СБОРНИК **НАУЧНЫХ ТРУДОВ НГТУ. – 2012. – № 1(67) –** 107–114

УДК 519.24

ИССЛЕДОВАНИЕ МОЩНОСТИ КРИТЕРИЯ ПОКАЗАТЕЛЬНОСТИ БОЛЬШЕВА*

Б.Ю. ЛЕМЕШКО, А.П. РОГОЖНИКОВ

Рассматривается критерий Л.Н. Большева для проверки гипотезы о показательном распределении наблюдений в нескольких малых выборках. Методами статистического моделирования исследуется мощность критерия при использовании в качестве вспомогательных различных критериев согласия, проводится сравнение с мощностью критериев согласия.

Ключевые слова: показательное распределение, мощность, критерий Большева.

ВВЕДЕНИЕ

В теории надежности, статистическом управлении качеством, в задачах массового обслуживания среди множества других задач рассматривается задача о проверке вида закона распределения по малым выборкам. В частности, требуется проверить, принадлежат ли случайные величины в некоторой совокупности выборок показательному семейству распределений. Один из подходов к проверке такой гипотезы дается в работе Л.Н. Большева [1], причем предложенный критерий не требует определения оценок параметров распределений случайных величин в малых выборках. Наш интерес к данному критерию пробудил ученик Л.Н. Большева профессор М.С. Никулин, с одной стороны, убежденный в полезности этого критерия, а с другой – удивленный, что в литературе нет упоминаний о практическом его использовании.

1. КРИТЕРИЙ БОЛЬШЕВА

Критерий показательности Большева предназначен для проверки гипотезы о принадлежности совокупности малых выборок показательному закону распределения вероятностей.

Пусть $\xi_{i1},\ldots,\xi_{in_i}$ $\left(n_i \geq 2;\ i=\overline{1,N}\right)$ – взаимно независимые случайные величины. Требуется проверить гипотезу H_0: величины ξ_{ij} подчиняются показательным распределениям с плотностями вероятностей $a_i \exp\left(-a_i x\right)$,

---

* Статья получена 21 октября 2011 г.

$\left(x > 0,\ j = \overline{1, n_i};\ i = \overline{1, N}\right)$; значения параметров a_i неизвестны и, возможно, различны для разных i.

Рассмотрим выборку с номером i и положим $\zeta_{ir} = \sum_{j=1}^{r} \xi_{ij} \Big/ \sum_{j=1}^{r+1} \xi_{ij}$ $\left(r = \overline{1, n_i - 1}\right)$. В [1] показано, что при справедливости гипотезы H_0 статистики ζ_{ir} взаимно независимы и подчиняются B-распределениям с параметрами r и 1. Отсюда, в свою очередь, следует, что ζ_{ir}^{r} взаимно независимы и одинаково равномерно распределены на отрезке [0,1]. Также в [1] показано, что при $n \geq 3$ независимость статистик ζ_{ir} и их подчинение B-распределениям с указанными параметрами является свойством, характеризующим Г-распределение, частным случаем которого является показательное распределение. К статистикам ζ_{ir}^{r} $\left(\text{в количестве } \sum_{i=1}^{N} n_i - N\right)$ для проверки простой гипотезы о принадлежности равномерному закону $U(0,1)$ можно применять различные критерии согласия. Далее мы использовали непараметрические критерии Колмогорова, ω^2 Крамера–Мизеса–Смирнова и Ω^2 Андерсона–Дарлинга [2, 3].

При справедливости проверяемой гипотезы распределения статистик критериев согласия при общем количестве наблюдений $N \leq n \geq 25$ практически не отличаются (рис. 1) от теоретических (предельных). Здесь и далее объем моделируемых выборок статистик критериев согласия равен 100 000 наблюдений.

