

УДК 620.179

В.П. Малайчук, Н.А. Лысенко, А.И. Федорович

КРИТЕРИИ НЕПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ КЛОТЦА И БУША-ВИНДА В ЗАДАЧАХ ПЕРИОДИЧЕСКОГО КОНТРОЛЯ ТЕХНИЧЕСКИХ ОБЪЕКТОВ

Аннотация. Исследуются возможности обнаружения изменений статистических закономерностей выборок измерений, содержащих информацию о состоянии технических объектов, с помощью критериев непараметрической статистики Клотца и Буша-Винда в условиях полного незнания законов распределения вероятностей и ограничениям на объем измерений.

Ключевые слова: выборка измерений, непараметрическая статистика, критерий, вероятностные характеристики, сдвиг, масштаб.

Постановка задачи

Наблюдение за техническими объектами, технологическими процессами, окружающей средой методами и средствами неразрушающего контроля осуществляется путем периодического измерения их параметров, прямо или косвенно содержащих информацию об их состоянии. Выборки измерений запоминаются, сравниваются между собой и по результатам сравнения готовятся данные для поддержки принятия информационных решений, отвечающих на вопрос «Что правда?» [1]. Выборки измерений, как правило, представляют собой последовательности многомерных случайных величин, статистические закономерности которых точно неизвестны и могут изменяться во времени и пространстве при наблюдении за точечными или пространственно распределенными объектами. В условиях такой априорной неопределенности для сравнения выборок измерений можно воспользоваться критериями Клотца, предназначенного для проверки гипотезы о равенстве параметров масштаба, и комбинированного Критерия Буша-Винда для проверки гипотезы о равенстве параметров сдвига и масштаба двух выборок измерений [2].

Сравнение двух выборок измерений в задачах наблюдения (мониторинга) за состоянием технических объектов и

технологических процессов имеет своей целью обнаружение изменений статистических закономерностей путем оценки параметров сдвига и масштаба. Эта задача решалась путем проведения вычислительных экспериментов на моделях случайных величин с различными законами распределения вероятности и параметрами, характеризующими сдвиг и масштаб выборок измерений. Рассмотрены реакция критериев Клотца и Буша-Винда на изменение параметров законов распределения вероятности и размеров сравниваемых выборок. Исследовались случайные величины с симметричными и асимметричными законами распределения вероятности.

Критерии Клотца и Буша-Винда в задачах проверки гипотез о равенстве параметров сдвига и масштаба

Хорошо известен критерий Ван-дер-Вардена для проверки гипотезы об отсутствии сдвига (равенстве математических ожиданий) двух выборок случайных величин с одинаковыми дисперсиями. Критерий Клотца является аналогом критерия Ван-дер-Вардена при проверке гипотезы об отсутствии изменений масштаба (дисперсий) двух выборок x_{1i} и x_{2i} , $i=1,2,\dots,n$ с равными и известными математическими ожиданиями.

Критерий Клотца вычисляется по формуле

$$L = \sum_{i=1}^n \Psi^2 \left[\frac{R(x_{2i})}{2n+1} \right], \quad R(x_{2i}) = \sum_{j=1}^n \text{sign}(x_{2i} - \xi_j),$$

где $\Psi(z)$ - функция, обратная интегралу вероятности Гаусса, $R(x_{2i})$ - ранг измерения x_{2i} в общем упорядоченном по возрастанию ряду $\xi_1 < \xi_2 < \dots < \xi_i < \dots < \xi_{2n}$ двух исследуемых выборок, $\text{sign}(z)$ - функция единичного скачка. Для обратной функции можно использовать аппроксимацию

$$\Psi \left[\frac{R(x_{2i})}{2n+1} \right] = 4,91 \left[\left(\frac{R(x_{2i})}{2n+1} \right)^{0.14} - \left(1 - \frac{R(x_{2i})}{2n+1} \right)^{0.14} \right].$$

