

Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. I

Б.Ю. Лемешко, С.Б. Лемешко

Новосибирский государственный технический университет

E-mail: Lemeshko@fpm.ami.nstu.ru

При проверке сложных гипотез, когда скалярный или векторный параметр закона распределения вероятностей вычисляется по той же выборке, непараметрические критерии согласия Колмогорова, Крамера-Мизеса-Смирнова и Андерсона-Дарлингга теряют свободу от распределения. При проверке сложных гипотез условные распределения статистик критериев зависят от ряда факторов.

В работе приводятся уточненные результаты (таблицы процентных точек и модели распределений статистик) для непараметрических критериев согласия при проверке сложных гипотез относительно ряда законов распределения в случае использования оценок максимального правдоподобия.

Ключевые слова: критерии согласия, проверка сложных гипотез, критерий Колмогорова, критерий Крамера-Мизеса-Смирнова, критерий Андерсона-Дарлингга.

1. Введение

Практика применения непараметрических критериев согласия в задачах статистического анализа изобилует примерами некорректного использования классических результатов, имеющих силу при проверке простых гипотез, в ситуациях, соответствующих проверке сложных гипотез.

При проверке сложных гипотез вида $H_0: F(x) \in \{F(x, \theta), \theta \in \Theta\}$, где оценка $\hat{\theta}$ скалярного или векторного параметра распределения $F(x, \theta)$ вычисляется по той же самой выборке, непараметрические критерии согласия

Колмогорова, ω^2 Крамера-Мизеса-Смирнова, Ω^2 Андерсона-Дарлинга теряют свойство свободы от распределения.

В критерии Колмогорова в качестве расстояния между эмпирическим и теоретическим законами распределения используется величина

$$D_n = \sup_{|x| < \infty} |F_n(x) - F(x, \theta)|,$$

где $F_n(x)$ – эмпирическая функция распределения, n – объем выборки. При $n \rightarrow \infty$ функция распределения статистики $\sqrt{n}D_n$ при справедливости проверяемой гипотезы сходится равномерно к функции распределения Колмогорова [1]

$$K(S) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 s^2}$$

При проверке гипотез с применением критерия Колмогорова рекомендуется использовать статистику с поправкой Большева [2,3] в форме [4]

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}}, \quad (1)$$

где $D_n = \max(D_n^+, D_n^-)$,

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i, \theta) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ F(x_i, \theta) - \frac{i-1}{n} \right\},$$

n – объем выборки, x_1, x_2, \dots, x_n – упорядоченные по возрастанию элементы выборки. При справедливости простой проверяемой гипотезы статистика (1) также подчиняется закону распределения Колмогорова [4], причем сходится к нему существенно быстрее, чем статистика $\sqrt{n}D_n$.

К большому сожалению, поправка Большева не привлекла внимания зарубежных специалистов. В работах, посвященных критерию Колмогорова до сих пор, как правило, используют статистику $\sqrt{n}D_n$. Вследствие этого при

ограниченных значениях n вынуждены учитывать существенную зависимость распределения статистики от величины n .

В критерии ω^2 Крамера-Мизеса-Смирнова используется статистика вида [4]

$$S_{\omega} = n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(x_i, \theta) - \frac{2i-1}{2n} \right\}^2, \quad (2)$$

а в критерии типа Ω^2 Андерсона-Дарлинга [5, 6] – статистика в форме

$$S_{\Omega} = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n} \right) \ln(1 - F(x_i, \theta)) \right\}. \quad (3)$$

При проверке простых гипотез статистика (2) подчиняется распределению вида [4]

$$a1(S) = \frac{1}{\sqrt{2s}} \sum_{j=0}^{\infty} \frac{\Gamma(j+1/2)\sqrt{4j+1}}{\Gamma(1/2)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2}{16S}\right\} \times \left\{ I_{-\frac{1}{4}}\left[\frac{(4j+1)^2}{16S}\right] - I_{\frac{1}{4}}\left[\frac{(4j+1)^2}{16S}\right] \right\},$$

где $I_{-\frac{1}{4}}(\cdot)$, $I_{\frac{1}{4}}(\cdot)$ - модифицированные функции Бесселя,

$$I_{\nu}(z) = \sum_{k=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \quad |z| < \infty, \quad |\arg z| < \pi,$$

а статистика (3) подчиняется распределению (Bolshev and Smirnov, вида [4])

$$a2(S) = \frac{\sqrt{2\pi}}{S} \sum_{j=0}^{\infty} (-1)^j \frac{\Gamma\left(j+\frac{1}{2}\right)(4j+1)}{\Gamma\left(\frac{1}{2}\right)\Gamma(j+1)} \exp\left\{-\frac{(4j+1)^2\pi^2}{8S}\right\} \times \int_0^{\infty} \exp\left\{\frac{S}{8(y^2+1)} - \frac{(4j+1)^2\pi^2 y^2}{8S}\right\} dy.$$

2. Распределения статистик при проверке сложных гипотез

При проверке сложных гипотез условные распределения статистик $G(S|H_0)$ зависят от ряда факторов: от вида наблюдаемого закона $F(x, \theta)$,

соответствующего справедливой проверяемой гипотезе H_0 ; от типа оцениваемого параметра и числа оцениваемых параметров; в некоторых случаях от конкретного значения параметра (например, в случае семейств гамма- и бета-распределений); от метода оценивания параметров. Различия в предельных распределениях той же самой статистики при проверке простых и сложных гипотез настолько существенны, что пренебрегать этим ни в коем случае нельзя. Например, на рисунке 1 показаны распределения статистики Андерсона-Дарлинга (3) при проверке сложных гипотез относительно различных законов распределения при использовании оценок максимального правдоподобия (ОМП) двух параметров.

