О ПРИМЕНЕНИИ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ СОГЛАСИЯ К ПРОВЕРКЕ ГИПОТЕЗ ОТНОСИТЕЛЬНО ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Шашок Г.А., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: shashok.galina@gmail.com

С анализом дискретных случайных величин сталкиваются при моделировании транспортных потоков, процессов в системах связи, в различных приложениях теории массового обслуживания, при статистическом контроле качества. Несмотря на то, что с дискретными случайными величинами часто имеют дело в приложениях, в литературе довольно сложно найти рекомендации по применению критериев согласия в дискретном случае.

Проверка адекватности закона распределения вероятностей наблюдаемой выборке осуществляется с использованием критериев согласия. В случае непрерывных случайных величин проверка гипотезы о принадлежности выборки закону с функцией распределения $F(x,\theta)$ может осуществляться с использованием критериев типа χ^2 или целого ряда непараметрических критериев согласия, например, Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга. При этом следует различать проверку простых и сложных гипотез.

Применение критериев согласия к анализу выборок дискретных случайных величин сопряжено с определенными проблемами. Очевидно, что нет принципиальных трудностей для использования критерия χ^2 Пирсона. А вот непосредственное применение непараметрических критериев согласия Колмогорова, Крамера-Мизеса-Смирнова, Андерсона-Дарлинга невозможно в силу очевидных причин, связанных с видом классических статистик.

Один из подходов, который позволяет применять классические критерии, заключается в использовании преобразования Смирнова и последующей рандомизации, в результате которых выборка дискретной величины, принадлежащей $F(x,\theta)$, преобразуется в выборку непрерывной, равномерно распределенной на интервале (0,1). А далее классический критерий применяется для проверки согласия полученной выборки с равномерным законом. Такой подход хорошо изложен в [1].

Второй подход связан с построением модификаций непараметрических критериев согласия к проверке статистических гипотез по дискретным данным [2]. Например, статистика модифицированного критерия Крамера-Мизеса-Смирнова имеет вид

$$S = n^{-1}\sum_{i=1}^{k} Z_i^2 p_i , \tag{1}$$

где $Z_i = \sum_{j=1}^{i}(O_j - E_j)$, O_j, E_j – наблюдаемое и ожидаемое соответственно число попаданий в j интервал, p_j – вероятность попадания в интервал, k - количество интервалов, n – объем выборки.

В данной работе методами статистического моделирования исследовалась сходимость распределений статистик модифицированных непараметрических критериев согласия к предельным законам в зависимости от объемов выборок и других факторов, в частности, от значений параметров законов.

В качестве примера, приведены результаты исследования сходимости распределения модифицированной статистики Крамера-Мизеса-Смирнова к предельному распределению $a1(x)$ при проверке простой гипотезы относительно распределения Пуассона $P\{\xi(\omega)=k\}=\dfrac{\lambda^k}{k!}\cdot e^{-\lambda}$. Объем выборки n варьировался от 10 до 1000, параметр λ – от 10 до 1000. Выборки статистик ограничились объемом $N=10^4$.

Группирование выборки осуществлялось по заданным граничным точкам. Ситуация на приводимых рисунках соответствует количеству интервалов k = 5.

На рисунках 1, 2 показана сходимость эмпирических распределений $G(S_n|H_0)$ статистики Крамера-Мизеса-Смирнова при справедливости проверяемой гипотезы H_0 к теоретическому предельному закону.