При равенстве дисперсий исследуемых выборок и $n > 10$ закон распределения L как случайной величины удовлетворительно аппроксимируется нормальным с математическим ожиданием $M[L]$ и дисперсией $D[L]$

$$M[L] = \frac{1}{2} \sum_{i=1}^{2n} \Psi^2 \left(\frac{i}{2n-1} \right),$$

$$D[L] = \frac{1}{2(2n-1)} \sum_{i=1}^{2n} \left(n \Psi^4 \left(\frac{i}{2n-1} \right) - \Psi^2 \left(\frac{i}{2n-1} \right) \right).$$

Предполагается, что критерий Клотца оптимален для симметричных распределений, близких к нормальному. Если провести нормировку критерия Клотца $l = \frac{L - M[L]}{\sqrt{D[L]}}$, то решение о

равенстве дисперсий двух выборок принимается, если модуль нормированного критерия Клотца удовлетворяет неравенству $|l| < \Psi \left(\frac{1+P}{2} \right)$, где $1-P = P_o$ - допустимая вероятность принятия ошибочного решения.

Критерий Буша-Винда, предназначенный для проверки гипотезы о равенстве неизвестных математических ожиданий и дисперсий двух выборок измерений, является комбинацией аналогов критериев Ван-дер-Вардена S и Клотца T

$$S = \sqrt{2 \left(2 - \frac{1}{n} \right)} \frac{\sum_{i=1}^n \Psi \left[\frac{R(x_{2i})}{2n+1} \right]}{\sqrt{\sum_{i=1}^{2n} \Psi^2 \left[\frac{i}{m+1} \right]}},$$

$$T = \sqrt{2 \left(2 - \frac{1}{n} \right)} \frac{\sum_{i=1}^n \Psi^2 \left(\frac{R(x_{2i})}{2n+1} \right) - \sum_{i=1}^{2n} \Psi^2 \left(\frac{i}{2n+1} \right)}{\sqrt{\sum_{i=1}^{2n} \Psi^4 \left(\frac{i}{2n+1} \right)}}.$$

Критерий W равен комбинации S и T вида

$$W = -2 \ln \left[2 \left(1 - \Phi(|S|) \right) - 2 \ln \left[2 \left(1 - \Phi(|T|) \right) \right] \right],$$

где $\Phi(x)$ - интеграл вероятности Гаусса, для вычисления которого можно применить аппроксимацию

$$\Phi^*(z) = 1 - 0.852 \exp \left[- \left(\frac{z + 1.5774}{2.0637} \right)^{2.34} \right], \quad z \geq 0, \quad \Phi^*(-z) = 1 - \Phi^*(z).$$

Показатели Буша-Винда W являются случайными величинами и при $n > 30$ справедлива модель их статистических закономерностей законом распределения хи-квадрат с четырьмя степенями свободы, если математическое ожидания и дисперсия двух сравниваемых выборок равны. В этом случае с вероятностью $P = 0,95$ будет выполняться неравенство $W \leq 9,25$ ($n > 30$) и $W \leq 9,32$ ($n > 50$). Путем проведения вычислительных экспериментов исследовано влияние объема измерений и видов законов распределения вероятностей двух выборок с равными математическими ожиданиями и дисперсиями на статистические закономерности критериев Клотца и Буша-Винда.

Для проведения исследований была разработана компьютерная модель формирования выборок измерений с симметричными и асимметричными законами распределения вероятностей с заданными параметрами, оценкой критериев Клотца и Буша-Винда и их статистического анализа. Установлено, что если две исследуемые выборки случайных величин имеют одинаковые законы распределения вероятностей и одинаковые параметры сдвига и масштаба, то статистические закономерности критериев Клотца и Буша-Винда соответствуют выводам их авторов: нормированный критерий Клотца имеет нормальное распределение с нулевым математическим ожиданием и единичной дисперсией, а критерий Буша-Винда – закон распределения хи-квадрат с четырьмя степенями свободы (гистограммы на рис.1).

