https://doi.org/10.21683/1729-2646-2023-23-1-52-55

Вариационный критерий равномерности
Variational criterion of evenness

Воловик А.В.[1]
Volovik A.V.[1]

[1]АО «ОДК-Климов», Санкт-Петербург, Российская Федерация
[1]UEC-Klimov JSC, Saint Petersburg, Russian Federation

volovik_aleksandr@mail.ru

Воловик А.В.

Резюме. Цель. *Разработан критерий для проверки гипотезы о равномерном законе распределения наблюдений случайных величин в выборках малого объема. Формирование критерия заключается в построении из выборочных наблюдений вариационного ряда в порядке возрастания и делении каждого предыдущего члена этого ряда на крайний член с последующим его отбрасыванием. С полученным новым вариационным рядом поступают аналогично до тех пор, пока не останется одно частное, являющееся значением критерия.* **Методы.** *В статье применяются методы теории вероятностей и математической статистики.* **Результаты.** *Предложенный критерий обладает достаточной эффективностью для различения выборок минимального объема для статистически близких гипотез, таких как гипотеза о равномерном законе распределения и гипотеза о бета-распределении 1-го рода.* **Выводы.** *Предлагаемый в статье подход позволяет достаточно просто реализовать процедуру последовательного анализа (обнаружение «разладки» процесса). Такая процедура позволяет выявлять «разладку» (отклонение распределения наблюдений от равномерного закона) с достаточной для практики интенсивностью.*

Abstract. Aim. *The author has developed a criterion to test the hypothesis of a uniform distribution for random variable observations in small samples. The criterion is built by using sample observations to construct a variation series in ascending order and dividing each previous term of this series by the extreme term, then discarding it. The resulting new variation series is processed similarly until there is only one quotient left that is the criterion value.* **Methods.** *The paper uses methods of the probability theory and mathematical statistics.* **Results.** *The suggested criterion is sufficiently efficient for distinguishing between samples of minimal size for statistically similar hypotheses, such as the hypothesis of a uniform distribution law and the hypothesis of a beta distribution of the first kind.* **Conclusions.** *The approach suggested in the paper makes it quite simple to implement the sequential analysis procedure (detection of a “dissonance” in a process). Such a procedure allows detecting a “dissonance” (deviation of the distribution of observations from the uniform law) with a practically sufficient rate.*

Ключевые слова: *вариационный ряд, плотность распределения, статистика, гипотеза, достигаемый уровень значимости, последовательный анализ.*
Keywords: *ordered sample, frequency distribution, statistics, hypothesis, achievable significance level, sequential analysis.*

Для цитирования: *Воловик А.В. Вариационный критерий равномерности // Надежность. 2023. №1. С. 52-55. https://doi.org/10.21683/1729-2646-2023-23-1-52-55*
For citation: *Volovik A.V. Variational criterion of evenness. Dependability 2023;1:52-55. https://doi.org/10.21683/1729-2646-2023-23-1-52-55*

Поступила *21.10.2022/* **После доработки** *12.12.2022* / **К печати** *14.03.2023*
Received on: *21.10.2022/* **Revised on:** *12.12.2022* / **For printing:** *14.03.2023*

Введение

Множество приложений, в которых сталкиваются с использованием модели равномерного закона распределения вероятностей, объясняет повышенный интерес к выбору простых в вычислительном отношении и эффективных критериев проверки гипотез о принадлежности анализируемых выборок равномерному закону [1]. Таких критериев разработано достаточно много. Однако при обработке малых выборок наблюдений, которые характеризуют высоконадежные или уникальные изделия, большинство из них дают неоднозначный результат. Для повышения достоверности используют мультипликативное объединение известных критериев [2]. При этом усложняется процедура проверки.

Метод

Сложность постановки и решения задач построения наилучших критериев при данном объеме статистического материала обусловлена тем обстоятельством, что искомое решение часто в сильной степени зависит от конкретного типа распределения, объема выборки и не может быть объектом достаточно общей математической теории [3]. Одним из способов формирования критерия для проверки равномерности распределения случайных величин является использование неубывающего вариационного ряда [4] для построения статистики критерия.

