2.1–2.8 повторяются для  $\text{Log}N(0,44; 1,32)$ ,  $\text{Log}N(1,53; 0,52)$ ,  $\text{Log}N(1,74; 0,1)$ ,  $W(0,75; 5)$ ,  $W(1,24; 3)$ ,  $W(2,6; 3)$ .

По описанным выше методам построены (для  $n = 10$ ) оценки  $\alpha_{\bar{X}}$ -вероятности выхода за  $UCL_{\bar{X}}$

и  $LCL_{\bar{X}}$   $\bar{X}$ -карт и оценки  $\alpha_S$ -вероятности выхода за  $UCL_S$  и  $LCL_S$   $S$ -карт (и сумма этих оценок), представленные в табл. 1 и 2.

Значения оценок  $\alpha$ -вероятности выхода за  $UCL$  и  $LCL$  (отдельно для верхнего и нижнего пределов) КК, построенных в соответствии с ПБ-методом, наиболее близки к оптимальным. На практике именно этот критерий является основным при выборе оптимального метода построения КК.

В качестве примера на рис. 1 представлены графические зависимости оценок  $\alpha_{\bar{X}}$ -вероятности выхода за  $LCL_{\bar{X}}$   $\bar{X}$ -карты (для каждого из рассмотренных методов) от величины  $P_x$ , где  $P_x$  — вероятность того, что  $X \leq M[X]$ .

Из рис. 1 видно, что для всех рассмотренных методов, за исключением ПБ-метода, по мере увеличения асимметрии распределения значительно возрастает отклонение  $\alpha_{\bar{X}}$  от требуемого значения 0,135 %. Несмотря на то, что при  $P_x = 0,72$  использование метода Шухарта и WV-

■ Таблица 1. Оценки, %,  $\alpha_{\bar{X}}$ -вероятности выхода за  $UCL_{\bar{X}}$  и  $LCL_{\bar{X}}$  (и сумма этих оценок)  $\bar{X}$ -карт при  $n = 10$

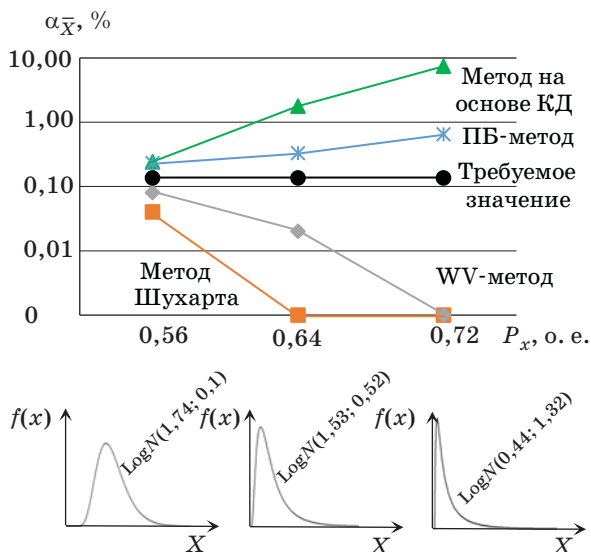
Распределение	$P_x$	$K_3[X]^1$	$K_4[X]^2$	$M[X]$	$SD[X]$	Метод Шухарта		WV-метод		Метод на основе КД		ПБ-метод	
						LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
Логарифмически нормальное	0,72	7,76	122,75	3	5	0,00	4,10	0,00	1,74	7,36	2,08	0,65	0,76
						4,10		1,74		9,44		1,41	
	0,64	2,94	16,95	6	5	0,00	2,22	0,02	1,07	1,77	1,00	0,33	0,45
						2,22		1,09		2,77		0,78	
	0,56	1,02	1,90	6	2	0,04	0,61	0,08	0,42	0,24	0,31	0,23	0,31
						0,65		0,50		0,55		0,54	
Вейбулла	0,68	3,11	15,44	6	8	0,00	2,72	0,00	1,20	1,11	1,06	0,33	0,47
						2,72		1,20		2,17		0,80	
	0,60	1,43	2,77	2,8	2,25	0,02	1,33	0,05	0,75	0,35	0,56	0,21	0,40
						1,35		0,80		0,91		0,61	
	0,52	0,31	-0,20	2,66	1,1	0,13	0,42	0,15	0,36	0,19	0,33	0,28	0,26
						0,55		0,51		0,52		0,54	

<sup>1</sup>  $K_3$  — коэффициент асимметрии.