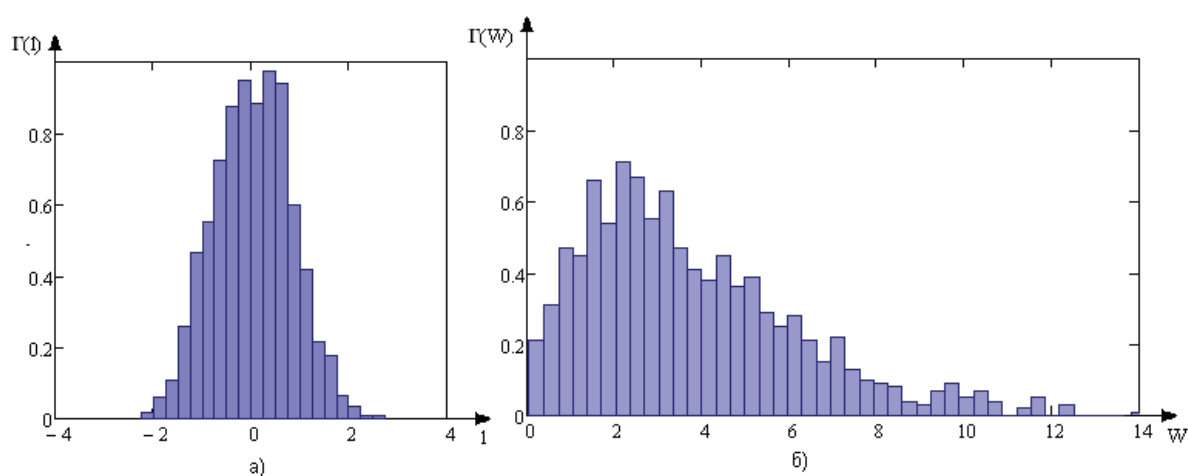


Рис. 1 – Гистограммы критерия Клотца (а) и Буша-Винда (б) для однородных выборок измерений с одинаковым сдвигом и масштабом

Такие же распределения получены для других симметричных (логистического, Лапласа) и асимметричных (экспоненциального,

Релея) законов распределения исследуемых выборок. Однако путем исследования двух выборок измерений с одинаковыми сдвигами и масштабами, но отличающиеся видами их законов распределения (нормальный и логистический, нормальный и Лапласа), было установлено, что статистические закономерности критериев Клотца и Буша-Винда существенно отличаются от теоретических. На рис. 2 представлены гистограммы критерия Буша-Винда при сравнении двух выборок с одинаковыми математическими ожиданиями и дисперсиями с законами распределения нормальным и логистическим, нормальным и Лапласа.

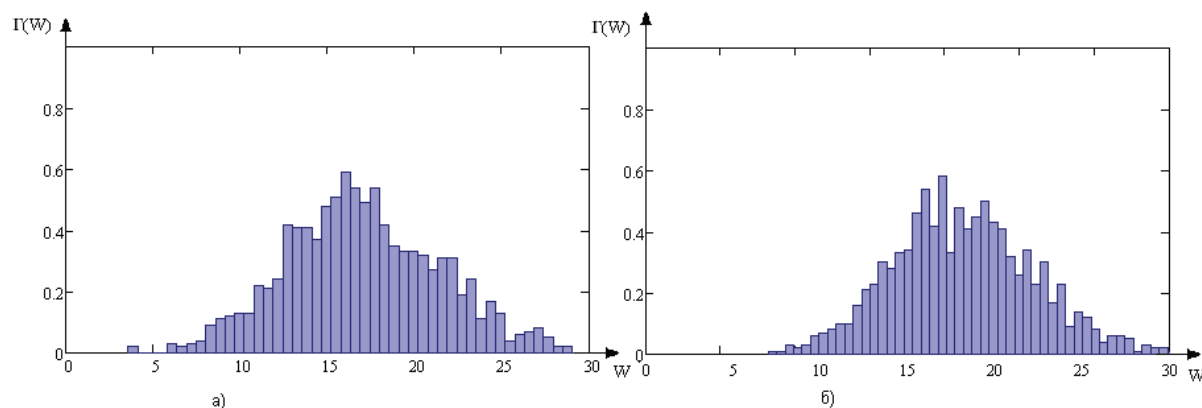


Рис. 2 – Гистограммы критерия Буша-Винда при сравнении выборок с нормальным и логистическим распределениями (а) и нормальным распределением и распределением Лапласа (б)

Таким образом, критерий Клотца и Буша-Винда являются критериями сдвига и масштаба только при сравнении выборок с одинаковыми видами законов распределения, то есть это критерии однородности выборок измерений случайных величин, и могут использоваться для обнаружения изменений статистических закономерностей выборок измерений при полном отсутствии информации о законах распределения вероятности и их параметрах.