Для наблюдаемых значений, составляющих случайную выборку $\{x_1, x_2, ..., x_n\}$ независимых непрерывных случайных величин из одной генеральной совокупности, при сортировке в неубывающем порядке получен вариационный ряд $x_{(1)} \le ... \le x_{(k)} \le ... \le x_{(n)}$, члены которого представляют собой наблюдаемые значения $(x_{(1)}, ..., x_{(k)}, ..., x_{(n)})$ порядковой статистики $(X_{(1)}, ..., X_{(k)}, ..., X_{(n)})$. Разделив первые $n-1$ членов вариационного ряда на $x_{(n)}$, получим вариационный ряд

$$\frac{x_{(1)}}{x_{(n)}} \le \ldots \le \frac{x_{(k)}}{x_{(n)}} \le \ldots \le \frac{x_{(n-1)}}{x_{(n)}}. \qquad (1)$$

Аналогичным образом из ряда (1), можно получить ряд

$$\frac{x_{(1)}}{x_{(n-1)}} \le \ldots \le \frac{x_{(k)}}{x_{(n-1)}} \le \ldots \le \frac{x_{(n-2)}}{x_{(n-1)}}. \qquad (2)$$

Продолжая далее ...

$$\frac{x_{(1)}}{x_{(n-m)}} \le \ldots \le \frac{x_{(k)}}{x_{(n-m)}} \le \ldots \le \frac{x_{(n-m-1)}}{x_{(n-m)}}, \qquad (3)$$

… получим в итоге безразмерную статистику

$$\upsilon = \frac{x_{(1)}}{x_{(2)}}, \qquad (4)$$

которая полностью характеризует исходную выборку $\{x_1, x_2, ..., x_n\}$, т.к. в ее формировании участвовали все n наблюдений путем последовательных делений (1), (2),…, (3),… и соответствующих сокращений. И, что характерно, из всей совокупности наблюдаемых значений $(x_{(1)}, ..., x_{(k)}, ..., x_{(n)})$ достаточно двух наименьших наблюдений $x_{(1)}$ и $x_{(2)}$.

Таким образом, для выборки любого размера статистика υ представляет собой частное от деления первой наимéньшей из n независимых одинаково распределенных случайных величин на вторую наимéньшую. Докажем для этого случая следующую теорему [5].

Теорема. Пусть дана выборка, состоящая из двух независимых случайных величин $\{x_1, x_2\}$ равномерно распределенных в интервале [0;1], и пусть из них составлен вариационный ряд $x_{(1)} \le x_{(2)}$. Тогда отношение $\upsilon = x_{(1)}/x_{(2)}$ распределено равномерно в интервале [0;1].

Доказательство. Плотность вероятности совместного распределения упорядоченных случайных величин $x_{(1)} \le x_{(2)}$, изначально независимых равномерно распределенных в интервале [0;1], запишется следующим образом [6]

$$f_{x_{(1)},x_{(2)}}\left(x_{(1)}, x_{(2)}\right) = 2!\prod_{i=1}^{2} f_{x_{(i)}}\left(x_{(i)}\right) = 2!, \qquad (5)$$

где $f_{x_{(i)}}\left(x_{(i)}\right)$ – плотность распределения i-й порядковой статистики.

Введем в рассмотрение две статистики (по числу членов вариационного ряда)

$$\upsilon_1 = \frac{x_{(1)}}{x_{(2)}} \quad \text{и} \quad \upsilon_2 = x_{(2)}. \qquad (6)$$

Так как обратные преобразования $x_{(1)} = \upsilon_1\upsilon_2$ и $x_{(2)} = \upsilon_2$ случайных величин (6) взаимно однозначны, то плотность совместного распределения

$$f_{\upsilon_1,\upsilon_2}(\upsilon_1, \upsilon_2) = f_{x_{(1)},x_{(2)}}\left(x_{(1)}, x_{(2)}\right)_{x_{(i)} = x_{(i)}(\upsilon_1, \upsilon_2)} \cdot |J|, \qquad (7)$$

где $J = \dfrac{\partial\left(x_{(1)}, x_{(2)}\right)}{\partial(\upsilon_1, \upsilon_2)}$ – якобиан.