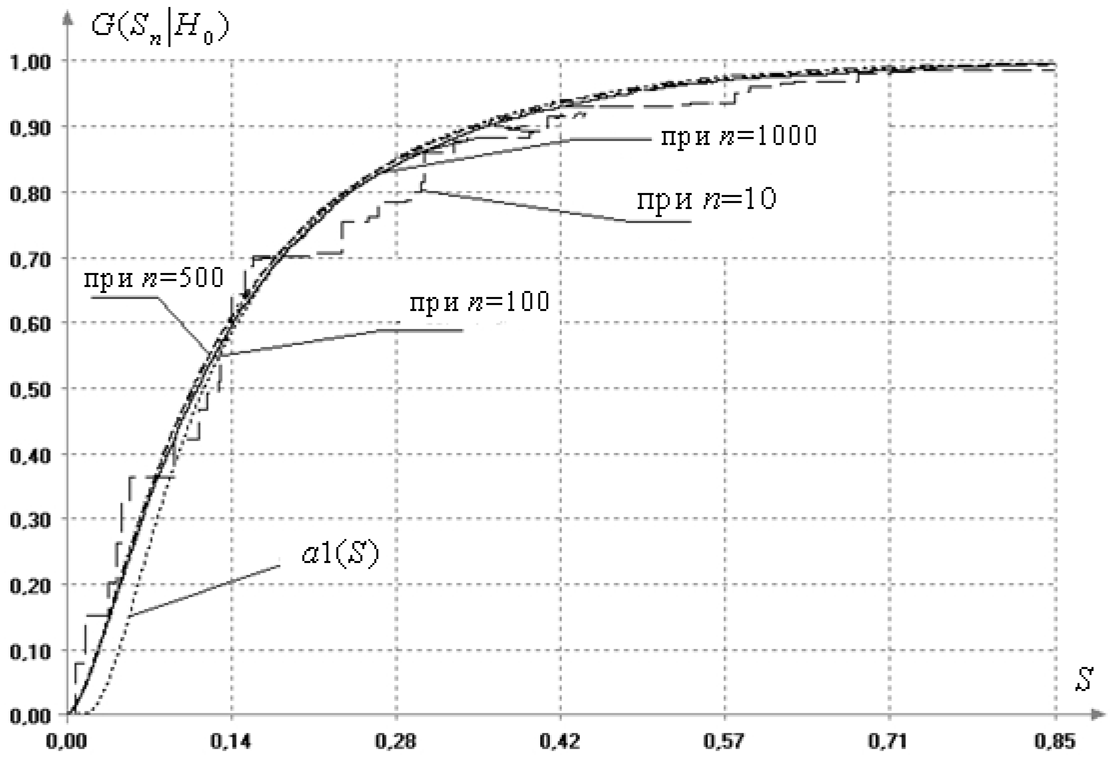

Рис. 1. Сходимость распределения статистики Крамера-Мизеса-Смирнова к предельному закону в зависимости от объема выборок $\lambda=10$

По рисунку 1 хорошо видно, что при малых объемах выборок распределение статистики обладает существенной дискретностью.

К сожалению, как видно на рисунке 2, распределение статистики критерия (его сходимость к предельному) зависит от параметра закона распределения. Наибольшая близость эмпирического распределения статистики к предельному $a1(x)$ наблюдается при значении параметра распределения Пуассона λ = 10.

Распределение статистики зависит от числа интервалов. При увеличении числа интервалов группирования (k=5; 10; 15) и достаточных объемах выборок эмпирические распределения статистики Крамера-Мизеса-Смирнова сходятся к распределению $a1(x)$.

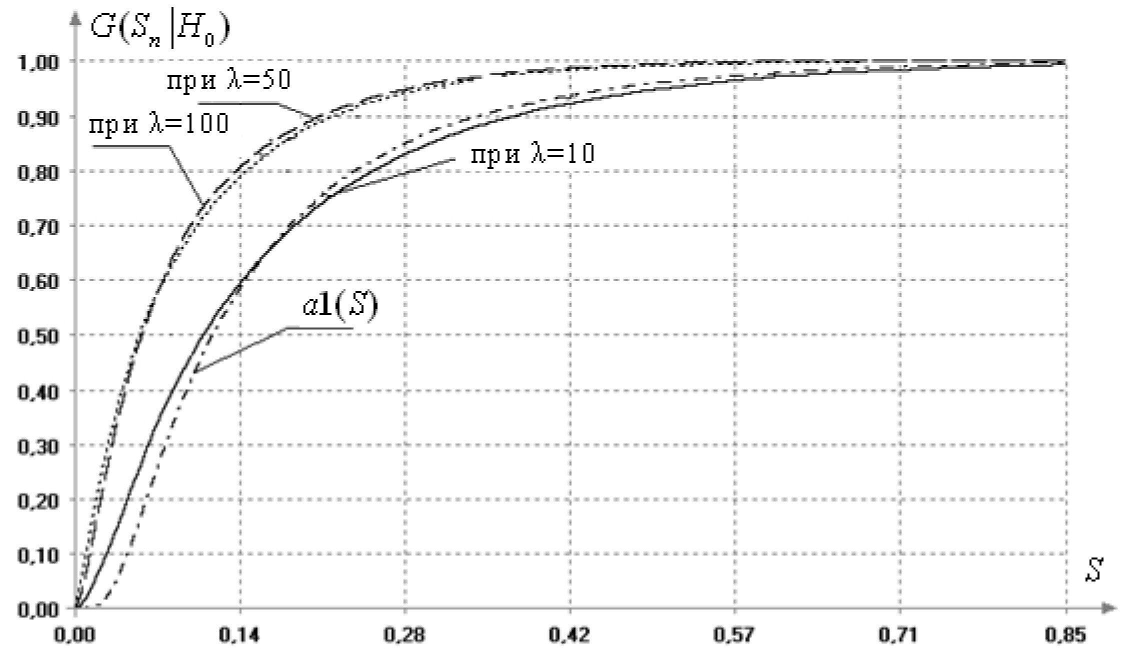

Рис. 2. Сходимость распределения статистики Крамера-Мизеса-Смирнова к предельному закону в зависимости от параметра λ, $n = 1000$

По полученным результатам можно сделать вывод, что распределение статистики Крамера-Мизеса-Смирнова зависит от закона распределения, а также от количества интервалов группирования. При большом количестве интервалов распределение модифицированной статистики (1) критерия приближаются к предельному распределению статистики. В этом случае при проверке гипотез о согласии получаемых в результате моделирования эмпирических распределений статистики (1) с предельным законом $a1(x)$ не было оснований для их отклонения гипотезы по всем применяемым критериям согласия [3].

Аналогичные результаты были получены при исследовании распределений статистики критерия Андерсона-Дарлинга.

Литература

1. Greenwood P. E., Nikulin M. S. A Guide to Chi-Squared Testing. N. Y.: John Wiley & Sons, 1996.
2. Steele M., Chaseling J. Powers of Discrete Goodness-of-Fit Test Statistics for a Uniform Null Against a Selection of Alternative Distributions // Communications in Statistics – Simulation and Computation, 2006, 35: 1067–1075.
3. 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.