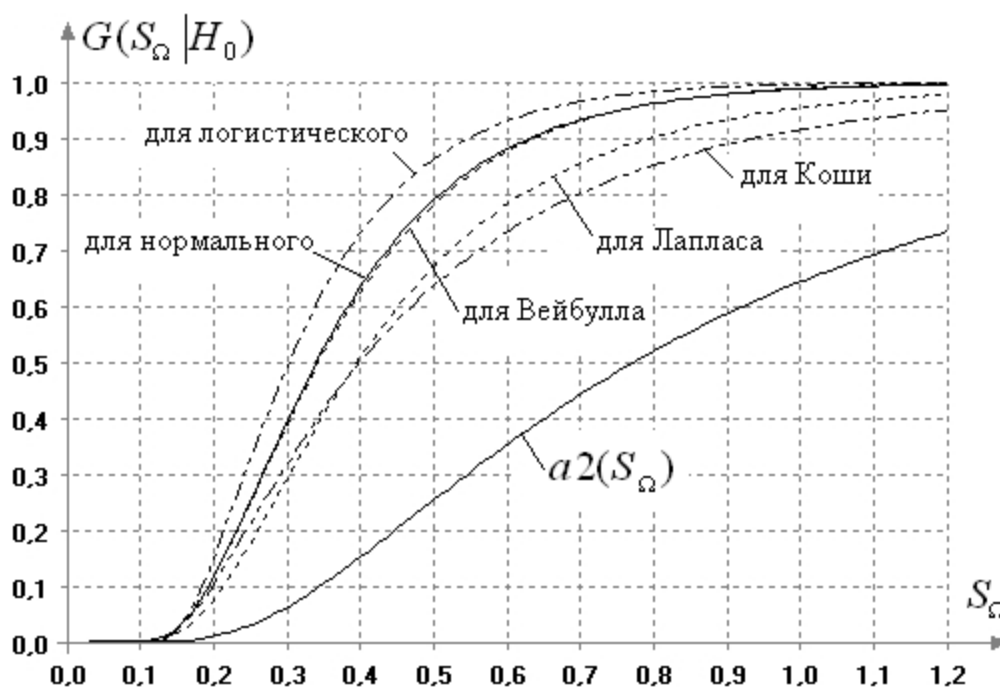


Рис. 1. Распределения статистики (3) при проверке сложных гипотез с вычислением ОМП двух параметров закона

Рисунок 2 иллюстрирует зависимость распределения статистики критерия Колмогорова (1) от типа и числа оцениваемых параметров на примере закона распределения Su -Джонсона.

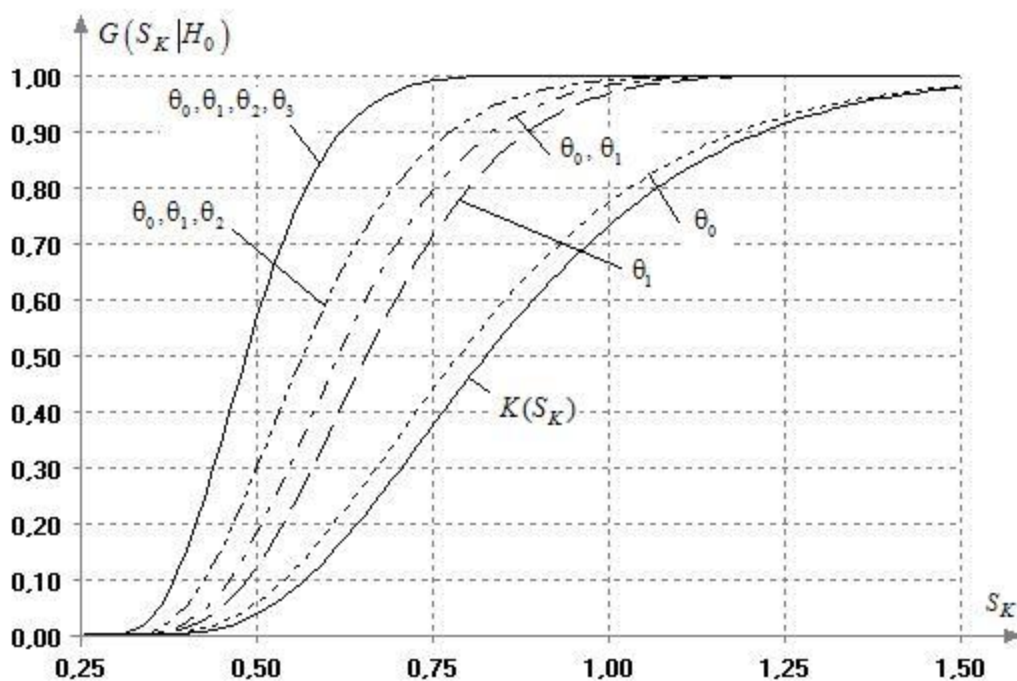


Рис. 2. Распределения статистики Колмогорова (1) при проверке сложных гипотез с вычислением ОМП параметров распределения *Si*-Джонсона

Точкой отсчета, с которой были начаты исследования предельных распределений статистик непараметрических критериев согласия при сложных гипотезах, послужила работа [7]. В дальнейшем для решения данной проблемы использовались различные подходы: предельные распределения статистик исследовались аналитическими методами [8, 9], [10, 11], [12-14], [15, 16], [15], [18-20]; процентные точки распределений строились методами статистического моделирования [21], [22, 23], [24]; строились формулы, дающие достаточно хорошие приближения при малых значениях соответствующих вероятностей [25, 26], [27].