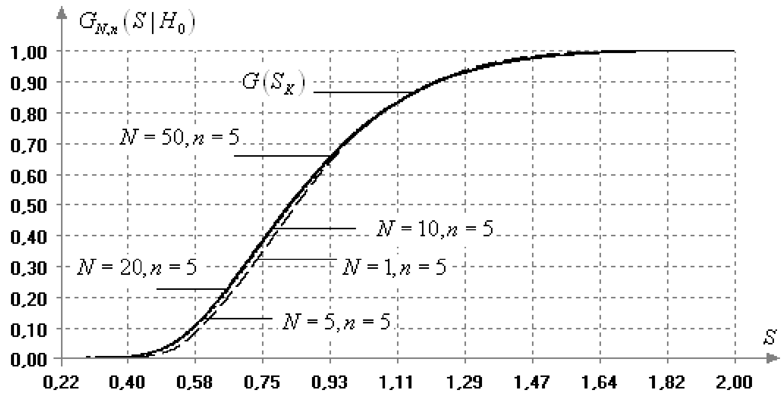

Рис. 1. Зависимость распределения статистики Колмогорова от объема выборок и их числа при проверке простой гипотезы $H : \zeta_{ir}^{r} \sim U(0,1)$ $\left(r = \overline{1, n-1}; i = \overline{1, N}\right)$

2. МОЩНОСТЬ КРИТЕРИЯ БОЛЬШЕВА ОТНОСИТЕЛЬНО РАЗЛИЧНЫХ АЛЬТЕРНАТИВ

В качестве проверяемой рассмотрим гипотезу H_0: $f_{\xi_{ij}}(x) = f_{\exp}(x)$, $\left(i = \overline{1,N},\ j = \overline{1,n}\right)$, где $f_{\exp}(x) = \exp\{-x/\theta_1\}/\theta_1$, $x \geq 0$ – плотность показательного распределения, $\theta_1 = 1$.

Показательный закон является частным случаем распределения Вейбулла $f(x) = \theta_2 x^{\theta_2 - 1} / \theta_1^{\theta_2} \exp\left\{-\left(x/\theta_1\right)^{\theta_2}\right\}$, $x \geq 0$, при параметре формы $\theta_2 = 1$. Поэтому в качестве близких конкурирующих гипотез $H_1, ..., H_6$ возьмем распределение Вейбулла при $\theta_2 \in \{0.7,\ 0.8,\ 0.95,\ 1.05,\ 1.2,\ 1.3\}$. На рис. 2 представлены плотности распределения данных законов.

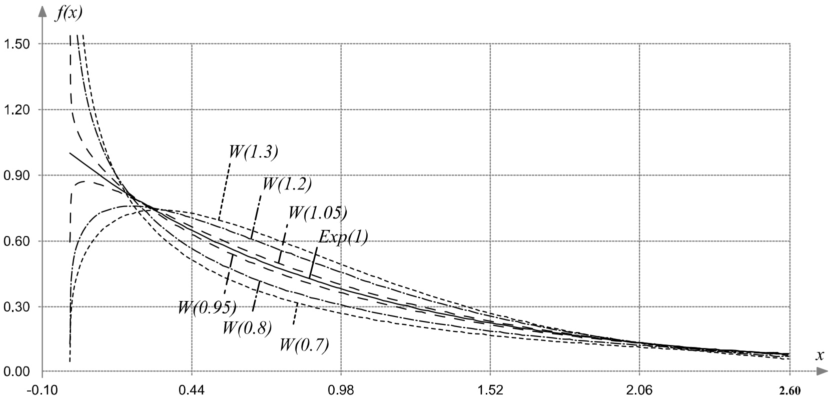

Рис. 2. Зависимость плотности распределения Вейбулла от параметра формы

В качестве менее близких конкурирующих гипотез рассмотрим ряд распределений с легкими «хвостами», плотность распределения которых сосредоточена в окрестности нуля. Чтобы получить распределение подобного вида максимально близким к рассматриваемому показательному, параметры данных распределений мы оценили методом максимального правдоподобия по выборке показательного распределения в 10’000 наблюдений. Вычисления проводились в программной системе ISW 4.2 [4].