Тогда, с учетом (5) плотность совместного распределения (7) запишется следующим образом

$$f_{\upsilon_1,\upsilon_2}(\upsilon_1, \upsilon_2) = 2!\,\upsilon_2. \qquad (8)$$

Исключая вспомогательную переменную υ_2 путем интегрирования выражения (8) по области значений υ_2, получим плотность переменной υ_1

$$f_{\upsilon_1}(\upsilon_1) = \int_0^1 f_{\upsilon_1,\upsilon_2}(\upsilon_1, \upsilon_2)\,d\upsilon_2 = \int_0^1 2\upsilon_2\,d\upsilon_2 = 1! \qquad (9)$$

Результат (9) свидетельствует о равномерном законе распределения переменной υ_1 в интервале [0;1].

На рис. 1 приведен результат статистической проверки гипотезы о равномерном распределении отношения (4) при объеме выборки $n=2$ и числе реализаций $N=5000$ с помощью стандартного пакета STATISTICA.

На данном и последующих рисунках над гистограммой выведен заголовок, в котором указана анализируемая переменная Var1, предполагаемый закон распределения (Rectangular – равномерный), значение критерия

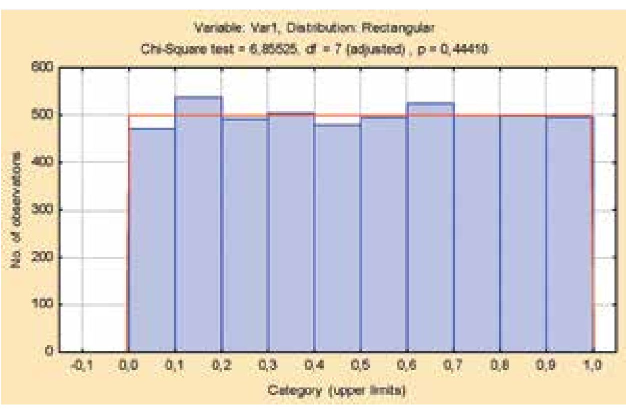

Рис. 1. Проверка гипотезы о равномерном законе распределения статистики υ при $n=2$

χ^2=6,86, число степеней свободы df = 7 и p-уровень значимости критерия, который определяет вероятность ошибки при отклонении гипотезы от равномерного закона. Так как вероятность ошибки достаточно велика, примерно 0,44 (что значительно больше 0,05), отвергнуть гипотезу о соответствии закона распределения равномерному оснований нет [7].

Таким образом, равномерность распределения статистики (4) критерия подтверждена статистически. При этом независимо от объема n исходной выборки из генеральной совокупности с равномерным законом распределения, статистика (4) будет распределена равномерно в интервале [0;1].

На рис. 2 приведен результат статистической проверки гипотезы о равномерном распределении отношения (4) при объеме выборки n=10 и числе реализаций N=5000.

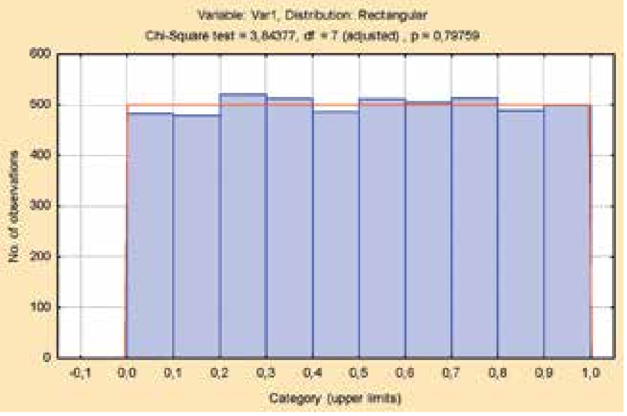

Рис. 2. Проверка гипотезы о равномерном законе распределения статистики υ при n=10

Так как вероятность ошибки при отклонении гипотезы от равномерного распределения достаточно велика, примерно 0,8 (что значительно больше 0,05), то отвергнуть гипотезу о соответствии закона распределения равномерному также оснований нет.

Результат. Для оценки эффективности критерия в качестве альтернативы равномерному закону (гипотеза H_0) рассмотрим гипотезу H_1 о бета-распределении 1-го рода с параметрами $\alpha = \beta$= 1,5 на интервале [0;1], как это рассматривалось в [1]. Заметим, что при $\alpha = \beta = 1,0$ бета-распределение вырождается в равномерное.