В наших исследованиях [28-35] распределения статистик непараметрических критериев согласия исследовались методами статистического моделирования. Далее, опираясь на полученные эмпирические распределения статистик, строились приближенные аналитические модели законов распределения

статистик. Часть полученных результатов были использованы нами при подготовке рекомендаций по стандартизации Р 50.1.037-2002 [36].

3. Уточнение моделей распределений статистик непараметрических критериев согласия

В настоящей работе представлены уточненные результаты (таблицы процентных точек и модели распределений статистик) для непараметрических критериев согласия при проверке сложных гипотез в случае использования ОМП. Данные исследования опираются на развиваемое программное обеспечение, позволяющее исследовать вероятностные закономерности методами статистического моделирования и строить приближенные аналитические модели для этих закономерностей.

Таблица 1 содержит список законов распределения, относительно которых можно проверять сложные гипотезы, используя построенные приближения для предельных распределений непараметрических критериев согласия.

Таблица 1. Законы распределения случайных величин.

Наименование закона	Функция плотности $f(x, \theta)$	Наименование закона	Функция плотности $f(x, \theta)$
Экспоненциальный	$\frac{1}{\theta_0} e^{-x/\theta_0}$	Лапласа	$\frac{1}{2\theta_0} e^{- x-\theta_1 /\theta_0}$
Полунормальный	$\frac{2}{\theta_0 \sqrt{2\pi}} e^{-x^2/2\theta_0^2}$	Нормальный	$\frac{1}{\theta_0 \sqrt{2\pi}} e^{-\frac{(x-\theta_1)^2}{2\theta_0^2}}$
Рэлея	$\frac{x}{\theta_0^2} e^{-x^2/2\theta_0^2}$	Логарифмически нормальный	$\frac{1}{x\theta_0 \sqrt{2\pi}} e^{-(\ln x - \theta_1)^2 / 2\theta_0^2}$
Максвелла	$\frac{2x^2}{\theta_0^3 \sqrt{2\pi}} e^{-x^2/2\theta_0^2}$	Коши	$\frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]}$
Наименование закона	Функция плотности $f(x, \theta)$		
Логистический	$\frac{\pi}{\theta_0 \sqrt{3}} \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x - \theta_1)}{\theta_0 \sqrt{3}}\right\}\right]^2$		

Экстремального значения (maximum)	$\frac{1}{\theta_0} \exp\left\{-\frac{x-\theta_1}{\theta_0} - \exp\left(-\frac{x-\theta_1}{\theta_0}\right)\right\}$
Экстремального значения (minimum)	$\frac{1}{\theta_0} \exp\left\{\frac{x-\theta_1}{\theta_0} - \exp\left(\frac{x-\theta_1}{\theta_0}\right)\right\}$
Вейбулла	$\frac{\theta_0 x^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x}{\theta_1}\right)^{\theta_0}\right\}$
<i>Sb</i> -Джонсона $Sb(\theta_0, \theta_1, \theta_2, \theta_3)$	$\frac{\theta_1 \theta_2}{(x-\theta_3)(\theta_2+\theta_3-x)} \exp\left\{-\frac{1}{2}\left[\theta_0 - \theta_1 \ln \frac{x-\theta_3}{\theta_2+\theta_3-x}\right]^2\right\}$
<i>Sl</i> -Джонсона $Sl(\theta_0, \theta_1, \theta_2, \theta_3)$	$\frac{\theta_1}{x-\theta_3} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln \frac{x-\theta_3}{\theta_2}\right]^2\right\}$
<i>Su</i> -Джонсона $Su(\theta_0, \theta_1, \theta_2, \theta_3)$	$\frac{\theta_1}{\sqrt{2\pi} \sqrt{x-\theta_3}^2 + \theta_2^2} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln \left\{\frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2 + 1}\right\}\right]^2\right\}$
Гамма-распределение $\gamma(\theta_0, \theta_1, \theta_2)$	$\frac{1}{\theta_1^{\theta_0} \Gamma(\theta_0)} (x-\theta_2)^{\theta_0-1} e^{-x-\theta_2/\theta_1}$
Двусторонний экспоненциальный	$\frac{\theta_0}{2\theta_1 \Gamma(1/\theta_0)} \exp\left\{-\left(\frac{ x-\theta_2 }{\theta_1}\right)^{\theta_0}\right\}$

Таблицы процентных точек и модели распределений статистик критериев строились по смоделированным выборкам статистик объемом $N=10^6$. При таких N величина разности между истинным законом $G(S|H_0)$ распределения статистики и смоделированным эмпирическим $G_N(S|H_0)$ по модулю не превышает величины 10^{-3} . При этом значения статистик критериев вычислялись по выборкам псевдослучайных величин, генерируемых в соответствии с наблюдаемым законом $F(x, \theta)$, объемом $n=10^3$. В такой ситуации распределение $G(S_n|H_0)$ практически совпадает с предельным $G(S|H_0)$.

Распределения $G(S|H_0)$ статистики Колмогорова наилучшим образом аппроксимируются семейством гамма распределений $\gamma(\theta_0, \theta_1, \theta_2)$ (см. таблицу 1) или семейством бета-распределений III-го рода с функцией плотности

$$B_3(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4) = \frac{\theta_2^{\theta_0}}{\theta_3 B(\theta_0, \theta_1)} \frac{\left(\frac{x - \theta_4}{\theta_3}\right)^{\theta_0 - 1} \left(1 - \frac{x - \theta_4}{\theta_3}\right)^{\theta_1 - 1}}{\left[1 + (\theta_2 - 1) \frac{x - \theta_4}{\theta_3}\right]^{\theta_0 + \theta_1}}.$$

А распределения статистик Крамера-Мизеса-Смирнова и Андерсона-Дарлингга хорошо аппроксимируются семейством распределений Sb -Джонсона $Sb(\theta_0, \theta_1, \theta_2, \theta_3)$ (см. таблицу 1) или семейством бета-распределений III-го рода $B_3 \theta_0, \theta_1, \theta_2, \theta_3, \theta_4$.