Таблица 1

Мощность критерия Большева по отношению к гипотезам H_1, H_2, H_3 при использовании критерия Колмогорова

$N(n-1)$	N	n	α (H_1)			α (H_2)			α (H_3)		
			0.15	0.05	0.01	0.15	0.05	0.01	0.15	0.05	0.01
4	1	5	.220	.102	.026	.190	.079	.019	.160	.056	.012
9	1	10	.275	.125	.041	.212	.086	.023	.159	.056	.012
19	1	20	.384	.206	.081	.258	.115	.034	.162	.057	.012
20	5	5	.336	.156	.050	.234	.095	.025	.160	.056	.011
40	10	5	.475	.249	.090	.289	.127	.035	.164	.057	.012
45	5	10	.552	.324	.146	.330	.153	.050	.165	.057	.013
49	1	50	.662	.449	.247	.392	.204	.081	.172	.062	.014
80	20	5	.710	.446	.195	.403	.192	.059	.169	.059	.012
90	10	10	.796	.582	.325	.472	.254	.095	.173	.062	.014
95	5	20	.847	.666	.425	.523	.301	.128	.178	.065	.014
180	20	10	.972	.890	.694	.703	.469	.228	.189	.069	.016
190	10	20	.985	.935	.805	.758	.543	.303	.198	.073	.018
200	50	5	.976	.879	.627	.696	.425	.178	.186	.064	.014
245	5	50	.998	.988	.942	.882	.719	.469	.220	.086	.021
380	20	20	.999	.999	.991	.957	.859	.659	.239	.094	.023
450	50	10	.999	.999	.995	.972	.879	.669	.238	.091	.022
490	10	50	.999	.999	.999	.991	.957	.852	.279	.117	.032
950	50	20	.999	.999	.999	.999	.999	.991	.364	.164	.051
980	20	50	.999	.999	.999	.999	.999	.996	.393	.190	.062
2450	50	50	.999	.999	.999	.999	.999	.999	.674	.428	.189

Гипотеза H_7 соответствует распределению Максвелла с плотностью распределения $f(x)=\sqrt{2/\pi}\left(x-\theta_0\right)^2\exp\left\{-\left(x-\theta_0\right)^2\Big/2\theta_1^2\right\}\Big/\theta_1^3$, x i θ_0, с параметром сдвига $\theta_0=-0.92$ и параметром масштаба $\theta_1=1.23$; H_8 – распределению Рэлея с плотностью $f(x)=\left(x-\theta_0\right)\exp\left\{-\left(x-\theta_0\right)^2\Big/2\theta_1^2\right\}\Big/\theta_1^2$, x i θ_0, с параметром сдвига $\theta_0=-0.46$ и параметром масштаба $\theta_1=1.23$; H_9 – распределению Парето с плотностью распределения $f(x)=\theta_0\theta_1^{\theta_2}\Big/x^{\theta_0+1}$, x i θ_1 с параметром формы $\theta_0=1.005$ и параметром положения $\theta_1=0.005$ (рис. 3).

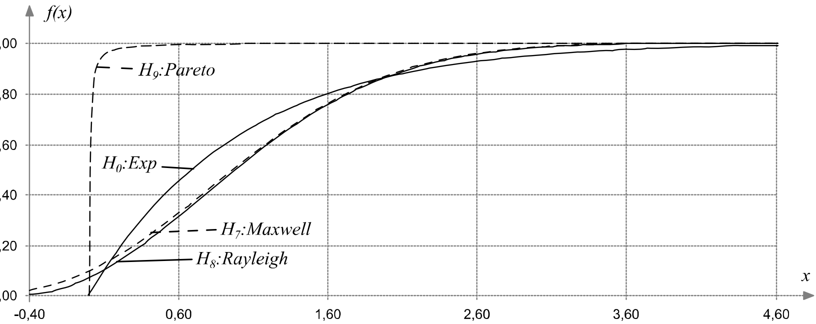

Рис. 3. Плотности распределений, соответствующие гипотезам H_7 – H_9

Распределения статистик и мощность критериев относительно рассматриваемых конкурирующих гипотез исследовались для объемов выборок n = 5, 10, 20, 50 и количества выборок N = 1, 5, 10, 20, 50. Оценки мощности критерия Большева относительно гипотез H_1, H_2, H_3 при использовании для проверки равномерности критерия Колмогорова приведены в табл. 1. Нетрудно заметить, что мощность в первую очередь зависит от эффективного объема выборки $N(n-1)$ и в меньшей степени – от объема малых выборок.