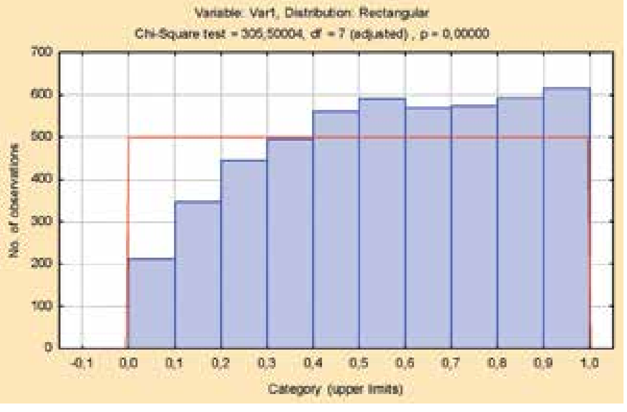

Рис. 3. Проверка гипотезы о равномерном распределении статистики υ для H_1 при n=2

На рис. 3 приведен результат проверки гипотезы о равномерном распределении статистики υ критерия для H_1 при объеме выборки n=2.

Поскольку вероятность ошибки при отклонении гипотезы о равномерном законе p = 0, что меньше значения 0,05, то имеются основания отвергнуть гипотезу о соответствии закона равномерному распределению.

На рис. 4 приведена проверка гипотезы о равномерном распределении статистики υ критерия для H_1 при объеме выборки n=10.

Рис. 4. Проверка гипотезы о равномерном распределении статистики υ для H_1 при n=10

Поскольку вероятность ошибки при отклонении гипотезы о равномерном законе p = 0, что меньше значения 0,05, то имеются основания отвергнуть гипотезу о соответствии закона равномерному распределению.

Таким образом, предлагаемый критерий позволяет различать достаточно близкие гипотезы H_0 и H_1, начиная с выборки минимального объема. Количественная оценка мощности критерия требует отдельного исследования.

Использование υ-статистики позволяет достаточно просто организовать процедуру последовательного анализа [8] следующим образом. Из первых двух наблюдений x_1 и x_2 составляется вариационный ряд $x_{(1)}<x_{(2)}$ и рассчитывается значение статистики $\upsilon=x_{(1)}/x_{(2)}$. Это значение сравнивается с критическим значением $\upsilon_{кр}$. Если $\upsilon \geq \upsilon_{кр}$, то наблюдения прекращают т.к. наступила «разладка» процесса. Если $\upsilon<\upsilon_{кр}$, то наблюдения продолжают. При этом $x_1=\upsilon$ т.к. случайная величина υ также распределена равномерно в интервале [0;1], а $x_2=x_3$, где x_3 – следующее наблюдение.

Далее снова составляется вариационный ряд $x_{(1)}<x_{(2)}$ и рассчитывается значение статистики $\upsilon=x_{(1)}/x_{(2)}$. Новое значение сравнивается с критическим значением $\upsilon_{кр}$ и т.д. до тех пор, пока не наступит событие $\upsilon \geq \upsilon_{кр}$, при котором наблюдения прекращают, или они закончатся в выборке.

Заключение

Необходимо отметить, что практическая применимость предлагаемого критерия может быть существенно расширена при использовании вероятностного интегрального преобразования [6] $y=F_x(x)$, где $F_x(x)$ – функция распределения случайной величины x. Случайная величина y подчинена равномерному в интервале [0;1] закону распределения. В этом случае проверка гипотезы $H_0:F_n(x)=F_x(x)$ равносильна проверке гипотезы $H_0:F_y(y)=R(0;1)$, где $F_n(x)$ и $F_x(x)$– эмпирическая и теоретическая функции распределения случайной величины x; $F_y(y)$ – эмпирическая функция распределения случайной величины y; $R(0;1)$ – функция равномерного в интервале [0;1] закона распределения.

Библиографический список

1. Лемешко Б.Ю., Блинов П.Ю. Критерии проверки отклонения распределения от равномерного закона. Новосибирск: Изд-во НГТУ, 2015. 182с.