Верхние процентные точки и построенные модели для предельных распределений статистики критерия Колмогорова в случае использования ОМП представлены в таблице 2 [для законов: экспоненциального, полунормального, Рэлея, Максвелла, Лапласа, нормального, логнормального, Коши, логистического, экстремальных значений (maximum и minimum), Вейбулла].

Для тех же законов распределения верхние процентные точки и построенные модели для предельных распределений статистики критерия Крамера-Мизеса-Смирнова представлены в таблице 3, для критерия Андерсона-Дарлингга – в таблице 4.

В таблице 5 приведены верхние процентные точки и модели предельных распределений статистик непараметрических критериев согласия в случае проверки сложных гипотез относительно законов распределения Sb -Джонсона (при использовании ОМП), в таблице 6 – относительно законов Sl -Джонсона, в таблице 7 – относительно законов Su -Джонсона.

Во всех выше перечисленных случаях распределения $G(S|H_0)$ статистик

критериев согласия не зависят от конкретных значений неизвестных параметров законов $F(x, \theta)$.

Таблица 2. Верхние процентные точки и модели предельных распределений статистики критерия Колмогорова в случае использования ОМП

Наименование закона	Оцениваемые параметры	Процентные точки			Модель
		0.1	0.05	0.01	
Экспоненциально-Рэлея	Масштаба	0.995	1.094	1.292	$\gamma(5.1092, 0.0861, 0.2950)$
Полунормальное	Масштаба	1.051	1.160	1.381	$\gamma(4.5462, 0.1001, 0.3100)$
Максвелла	Масштаба	0.969	1.062	1.251	$\gamma(5.4566, 0.0794, 0.2870)$
Лапласа	Масштаба	1.177	1.313	1.587	$B_3(4.4680, 4.8450, 3.9105, 2.3784, 0.324)$
	Сдвига	0.957	1.045	1.223	$B_3(5.3541, 7.2519, 2.5630, 1.7652, 0.302)$
	2 параметра	0.863	0.940	1.095	$\gamma(6.2949, 0.0624, 0.2613)$
Нормальное & логарифмически нормальное	Масштаба	1.190	1.327	1.600	$B_3(4.8849, 5.2341, 3.6279, 2.3872, 0.303)$
	Сдвига	0.888	0.963	1.114	$B_3(5.2604, 7.4327, 2.1872, 1.4774, 0.30)$
	2 параметра	0.835	0.909	1.057	$\gamma(6.4721, 0.0580, 0.2620)$
Коши	Масштаба	1.137	1.275	1.550	$\gamma(3.0987, 0.1463, 0.3350)$
	Сдвига	0.975	1.070	1.260	$\gamma(5.9860, 0.0780, 0.2528)$
	2 параметра	0.815	0.893	1.048	$\gamma(5.3642, 0.0654, 0.2600)$
Логистическое	Масштаба	1.180	1.316	1.589	$\gamma(3.4954, 0.1411, 0.3325)$
	Сдвига	0.837	0.907	1.046	$\gamma(7.6325, 0.0531, 0.2368)$
	2 параметра	0.746	0.805	0.923	$\gamma(7.5402, 0.0451, 0.2422)$
Экстремальных значений & Вейбулла	Масштаба ¹⁾	1.182	1.316	1.583	$\gamma(3.6805, 0.1355, 0.3350)$ ¹⁾
	Сдвига ²⁾	0.995	1.093	1.292	$\gamma(5.2194, 0.0848, 0.2920)$ ²⁾
	2 параметра	0.824	0.895	1.037	$\gamma(6.6012, 0.0563, .2598)$

Замечание. ¹⁾ – при оценивании параметра формы распределения Вейбулла, ²⁾ - при оценивании параметра масштаба распределения Вейбулла.

Таблица 3. Верхние процентные точки и модели предельных распределений статистики критерия Крамера-Мизеса-Смирнова в случае использования ОМП

Наименование закона	Оцениваемые параметры	Процентные точки			Модель
		0.1	0.05	0.01	
Экспоненциально-Рэлея	Масштаба	0.174	0.221	0.337	$Sb(3.3738, 1.2145, 1.0792, 0.011)$
Полунормальное	Масштаба	0.205	0.266	0.415	$Sb(3.527, 1.1515, 1.5527, 0.012)$
Максвелла	Масштаба	0.162	0.204	0.306	$Sb(3.353, 1.220, 0.9786, 0.0118)$
Лапласа	Масштаба	0.323	0.439	0.719	$B_3(3.9800, 1.4667, 38.0035, 1.13, 0.0111)$
	Сдвига	0.152	0.187	0.268	$B_3(3.3130, 3.8338, 10.097, 0.7517, 0.011)$
	2 параметра	0.115	0.144	0.213	$B_3(4.489, 3.7706, 17.577, 0.7065, 0.0085)$