<sup>2</sup>  $K_4$  — коэффициент эксцесса.

■ Таблица 2. Оценки, %,  $\alpha_S$ -вероятности выхода за  $UCL_S$  и  $LCL_S$  (и сумма этих оценок)  $S$ -карт при  $n = 10$

Распределение	$P_x$	$K_3[X]$	$K_4[X]$	$M[X]$	$SD[X]$	Метод Шухарта		WV-метод		Метод на основе КД		ПБ-метод	
						LCL	UCL	LCL	UCL	LCL	UCL	LCL	UCL
Логарифмически нормальное	0,72	7,76	122,75	3	5	8,37	13,96	0,00	2,39	9,80	6,50	0,21	0,70
						22,33		2,39		16,30		0,91	
	0,64	2,94	16,95	6	5	0,64	10,14	0,00	2,13	7,86	6,46	0,17	0,35
						10,78		2,13		14,32		0,52	
	0,56	1,02	1,90	6	2	0,05	2,62	0,01	1,20	6,85	5,43	0,22	0,36
						2,67		1,21		12,28		0,58	
Вейбулла	0,68	3,11	15,44	6	8	2,49	12,18	0,00	2,40	7,75	4,91	0,19	0,55
						14,67		2,40		12,66		0,74	
	0,60	1,43	2,77	2,8	2,25	0,11	5,25	0,01	1,63	6,10	7,06	0,16	0,41
						5,36		1,64		13,16		0,57	
	0,52	0,31	-0,20	2,66	1,1	0,01	0,54	0,03	0,49	4,24	3,97	0,15	0,33
						0,55		0,52		8,21		0,48	



■ Рис. 1. Зависимость оценок  $\alpha_{\bar{x}}$ -вероятности выхода за  $LCL_{\bar{x}}$   $\bar{X}$ -карт от величины  $P_x$  при  $n = 10$

метода приводит к отсутствию ошибок 1-го рода, данный факт сигнализирует о возможной неспособности этих методов вычисления  $LCL_{\bar{x}}$  к идентификации неуправляемой вариабельности процесса (что и будет подтверждено далее, см. рис. 2).

При сравнении оценок  $\beta$ -вероятности  $\bar{X}$ - и  $S$ -карт, построенных по представленным выше методам, необходимо учитывать, что карты с бóльшими значениями оценок  $\alpha$ -вероятности заведомо имеют преимущество. Поэтому основная задача данного исследования оценок  $\beta$ -вероятности — выяснить способность КК, построенных по ПБ-методу, к идентификации нестабильного состояния процесса (в принципе).

Наиболее распространенный метод моделирования распределения, характеризующего нестабильный процесс [3], заключается в сдвиге значений  $M[X]$  и  $SD[X]$  распределения стабильного процесса на заданное число сигма-единиц в соответствии с формулами

$$M[X]^* = M[X] + a \cdot SD[X];$$

$$SD[X]^* = b \cdot SD[X],$$

где  $M[X]^*$  и  $SD[X]^*$  — математическое ожидание и стандартное отклонение модели распределения нестабильного процесса;  $a$  и  $b$  — параметры, характеризующие величины сдвига значений  $M[X]$  и  $SD[X]$ .

Данный метод удобен при моделировании нормального (симметричного) распределения нестабильного процесса:

— при сдвигах значений  $M[X]$  и  $SD[X]$  форма распределения остается неизменной;

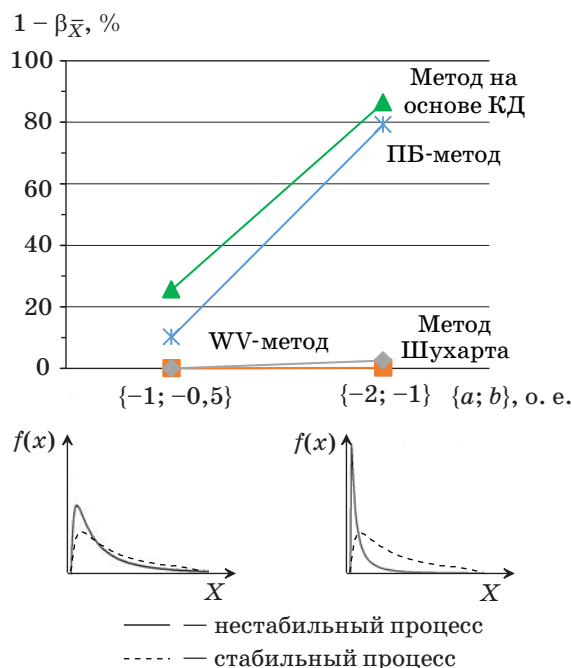
— кривые плотности распределения нестабильного процесса, сдвинутые на  $\pm a \cdot SD[X]$  от  $M[X]$ , симметричны относительно  $M[X]$ .