2. Иванов А.И., Куприянов Е.Н. Синтез новых более мощных статистических критериев через мультипликативное объединение классических критериев Фроцини и Мурота-Такеучи с критерием Херста для проверки гипотезы нормальности малых выборок // Надежность. 2022. Т. 22. №1. С. 52–55. DOI: 10.21683/1729-2646-2022-22-1-52-55

3. Ивченко Б.П., Мартыщенко Л.А., Табухов М.Е. Управление в экономических и социальных системах. Системный анализ. Принятие решений в условиях неопределенности. СПб.: «Нордмед-Издат», 2001. 248 с.

4. Лемешко Б.Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. Новосибирск: Изд-во НГТУ, 2011. 888 с.

5. Воловик А.В., Ефименко С.В., Клавдиев А.А., Клавдиев И.А. Вероятностно-статистическое обоснование метода выбора варианта приемочного контроля изделий // Ежемесячный научный журнал Международного независимого института Математики и Систем. 2014. № 8. С. 21.

6. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. Методы обработки данных. М.: Мир, 1980. 610 с.

7. Халафян А.А. STATISTICA 6. Статистический анализ данных. М.: ООО «Бином-Пресс», 2007. 512 с.

8. Житлухин М.В. Последовательные методы проверки статистических гипотез и обнаружения разладки: дис. ... канд. физ.-мат. наук : 01.01.05 / М.В. Житлухин ; Рос. акад. наук, Мат. ин-т им. В.А. Стеклова ; науч. рук. А.Н. Ширяев. Москва, 2013. 99 с.

References

1. Lemeshko B.Yu., Blinov P.Yu. [Criteria for testing a distribution for deviation from a uniform law]. Novosibirsk: NSTU; 2015. (in Russ.)

2. Ivanov A.I., Kupriyanov E.N. Synthesis of new, more powerful statistical tests through multiplicative clustering of classical Frozini and Murota-Takeuchi tests with the Hurst test for the purpose of testing small samples for normality. *Dependability* 2022;1:52-55. DOI: https://doi.org/10.21683/1729-2646-2022-22-1-…

3. Ivchenko B.P., Martyshchenko L.A., Tabukhov M.E. [Control in economic and social systems. Systems analysis. Decision-making under uncertainty]. Saint Petersburg: Nordmed-Izdat; 2001.

4. Lemeshko B.Yu., Lemeshko S.B., Postovalov S.N., Chimitova E.V. [Statistical analysis of data, simulation and research of probabilistic laws. A computer-based approach: a monograph]. Novosibirsk: NSTU Publishing; 2011. (in Russ.)

5. Volovik A.V., Yefimenko S.V., Klavdiev A.A., Klavdiev I.A. Probabilistic and statistical substantiation of the method for selecting a product acceptance testing procedure. Novosibirsk: MiS International Independent Institute of Mathematics and Systems 2014;8:21. (in Russ.)

6. Johnson N., Leone F. Statistics and Experimental Designs and Engineering and the Physical Sciences. Mathods of Data Processing. Moscow: Mir; 1980.

7. Khalafian A.A. [STATISTICA 6. Statistical data analysis]. Moscow: OOO Binom-Press; 2007. (in Russ.)

8. Zhitlukhin M.V. [Sequential methods of verifying statistical hypotheses and detecting dissonance; a dissertation]. Moscow: Steklov Mathematical Institute of RAS; 2013. (in Russ.)

Сведения об авторе

Воловик Александр Васильевич – кандидат технических наук, ведущий инженер-конструктор АО «ОДК-Климов», Санкт-Петербург, Российская Федерация, тел. +7-951-651-83-39, e-mail: volovik_aleksandr@mail.ru

About the author

Alexander V. Volovik, Candidate of Engineering, Lead Design Engineer, JSC “UEC-Klimov”, Saint Petersburg, Russian Federation, tel. +7 951 651 83 39, e-mail: volovik_aleksandr@mail.ru

Вклад автора в статью

Воловик А.В. Предложен критерий согласия на основе вариационного ряда, составленного из наблюдений в выборке малого объема. Проведены эксперименты по статистической оценке эффективности критерия по отношению к близкой альтернативе при разных объемах выборки, начиная с минимальной. Сделан вывод о целесообразности использования критерия для последовательного анализа и проверки гипотез о других законах распределения с помощью вероятностного интегрального преобразования.

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.