Нормальное & логарифмически нормальное	Масштаба	0.327	0.442	0.725	$Sb(3.153, 0.9448, 2.5477, 0.016)$
	Сдвига	0.134	0.165	0.238	$B_3(4.433, 3.6365, 13.920, 0.6632, 0.0084)$
	2 параметра	0.103	0.126	0.178	$B_3(4.1153, 4.1748, 11.035, 0.5116, 0.009)$
Коши	Масштаба	0.315	0.430	0.711	$Sb(3.1895, 0.9134, 2.690, 0.013)$
	Сдвига	0.172	0.216	0.319	$Sb((2.359, 1.0732, 0.595, 0.0129)$
	2 параметра	0.129	0.170	0.271	$Sb(3.4364, 1.0678, 1.000, 0.011)$
Логистическое	Масштаба	0.323	0.438	0.719	$Sb(3.264, 0.9581, 2.7046, 0.014)$
	Сдвига	0.119	0.148	0.216	$Sb(4.0026, 1.2853, 1.00, 0.0122)$
	2 параметра	0.081	0.098	0.135	$Sb(3.2137, 1.3612, 0.36, 0.0105)$
Экстремальных значений & Вейбулла	Масштаба ¹⁾	0.320	0.431	0.704	$Sb(3.343, 0.9817, 2.753, 0.015) 1)$
	Сдвига ²⁾	0.174	0.221	0.336	$Sb(3.498, 1.2236, 1.1632, 0.01) 2)$
	2 параметра	0.102	0.124	0.174	$Sb(3.3854, 1.4453, 0.4986, 0.007)$

Замечание. ¹⁾ – при оценивании параметра формы распределения Вейбулла, ²⁾ - при оценивании параметра масштаба распределения Вейбулла.

Таблица 4. Верхние процентные точки и модели предельных распределений статистики критерия Андерсона-Дарлинга в случае использования ОМП

Наименование закона	Оцениваемые параметры	Процентные точки			Модель
		0.1	0.05	0.01	
Экспоненциальной & Рэлея	Масштаба	1.060	1.319	1.954	$Sb(3.8386, 1.3429, 7.500, 0.090)$
Полунормальное	Масштаба	1.188	1.499	2.267	$Sb(4.2019, 1.2918, 11.500, 0.100)$
Максвелла	Масштаба	1.009	1.247	1.832	$Sb(3.9591, 1.3296, 7.800, 0.1010)$
Лапласа	Масштаба	1.725	2.290	3.685	$B_3(4.0842, 1.7532, 28.1434, 6.00, 0.105)$
	Сдвига	1.071	1.302	1.837	$B_3(4.0842, 1.7532, 28.1434, 6.00, 0.105)$
	2 параметра	0.798	0.982	1.439	$B_3(5.3576, 3.8690, 17.2148, 4.2386, 0.073)$
Нормальное & логарифмически нормальное	Масштаба	1.743	2.309	3.704	$B_3(3.4638, 2.330, 35.7115, 12.603, 0.105)$
	Сдвига	0.892	1.087	1.552	$B_3(4.1081, 5.0598, 16.9721, 7.9065, 0.09)$
	2 параметра	0.630	0.750	1.032	$B_3(4.7262, 4.6575, 9.4958, 2.717, 0.0775)$
Коши	Масштаба	1.716	2.277	3.673	$Sb(3.7830, 1.0678, 18.0, 0.11)$
	Сдвига	1.215	1.512	2.211	$Sb(3.4814, 1.2375, 7.810, 0.1)$
	2 параметра	0.948	1.226	1.913	$Sb(3.290, 1.129, 5.837, 0.099)$
Логистическое	Масштаба	1.724	2.285	3.682	$Sb(3.516, 1.054, 14.748, 0.117)$
	Сдвига	0.856	1.043	1.495	$Sb(5.1316, 1.5681, 10.0, 0.065)$
	2 параметра	0.562	0.665	0.903	$Sb(3.409, 1.434, 2.448, 0.095)$
Экстремальных значений & Вейбулла	Масштаба ¹⁾	1.723	2.273	3.634	$Sb(3.512, 1.064, 14.496, 0.125) 1)$
	Сдвига ²⁾	1.059	1.318	1.952	$Sb(4.799, 1.402, 13.0, 0.085) 2)$
	2 параметра	0.634	0.755	1.040	$Sb(3.4830, 1.5138, 3.00, 0.07)$

Замечание. ¹⁾ – при оценивании параметра формы распределения Вейбулла, ²⁾ - при оценивании параметра масштаба распределения Вейбулла.

Таблица 5. Верхние процентные точки и модели предельных распределений статистик непараметрических критериев согласия в случае проверки гипотез относительно распределений *Sb*-Джонсона при использовании ОМП

Оцениваемые параметры	Процентные точки			Модель
	0.9	0.95	0.99	
для критерия Колмогорова				
θ_0	0.888	0.963	1.115	$V_3(6.3484, 7.4913, 2.3663, 1.4790, 0.27)$
θ_1	1.189	1.326	1.600	$V_3(6.8242, 4.7737, 5.2621, 2.3878, 0.27)$
θ_0, θ_1	0.836	0.909	1.058	$V_3(6.6559, 8.1766, 2.9405, 1.6143, 0.27)$
для критерия Крамера-Мизеса-Смирнова				
θ_0	0.134	0.165	0.238	$V_3(4.2304, 3.8058, 13.1934, 0.6908, 0.0086)$
θ_1	0.327	0.442	0.724	$V_3(2.9153, 2.0048, 33.4135, 2.07821, 0.0114)$
θ_0, θ_1	0.104	0.126	0.179	$V_3(4.3897, 4.0574, 12.1009, 0.5119, 0.0086)$
для критерия Андерсона-Дарлинга				
θ_0	0.893	1.086	1.553	$V_3(4.2657, 4.3788, 11.4946, 4.6551, 0.084)$
θ_1	1.741	2.309	3.702	$V_3(4.1703, 2.3363, 42.0833, 12.6019, 0.088)$
θ_0, θ_1	0.631	0.751	1.034	$V_3(4.0891, 5.9708, 9.6497, 4.0000, 0.082)$