Однако данные условия не выполняются в случае моделирования асимметричного распределения нестабильного процесса, что усложняет выбор подходящих параметров  $a$  и  $b$ . В связи с этим для моделирования асимметричного распределения нестабильного процесса нами предложено использовать WV-метод (учитывающий асимметрию распределения). Алгоритм моделирования асимметричного распределения нестабильного процесса, основанный на WV-методе, и результаты вычисления оценок  $(1 - \beta)$ -вероятности  $\bar{X}$ - и  $S$ -карт, построенных по рассмотренным методам, можно получить от авторов статьи по запросу.

В качестве примера на рис. 2 представлены графические зависимости оценок  $(1 - \beta_{\bar{x}})$ -вероятности выхода за  $LCL_{\bar{x}}$   $\bar{X}$ -карты (для каждого из рассмотренных методов) от значений параметров  $\{a; b\}$  (для  $n = 10$ ). В качестве модели распределения значений  $X$  стабильного процесса рассмотрено  $\text{LogN}(0,44; 1,32)$  с  $P_x = 0,72$ . Моделями распределений нестабильного процесса являются  $\text{LogN}(-0,12; 1,43)$  при значениях параметров  $\{-1; -0,5\}$  и  $\text{LogN}(-1,41; 1,9)$  при значениях параметров  $\{-2; -1\}$ .

Представленные на рис. 2 результаты можно также интерпретировать при помощи показателя средней продолжительности серии (количества точек) до выхода точки за контрольный предел карты (*average run length* — ARL):

$$ARL_{\bar{x}} \approx \frac{1}{1 - \beta_{\bar{x}}}.$$



■ Рис. 2. Зависимость оценок  $(1 - \beta_{\bar{x}})$ -вероятности выхода за  $LCL_{\bar{x}}$   $\bar{X}$ -карт от значений параметров  $\{a; b\}$  при  $n = 10$  и  $P_x = 0,72$

Например, при  $1 - \beta_{\bar{X}} > 50\%$   $\bar{X}$ -карта в среднем выявляет неуправляемую вариабельность процесса ранее, чем на второй точке (чувствительность такой карты считается высокой).

Как видно из рис. 2,  $\bar{X}$ -карта, построенная в соответствии с ПБ-методом, способна практически сразу выявить неуправляемую вариабельность при сдвиге  $\{-2; -1\}$ . Однако  $\bar{X}$ -карты, построенные в соответствии с методом Шухарта и WV-методом, не способны выявить неуправляемую вариабельность при аналогичных сдвигах в положении и изменчивости значений характеристики процесса.

Данный пример содержит практическую ценность: рассмотренные в нем условия (значения  $n$ ,  $P_x$ ,  $\{a; b\}$ , формы распределения значений  $X$  стабильного и нестабильного процессов) схожи с условиями мониторинга показателя ACD в сфере телекоммуникаций.

### Заключение

В статье подробно описаны основные подходы к построению КК для асимметрично распределенных данных. На практике асимметричные

распределения значений характеристик многих процессов описываются теоретическими законами (в некотором приближении). Для контроля таких процессов применимы параметрические методы построения КК.

Предложенный нами ПБ-метод позволяет построить КК любой статистики для любого теоретического закона распределения значений  $X$  (с одним или двумя параметрами). Для него характерны низкая вероятность совершения ошибок 1-го и 2-го рода, наличие вычислительной робастности контрольных пределов, отсутствие нелинейных преобразований исходных данных (в то время как альтернативные методы построения КК не удовлетворяют одновременно всем этим критериям). В качестве оценок внутригрупповой изменчивости (для формирования теоретической модели распределения  $X$ ) предложено использовать статистику  $(\bar{S})^2$  в фазе I и статистику  $\bar{S}^2$  в фазе II.