3. СРАВНЕНИЕ С МОЩНОСТЬЮ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ СОГЛАСИЯ

Следует отметить, что непосредственное применение непараметрических критериев для проверки гипотез о распределении случайных величин по нескольким выборкам невозможно. Однако в рассматриваемом частном случае, когда наблюдения в малых выборках имеют показательное распределение с одинаковыми параметрами, можно сравнивать мощность критерия при подходе Большева с мощностью критериев согласия при объемах выборок, равных эффективным объемам (табл. 1).

Кроме того, исходная постановка задачи предполагает проверку сложной гипотезы (принадлежность малых выборок *семейству* показательных распределений), а решение об отклонении или неотклонении данной гипотезы принимается на основе проверки *простой* гипотезы $H: \zeta_{ir}^{r} \sim U(0,1)$.

В табл. 2 приведены значения мощности относительно конкурирующих гипотез H_1 и H_3 для критерия согласия Колмогорова при проверке простых и сложных гипотез о принадлежности выборки объемом n показательному закону.

Таблица 2

Мощность критерия Колмогорова по отношению к гипотезам H_1 и H_3

n	H_1 (H_0 простая)			H_1 (H_0 сложная)			H_3 (H_0 простая)			H_3 (H_0 сложная)		
	0.15	0.05	0.01	0.15	0.05	0.01	0.15	0.05	0.01	0.15	0.05	0.01
4	.221	.100	.023	.207	.077	.033	.158	.055	.011	.147	.049	.009
9	.261	.111	.031	.323	.176	.071	.157	.054	.011	.150	.052	.009
19	.331	.147	.046	.531	.357	.181	.162	.056	.012	.157	.057	.012
20	.337	.147	.043	.550	.378	.192	.160	.056	.011	.159	.055	.012
40	.478	.241	.083	.801	.654	.446	.162	.056	.012	.171	.062	.014
45	.507	.256	.093	.837	.699	.493	.162	.055	.011	.176	.061	.016
49	.540	.268	.090	.868	.732	.527	.167	.056	.011	.176	.058	.015
80	.704	.433	.178	.963	.907	.787	.166	.057	.012	.193	.073	.021
90	.754	.495	.213	.977	.939	.831	.171	.059	.012	.199	.075	.017
95	.777	.493	.240	.982	.946	.870	.173	.059	.012	.199	.082	.021
180	.963	.830	.493	.999	.997	.988	.187	.063	.012	.248	.109	.025
190	.972	.855	.566	.999	.998	.993	.182	.066	.014	.254	.111	.029
200	.975	.868	.601	.999	.999	.995	.182	.067	.014	.264	.114	.031
245	.991	.940	.717	.999	.999	.998	.189	.070	.014	.275	.127	.037
380	.999	.997	.954	.999	.999	.999	.213	.085	.015	.341	.168	.056
450	.999	.999	.988	.999	.999	.999	.220	.086	.019	.371	.190	.064
490	.999	.999	.992	.999	.999	.999	.230	.091	.018	.388	.217	.065
950	.999	.999	.999	.999	.999	.999	.294	.114	.029	.564	.349	.159
980	.999	.999	.999	.999	.999	.999	.293	.117	.032	.577	.363	.177
2450	.999	.999	.999	.999	.999	.999	.535	.261	.082	.881	.745	.527

В [3] отмечается, что мощность критериев согласия в случае простой гипотезы при применении оценок максимального правдоподобия оказывается существенно выше мощности тех же критериев при проверке простой гипотезы. Проведенные исследования показали, что мощность критерия Большева относительно близких конкурирующих гипотез H_1, …, H_6 сопоставима с мощностью используемых в нем непараметрических критериев согласия, которой обладают последние в ситуации проверки простой гипотезы (см. табл. 2).