Таблица 6. Верхние процентные точки и модели предельных распределений статистик непараметрических критериев согласия в случае проверки гипотез относительно распределений *Sl*-Джонсона при использовании ОМП

Оцениваемые параметры	Процентные точки			Модель
	0.9	0.95	0.99	
для критерия Колмогорова				
θ_0	0.888	0.963	1.115	$V_3(6.3484, 7.4913, 2.3663, 1.4790, 0.27)$
θ_1	1.189	1.326	1.600	$V_3(6.8242, 4.7737, 5.2621, 2.3878, 0.27)$
θ_2	0.888	0.963	1.115	$V_3(6.3484, 7.4913, 2.3663, 1.4790, 0.27)$
θ_0, θ_1	0.836	0.909	1.058	$V_3(6.6559, 8.1766, 2.9405, 1.6143, 0.27)$
θ_0, θ_2	0.888	0.963	1.115	$V_3(6.3484, 7.4913, 2.3663, 1.4790, 0.27)$
θ_1, θ_2	0.836	0.909	1.058	$V_3(6.6559, 8.1766, 2.9405, 1.6143, 0.27)$
$\theta_0, \theta_1, \theta_2$	0.836	0.909	1.058	$V_3(6.6559, 8.1766, 2.9405, 1.6143, 0.27)$
для критерия Крамера-Мизеса-Смирнова				
θ_0	0.134	0.165	0.238	$V_3(4.2304, 3.8058, 13.1934, 0.6908, 0.0086)$
θ_1	0.327	0.442	0.724	$V_3(2.9153, 2.0048, 33.4135, 2.07821, 0.0114)$

θ_2	0.134	0.165	0.238	$B_3(4.2304, 3.8058, 13.1934, 0.6908, 0.0086)$
θ_0, θ_1	0.104	0.126	0.179	$B_3(4.3897, 4.0574, 12.1009, 0.5119, 0.0086)$
θ_0, θ_2	0.134	0.165	0.238	$B_3(4.2304, 3.8058, 13.1934, 0.6908, 0.0086)$
θ_1, θ_2	0.104	0.126	0.179	$B_3(4.3897, 4.0574, 12.1009, 0.5119, 0.0086)$
$\theta_0, \theta_1, \theta_2$	0.104	0.126	0.179	$B_3(4.3897, 4.0574, 12.1009, 0.5119, 0.0086)$
для критерия Андерсона-Дарлинга				
θ_0	0.893	1.086	1.553	$B_3(4.2657, 4.3788, 11.4946, 4.6551, 0.084)$
θ_1	1.741	2.309	3.702	$B_3(4.1703, 2.3363, 42.0833, 12.6019, 0.088)$
θ_2	0.893	1.086	1.553	$B_3(4.2657, 4.3788, 11.4946, 4.6551, 0.084)$
θ_0, θ_1	0.631	0.751	1.034	$B_3(4.0891, 5.9708, 9.6497, 4.0000, 0.082)$
θ_0, θ_2	0.893	1.086	1.553	$B_3(4.2657, 4.3788, 11.4946, 4.6551, 0.084)$
θ_1, θ_2	0.631	0.751	1.034	$B_3(4.0891, 5.9708, 9.6497, 4.0000, 0.082)$
$\theta_0, \theta_1, \theta_2$	0.631	0.751	1.034	$B_3(4.0891, 5.9708, 9.6497, 4.0000, 0.082)$

Таблица 7. Верхние процентные точки и модели предельных распределений статистик непараметрических критериев согласия в случае проверки гипотез относительно распределений *Su*-Джонсона при использовании ОМП

Оцениваемые параметры	Процентные точки			Модель
	0.9	0.95	0.99	
для критерия Колмогорова				
θ_0	0.888	0.963	1.115	$B_3(6.3484, 7.4913, 2.3663, 1.4790, 0.27)$
θ_1	1.189	1.326	1.600	$B_3(6.8242, 4.7737, 5.2621, 2.3878, 0.27)$
θ_2	1.161	1.300	1.576	$B_3(5.3417, 4.6440, 4.7448, 2.3802, 0.29)$
θ_3	0.880	0.960	1.122	$B_3(6.6252, 7.4025, 3.0590, 1.6516, 0.27)$
θ_0, θ_1	0.836	0.909	1.058	$B_3(6.4792, 7.0243, 2.8437, 1.4260, 0.27)$
θ_0, θ_2	0.798	0.872	1.024	$B_3(6.4496, 6.7714, 3.3119, 1.4226, 0.27)$
θ_0, θ_3	0.802	0.875	1.023	$B_3(6.3069, 6.1065, 3.2916, 1.3317, 0.27)$
θ_1, θ_2	1.142	1.282	1.561	$B_3(5.9751, 4.4559, 5.6810, 2.4123, 0.27)$
θ_1, θ_3	0.792	0.858	0.994	$B_3(6.4839, 7.0152, 2.7376, 1.2838, 0.27)$
θ_2, θ_3	0.733	0.791	0.910	$B_3((6.2438, 6.9161, 2.5011, 1.0904, 0.27)$
$\theta_0, \theta_1, \theta_2$	0.776	0.851	1.007	$B_3(6.2414, 6.4027, 3.7458, 1.4361, 0.27)$
$\theta_0, \theta_1, \theta_3$	0.720	0.780	0.901	$B_3(6.4262, 6.9732, 2.7325, 1.1317, 0.26)$
$\theta_0, \theta_2, \theta_3$	0.658	0.706	0.806	$B_3(6.1239, 7.9516, 2.24033, 0.9839, 0.26)$
$\theta_1, \theta_2, \theta_3$	0.704	0.760	0.878	$B_3(7.1354, 8.0363, 2.7466, 1.1766, 0.25)$
$\theta_0, \theta_1, \theta_2, \theta_3$	0.622	0.666	0.755	$B_3(6.6889, 8.1712, 2.3857, 0.9291, 0.25)$