Предложенный нами ПБ-метод успешно используется на практике в телекоммуникационной отрасли для мониторинга характеристики ACD.

### Литература

1. Wheeler D. Advanced Topics in Statistical Process Control. — SPC Press, 1995. — 470 p.
2. Choobineh F., Ballard J. Control-Limits of QC Charts for Skewed Distributions Using Weighted-Variance // IEEE Transactions on Reliability. 1987. Vol. 36(4). P. 473–477.
3. Khoo M. B. C., Atta A. M. A., Chen C.-H. Proposed  $\bar{X}$  and  $S$  Control Charts for Skewed Distributions // Proc. of the 2009 IEEE IEEM. 2009. P. 389–393.
4. Chang Y. S., Bai D. S. Control Charts for Positively — Skewed Populations with Weighted Standard Deviations // Quality and Reliability Engineering International. 2001. Vol. 17(5). P. 397–406.
5. Chan L. K., Cui H. J. Skewness Correction  $\bar{X}$  and  $R$  Charts for Skewed Distributions // Naval Research Logistics. 2003. Vol. 50(6). P. 555–573.
6. Bajgier S. M. The Use of Bootstrapping to Construct Limits on Control Charts // Proc. of the Decision Science Institute, San Diego, CA, 1992. P. 1611–1613.
7. Jones L. A., Woodall W. H. The Performance of Bootstrap Control Charts // Journal of Quality Technology. 1998. Vol. 30(4). P. 362–375.
8. Castagliola P., Khoo M. 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.
9. Hill I., Hill R., Holder R. Fitting Johnson Curves by Moments // Applied Statistics. 1976. Vol. 25(2). P. 180–189.
10. Morrison J. The Lognormal Distribution in Quality Control // Applied Statistics. 1958. Vol. 7(3). P. 160–172.
11. Nelson R. P. Control Charts for Weibull Processes with Standards Given // IEEE Transactions on Reliability. 1979. Vol. 28(4). P. 283–298.
12. Nichols M. D., Padgett W. J. A Bootstrap Control Chart for Weibull Percentiles // Quality and Reliability Engineering International. 2005. Vol. 22(2). P. 141–151.
13. Lio Y. L., Park C. A Bootstrap Control Chart for Birnbaum–Saunders Percentiles // Quality and Reliability Engineering International. 2008. Vol. 24. P. 585–600.
14. Abbasi B., Guillen M. Bootstrap Control Charts in Monitoring Value at Risk in Insurance // Expert Systems with Applications. 2013. Vol. 40. P. 6125–6135.
15. Lukin V., Yaschenko V. Computation of Control Limits of  $\bar{X}$ -Chart Based on Pseudorandom Numbers Generation // Proc. of the IEEE Russia. North West Section. 2014. Vol. 6. P. 54–58.
16. Woodall W. H., Montgomery D. C. Research Issues and Ideas in Statistical Process Control // Journal of Quality Technology. 1999. Vol. 31(4). P. 376–386.



UDC 658.562.012.7, 519.24

doi:10.15217/issn1684-8853.2015.2.98

**A Parametric Bootstrap Method of Chart Control Limit Computation for Skewed Distributions**Lukin V. N.<sup>a, b</sup>, Researcher, Business Processes Analyst, vnlukin@gmail.comIashchenko V. V.<sup>a</sup>, PhD, Tech., Associate Professor, berater51@list.ru<sup>a</sup>Saint-Petersburg State Electrotechnical University «LETI», 5, Professora Popova St., 197376, Saint-Petersburg, Russian Federation<sup>b</sup>«Telesoft» Ltd., 25, Building «Zh», Krasnogo Kursanta St., 197110, Saint-Petersburg, Russian Federation