Оценка эффективности критерия Буша-Винда в задачах обнаружения изменений статистических закономерностей выборок случайных величин

Исследовалась эффективность критерия Буша-Винда путем сравнения решающих правил распознавания, оптимальных в смысле правила Неймана-Пирсона. Рассматривались две выборки нормальных случайных величин с различными математическими ожиданиями и равными дисперсиями (a_1, a_2, σ^2) . Оптимальное

решающее правило обнаружения сдвига второй выборки ($a_1 < a_2$) по правилу Неймана-Пирсона запишется в виде

$$z = \frac{1}{n} \sum_{k=1}^n x_k \geq a_1 + \frac{\sigma}{\sqrt{n}} \Psi(1 - P_{op}),$$

где $\Psi(x)$ -функция, обратная интегралу вероятности Гаусса, P_{op} - допустимое значение вероятности ошибочного решения, если сдвиг отсутствует ($a_1 = a_2$). Если $a_1 \neq a_2$, то вероятность обнаружения сдвига второй выборки P_{oc} определяется по формуле [3]

$$P_{oc} = \Phi \left[\frac{\sqrt{n} \Delta a}{\sigma} - \Psi(1 - P_{op}) \right],$$

где $\Phi(x)$ -интеграл вероятности Гаусса, $\Delta a = a_2 - a_1$.

Решающее правило обнаружения сдвига второй выборки относительно первой при использовании критерия Буша-Винда получим на основании следующих знаний. Если выборки не имеют сдвига и у них равные дисперсии, то критерий Буша-Винда при $n > 20$ это случайная величина с законом распределения хи-квадрат с четырьмя степенями свободы. Порог сравнения W_0 в этом случае зависит только от величины допустимой вероятности принятия оптимального решения о сдвиге. При $P_{op} = 0,05$ порог равен 9,5 то есть при отсутствии сдвига ($a_1 = a_2$) для критерия Буша-Винда с вероятностью 0,95 справедливо неравенство $0 \leq W \leq 9,5$. Решающее правило обнаружения сдвига запишется в виде неравенства $W > W_0 = 9,5$. На рис.3а оптимальные и эмпирические характеристики обнаружения сдвига при различных размерах выборок измерений.

Аналогично решалась задача обнаружения изменения масштаба и сдвига для релеевских выборок случайных величин. Оптимальное решающее правило обнаружения изменения параметра релеевского

распределения $W(x) = \frac{x}{b} e^{-\frac{x^2}{2b}}$, $0 \leq x \leq \infty$, по правилу Неймана-Пирсона

имеет вид $z = \sum_{k=1}^n x_k^2 \geq z_{01}$ [3]. Преобразуем это неравенство в

равносильное следующим образом. Разделим его на b_1 и на b_2 получим два неравенства

$$\frac{z}{b_1} = u_1 = \sum_{k=1}^n \frac{x_k^2}{b_1} \geq \frac{z_{01}}{b_1}; \frac{z}{b_2} = u_2 = \sum_{k=1}^n \frac{x_k^2}{b_2} \geq \frac{z_{01}}{b_2}.$$

Здесь u_1 и u_2 случайные величины с законами распределения хи-квадрат с $2n$ степенями свободы. Используя аппроксимацию Вилсона-Хилферти, порог сравнения по методу Неймана-Пирсона оценим по формуле

$$\frac{z_{01}}{b_1} = u_{01} = 2n \left[1 - \frac{1}{9n} + \Psi(1 - P_{op}) \sqrt{\frac{1}{9n}} \right]^3, \quad n \geq 15.$$

Вероятность обнаружения сдвига и масштаба вычисляем следующим образом

$$P_{oc} = 1 - \frac{1}{\sqrt{2}\Gamma(n)} \int_0^{y_0} y^{n-1} \exp\left(-\frac{y}{2}\right) dy,$$

где $y_0 = \frac{b_1}{b_2} 2n \left[1 - \frac{1}{9n} + \Psi(1 - P_{op}) \sqrt{\frac{1}{9n}} \right]^3, \quad b_2 \geq b_1.$

На рис. 3б представлены теоретические и эмпирические вероятностные характеристики обнаружения изменения параметра релеевского распределения.