В то же время исследования показали, что по отношению к далеким конкурирующим гипотезам критерий Большева заметно выигрывает в мощности у критерия Ω^2 Андерсона–Дарлинга.

Мощность критерия, использующего подход Большева, определяется мощностью критерия согласия, применяемого для проверки равномерности. Поэтому наиболее предпочтительным оказывается применение критерия согласия Ω^2 Андерсона–Дарлинга, который при проверке простых гипотез, как правило, обладает большей мощностью по сравнению с критериями Колмогорова и ω^2 Крамера–Мизеса–Смирнова [2].

ЗАКЛЮЧЕНИЕ

Подход, предложенный Большевым, является достаточно хорошим средством проверки показательности распределения по набору выборок малого объема. Проведенный сравнительный анализ мощности критерия с использованием данного подхода с мощностью в сопоставимых условиях непараметрических критериев согласия, применяемых для проверки показательности, позволяет заключить, что преобразования, проводимые при вычислении статистик критерия Большева, не приводят к потерям информации, содержащейся в выборках, либо приводят к несущественным потерям.

Мощность критерия, использующего подход Большева для проверки показательности, напрямую зависит от мощности применяемого критерия равномерности. Из рассмотренных критериев согласия можно рекомендовать использование непараметрического критерия Ω^2 Андерсона–Дарлинга, так как это обеспечивает максимальную мощность относительно как близких, так и далеких конкурирующих гипотез. В то же время целесообразно провести исследование и сравнить по мощности множество критериев, предназначенных непосредственно для проверки гипотезы о принадлежности выборки равномерному закону. Выбор наиболее мощного критерия придаст наилучшие свойства и критерию показательности.

[1] *Большев Л.* К вопросу о проверке «показательности» // Теория вероятностей и ее применения. – 1966. – Т. 11. – № 3. – С. 542–544.

[2] *Лемешко Б., Лемешко С., Постовалов С.* Сравнительный анализ мощности критериев согласия при близких конкурирующих гипотезах. I. Проверка простых гипотез // Сибирский журнал индустриальной математики. – 2008. – Т. 11. – № 2(34). – С. 96–111.

[3] *Лемешко Б., Лемешко С., Постовалов С.* Сравнительный анализ мощности критериев согласия при близких альтернативах. II. Проверка сложных

гипотез // Сибирский журнал индустриальной математики. – 2008. – Т. 11. – № 4(36). – С. 78–93.

[4] *Лемешко Б., Постовалов С.* Компьютерные технологии анализа данных и исследования статистических закономерностей: учеб. пособие. – Новосибирск: Изд-во НГТУ, 2004. – 120 с.

Лемешко Борис Юрьевич – декан факультета прикладной математики и информатики Новосибирского государственного технического университета, доктор технических наук, профессор кафедры прикладной математики. Основное направление научных исследований – компьютерные методы исследования вероятностных закономерностей, методы прикладной математической статистики. Имеет более 300 публикаций, в том числе 9 монографий. Тел. (383) 346-37-54. E-mail: Lemeshko@fpm.ami.nstu.ru.

Рогожников Андрей Павлович – аспирант кафедры прикладной математики Новосибирского государственного технического университета. Основное направление научных исследований – компьютерные методы исследования свойств статистических критериев. Тел. +7-923-234-03-42. E-mail: rogozhnikov.andrey@gmail.com.

B.Yu. Lemeshko, A.P. Rogozhnikov

Study of power of Bol’shev’s test for exponentiality

Bol’shev’s test for exponential distribution of observations in several small samples is considered. Power of the test is studied by means of statistical simulation for different goodness-of-fit tests used as auxiliary. Comparison with power of the goodness-of-fit tests is conducted.

Key words: exponential distribution, power, Bol’shev’s test.