**Purpose:** In many processes, empirical distributions of measurements can be approximated by some (skewed) theoretical models. The parametric control charting methods are the most accurate methods in identifying the state of such process (either stable or unstable). The control charting method should meet the following criteria: low type-I and type-II error rates, robust control limits, and input data free of nonlinear transformations. However, the existing (most precise) parametric control charting methods do not meet all these criteria simultaneously. The goal of the work is to develop a parametric control charting method meeting the above-mentioned criteria, and to compare the performance of the charts for average and standard deviations built by the proposed method with the performance of similar charts built by the conventional methods during the process monitoring (in Phase II). **Results:** The proposed parametric bootstrap method modifies the authors' pseudorandom-generation based approach to the evaluation of control limits by using an unbiased estimator of a within-subgroup variation (pooled variance) at the step of evaluating the distribution parameters, hence decreasing the probability of a false alarm in the process monitoring phase (Phase II). The use of an average statistic of within-subgroup variation increases the robustness of the control limits (to the presence of exceptional variation), helping to apply the proposed method to the retrospective analysis (Phase I). The method does not require nonlinear transformations of the data which may complicate the technical interpretation and application of the analysis results. The proposed parametric bootstrap method may be used to construct a control chart for any statistic when the process measurements are distributed in accordance with any (one- or two-parameter) theoretical law. The type-I error rates of averages and standard deviation charts calculated using the proposed method are closer to the required values than the same rates of the same charts calculated using other methods. The type-II error rates of these charts calculated by the proposed method enable them to quickly detect large shifts in the process location and variance. **Practical relevance:** The developed method has a successful practical application in monitoring the average call duration (ACD) of a telecommunicational process of voice traffic transmission.

**Keywords** — Statistical Process Control, Control Chart, Evaluation of Control Limits, Parametric Bootstrap Method, Skewed Distribution, Type-I and Type-II Error Rates.

**References**

1. Wheeler D. *Advanced Topics in Statistical Process Control*. SPC Press, 1995. 470 p.
2. Choobineh F., Ballard J. Control-Limits of QC Charts for Skewed Distributions Using Weighted-Variance. *IEEE Transactions on Reliability*, 1987, vol. 36(4), pp. 473–477.
3. Khoo M. B. C., Atta A. M. A., Chen C.-H. Proposed  $\bar{X}$  and S Control Charts for Skewed Distributions. *Proc. of the 2009 IEEE IEM*, 2009, pp. 389–393.
4. Chang Y. S., Bai D. S. Control Charts for Positively — Skewed Populations with Weighted Standard Deviations. *Quality and Reliability Engineering International*, 2001, vol. 17(5), pp. 397–406.
5. Chan L. K., Cui H. J. Skewness Correction  $\bar{X}$  and R Charts for Skewed Distributions. *Naval Research Logistics*, 2003, vol. 50(6), pp. 555–573.
6. Bajgier S. M. The Use of Bootstrapping to Construct Limits on Control Charts. *Proc. of the Decision Science Institute*, San Diego, CA, 1992, pp. 1611–1613.
7. Jones L. A., Woodall W. H. The Performance of Bootstrap Control Charts. *Journal of Quality Technology*, 1998, vol. 30(4), pp. 362–375.
8. Castagliola P., Khoo M. 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), pp. 1659–1674.
9. Hill I., Hill R., Holder R. Fitting Johnson Curves by Moments. *Applied Statistics*, 1976, vol. 25(2), pp. 180–189.
10. Morrison J. The Lognormal Distribution in Quality Control. *Applied Statistics*, 1958, vol. 7(3), pp. 160–172.
11. Nelson R. P. Control Charts for Weibull Processes with Standards Given. *IEEE Transactions on Reliability*, 1979, vol. 28(4), pp. 283–298.
12. Nichols M. D., Padgett W. J. A Bootstrap Control Chart for Weibull Percentiles. *Quality and Reliability Engineering International*, 2005, vol. 22(2), pp. 141–151.
13. Lio Y. L., Park C. A Bootstrap Control Chart for Birnbaum–Saunders Percentiles. *Quality and Reliability Engineering International*, 2008, vol. 24, pp. 585–600.
14. Abbasi B., Guillen M. Bootstrap Control Charts in Monitoring Value at Risk in Insurance. *Expert Systems with Applications*, 2013, vol. 40, pp. 6125–6135.
15. Lukin V., Yaschenko V. Computation of Control Limits of  $\bar{X}$ -Chart Based on Pseudorandom Numbers Generation. *Proc. of the IEEE Russia. North West Section*, 2014, vol. 6, pp. 54–58.
16. Woodall W. H., Montgomery D. C. Research Issues and Ideas in Statistical Process Control. *Journal of Quality Technology*, 1999, vol. 31(4), pp. 376–386.