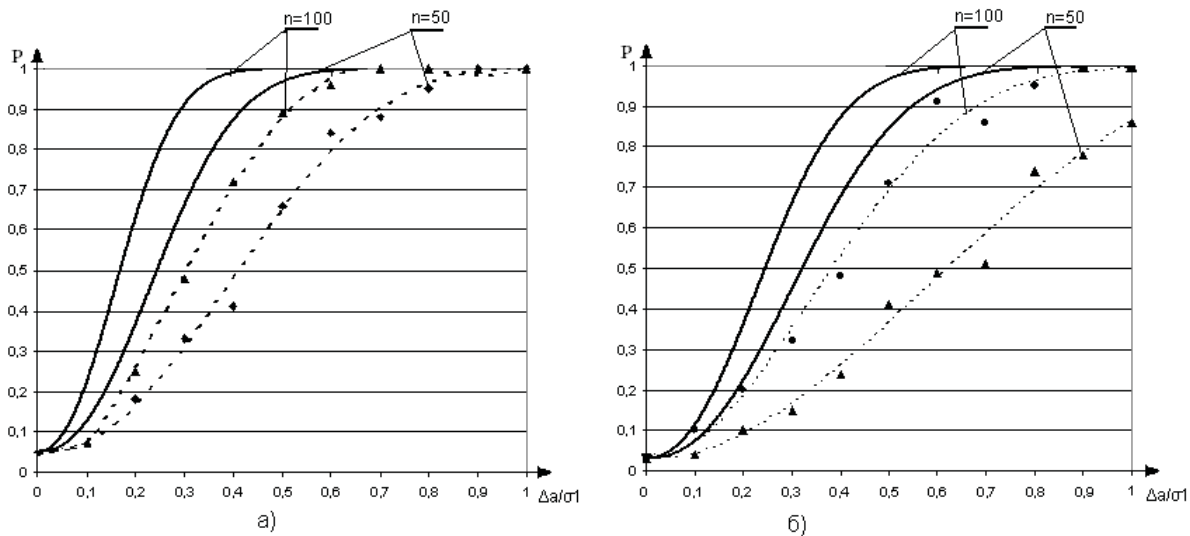


Рис. 3 – Вероятностные характеристики обнаружения изменения сдвига и масштаба

Выводы

1. Критерии Буша-Винда и Клотца являются критериями для проверки гипотез об отсутствии сдвига и равенстве масштаба только выборок с неизвестными, но одинаковыми законами распределения вероятностей.

2. Критерии Клотца и Буша-Винда относятся к классу критериев однородности и могут использоваться в задачах обнаружения изменений статистических закономерностей двух выборок измерений с неизвестными законами распределения вероятностей, используя правило Неймана-Пирсона.

3. На основе исследования вероятностных характеристик обнаружения изменений статистических закономерностей выборок измерений установлено снижение эффективности эмпирического решающего правила по сравнению с оптимальным как плату за незнание законов распределения вероятностей случайных величин: вероятность обнаружения уменьшается с 0,98 до 0,89 и 0,9 при $30 \leq n \leq 100$.

ЛИТЕРАТУРА

1. Огоренко В.В. Введение в психологию принятия решений / В.В. Огоренко, В.П. Малайчук – Днепропетровск. Системные технологии, 2009, - 150 с.
2. Кобзарь А.И. Прикладная математическая статистика/ А.И. Кобзарь. – М.: ФИЗМАТ ЛИТ, 2006. – 816 с.
3. Малайчук В.П., Мозговой А.В. Математическая дефектоскопия: Монография. – Днепропетровск: Системные технологии, 2005, - 180 с.