для критерия Крамера-Мизеса-Смирнова				
θ_0	0.134	0.165	0.238	$B_3(3.6736, 3.9355, 11.2146, 0.6908, 0.01)$
θ_1	0.327	0.442	0.724	$B_3(2.9153, 2.0048, 33.4135, 2.07821, 0.0114)$
θ_2	0.318	0.433	0.716	$B_3(2.2077, 1.7250, 28.4959, 1.75, 0.015)$
θ_3	0.125	0.154	0.225	$B_3(3.6990, 3.8775, 11.9942, 0.6601, 0.01)$
θ_0, θ_1	0.103	0.126	0.178	$B_3(4.3897, 4.0574, 12.1009, 0.5119, 0.0086)$
θ_0, θ_2	0.090	0.110	0.161	$B_3(5.2030, 3.9325, 15.6968, 0.4659, 0.0075)$
θ_0, θ_3	0.104	0.133	0.203	$B_3(5.9540, 3.1023, 30.6943, 0.6380, 0.0071)$
θ_1, θ_2	0.314	0.428	0.711	$B_3(2.4905, 1.6985, 45.9674, 2.3084, 0.012)$
θ_1, θ_3	0.094	0.113	0.158	$B_3(4.6011, 5.7370, 19.1580, 1.0, 0.0075)$
θ_2, θ_3	0.080	0.096	0.137	$B_3(4.7686, 4.6085, 11.1421, 0.3929, 0.0075)$
$\theta_0, \theta_1, \theta_2$	0.083	0.104	0.155	$B_3(5.2574, 3.6440, 19.9213, 0.4707, 0.0075)$
$\theta_0, \theta_1, \theta_3$	0.071	0.086	0.122	$B_3(5.7750, 4.7935, 18.1182, 0.4777, 0.0065)$
$\theta_0, \theta_2, \theta_3$	0.056	0.066	0.089	$B_3(7.3500, 5.4726, 13.7452, 0.2883, 0.0052)$
$\theta_1, \theta_2, \theta_3$	0.073	0.089	0.130	$B_3(5.6379, 4.0985, 18.5518, 0.42100, 0.007)$
$\theta_0, \theta_1, \theta_2, \theta_3$	0.048	0.056	0.075	$B_3(6.9739, 6.6406, 13.7433, 0.3151, 0.0052)$
для критерия Андерсона-Дарлинга				
θ_0	0.892	1.087	1.552	$B_3(4.2329, 4.5369, 10.8807, 4.6551, 0.082)$
θ_1	1.743	2.309	3.704	$B_3(4.1703, 2.3363, 42.0833, 12.6019, 0.088)$
θ_2	1.707	2.275	3.667	$B_3(2.6348, 1.9774, 21.3842, 7.75, 0.125)$
θ_3	0.952	1.161	1.648	$B_3(3.5597, 4.9656, 11.4180, 6.5202, 0.092)$
θ_0, θ_1	0.630	0.750	1.032	$B_3(4.0891, 5.9708, 9.6497, 4.0000, 0.082)$
θ_0, θ_2	0.576	0.689	0.961	$B_3(5.5368, 4.9114, 13.1278, 3.0625, 0.07)$
θ_0, θ_3	0.737	0.920	1.386	$B_3(5.6629, 3.4912, 25.1600, 4.5052, 0.07)$
θ_1, θ_2	1.666	2.232	3.627	$B_3(3.8896, 1.6253, 31.1820, 5.80, 0.09)$
θ_1, θ_3	0.694	0.842	1.200	$B_3(4.6199, 5.2874, 19.2708, 6.5610, 0.074)$
θ_2, θ_3	0.642	0.935	1.140	$B_3(4.4276, 4.30288, 14.6688, 3.7865, 0.08)$
$\theta_0, \theta_1, \theta_2$	0.518	0.627	0.898	$B_3(5.5158, 4.3512, 14.7750, 2.6199, 0.067)$
$\theta_0, \theta_1, \theta_3$	0.454	0.536	0.733	$B_3(5.3306, 5.8858, 10.7581, 2.5087, 0.065)$
$\theta_0, \theta_2, \theta_3$	0.395	0.459	0.606	$B_3(5.7098, 6.8325, 7.9837, 1.8803, 0.06)$
$\theta_1, \theta_2, \theta_3$	0.585	0.729	1.087	$B_3(5.1840, 3.2993, 19.3614, 2.7865, 0.073)$
$\theta_0, \theta_1, \theta_2, \theta_3$	0.329	0.378	0.488	$B_3(7.1015, 5.8708, 7.1323, 1.0517, 0.05)$

4. Пример

Проверить сложную гипотезу о принадлежности выборки нормальному закону. Упорядоченная выборка объемом 200 наблюдений имеет вид:

-1.270 -1.196 -1.043 -1.018 -1.011 -0.929 -0.916 -0.892 -0.886 -0.877 -0.827 -0.801
 -0.791 -0.782 -0.776 -0.766 -0.757 -0.736 -0.723 -0.714 -0.701 -0.700 -0.677 -0.673
 -0.667 -0.658 -0.627 -0.615 -0.615 -0.615 -0.604 -0.602 -0.572 -0.567 -0.565 -0.556
 -0.543 -0.542 -0.482 -0.475 -0.468 -0.462 -0.458 -0.450 -0.448 -0.437 -0.433 -0.416
 -0.410 -0.408 -0.403 -0.382 -0.370 -0.367 -0.365 -0.340 -0.337 -0.331 -0.326 -0.322
 -0.309 -0.304 -0.288 -0.282 -0.271 -0.252 -0.245 -0.234 -0.224 -0.224 -0.211 -0.205
 -0.178 -0.143 -0.131 -0.129 -0.127 -0.117 -0.115 -0.108 -0.095 -0.093 -0.078 -0.051
 -0.044 -0.035 -0.035 -0.034 -0.031 -0.012 -0.012 -0.010 -0.004 0.006 0.008 0.019
 0.023 0.037 0.045 0.055 0.056 0.083 0.086 0.092 0.101 0.104 0.114 0.121
 0.122 0.124 0.136 0.141 0.141 0.142 0.147 0.150 0.151 0.188 0.194 0.204
 0.206 0.211 0.211 0.228 0.231 0.232 0.242 0.252 0.252 0.257 0.270 0.278
 0.283 0.294 0.319 0.321 0.336 0.342 0.343 0.361 0.382 0.384 0.404 0.404
 0.404 0.405 0.408 0.435 0.468 0.474 0.480 0.482 0.510 0.514 0.533 0.541
 0.543 0.552 0.580 0.582 0.582 0.587 0.589 0.600 0.611 0.617 0.627 0.641
 0.643 0.643 0.645 0.651 0.668 0.677 0.678 0.679 0.693 0.714 0.737 0.737
 0.749 0.750 0.756 0.762 0.780 0.784 0.808 0.809 0.827 0.853 0.853 0.866
 0.877 0.883 0.902 0.960 0.994 1.048 1.075 1.114

Вычисленные по выборке ОМП параметров нормального закона: масштаба $\hat{\theta}_0 = 0.5383$, сдвига $\hat{\theta}_1 = 0.0294$.

Значение статистики (1) Колмогорова $S_K^* = 0.8152$ меньше критического значения $S_{K,0.1} = 0.835$ (см. таблицу 2), то есть проверяемая гипотеза не отклоняется при уровне значимости $\alpha = 0.1$. Вычисленный в соответствии с гамма-распределением $\gamma(6.4721, 0.0580, 0.2620)$ (таблица 2) достигаемый уровень значимости $P\{S_K > S_K^*\} \approx 0.119$.

Значение статистики (2) Крамера-Мизеса-Смирнова $S_\omega^* = 0.1451$ больше 0.126, но меньше 0.178 (см. таблицу 3), то есть проверяемая гипотеза не отклоняется при уровне значимости $\alpha = 0.01$. Вычисленный в соответствии с бета-распределением III-го типа $B_3(4.1153, 4.1748, 11.035, 0.5116, 0.009)$ (таблица 3) достигаемый уровень значимости $P\{S_\omega > S_\omega^*\} \approx 0.029$.

Значение статистики (3) Андерсона-Дарлинга $S_{\Omega}^* = 1.0157$ больше 0.750, но меньше 1.032 (см. таблицу 4). Вычисленный в соответствии с бета-распределением III-го типа $B_3(4.7262, 4.6575, 9.4958, 2.717, 0.0775)$ (таблица 4) достигаемый уровень значимости $P\{S_{\Omega} > S_{\Omega}^*\} \approx 0.0117$. Таким образом, проверяемая гипотеза не отклоняется при уровне значимости $\alpha = 0.01$.

Подчеркнем, что данная выборка сгенерирована по симметричному закону, существенно отличающемуся от нормального.

5. Заключение

Опираясь на компьютерные методы исследования статистических закономерностей, в основе которых лежит статистическое моделирование эмпирических распределений статистик и последующий анализ этих распределений, построены более точные модели распределений статистик непараметрических критериев согласия Колмогорова, ω^2 Крамера-Мизеса-Смирнова и Ω^2 Андерсона-Дарлинга при проверке сложных гипотез относительно ряда законов распределения вероятностей. Эти результаты уточняют и расширяют, рекомендации по стандартизации Р 50.1.037-2002 [36].

В следующей части работы будут представлены модели распределений статистик критериев и таблицы процентных точек для проверки сложных гипотез относительно семейств гамма-распределений и двустороннего экспоненциального (табл. 1).

Настоящие исследования выполнены при частичной поддержке Российского фонда фундаментальных исследований (проект № 06-01-00059а) и Федерального агентства по образованию Минобрнауки РФ в рамках Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" (проект № 2.1.2/3970).

Литература

1. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione // *Giornale dell' Istituto Italiano degli Attuari*. 1933. – Vol. 4. – № 1. – P. 83-91.
2. Большев Л.Н. Асимптотические пирсоновские преобразования // *Теория вероятностей и ее применения*. 1963. – Т. 8. – № 2. – С. 129-155.
3. Большев Л.Н. Теория вероятностей и математическая статистика / *Избранные труды*. Под ред. Ю.В. Прохорова. – М.: Наука, 1987. – 286 с.
4. Большев Л.Н., Смирнов Н.В. *Таблицы математической статистики*. – М.: Наука, 1983. – 416 с.
5. Anderson T.W., and Darling D.A. Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes // *Ann. Math. Statist.*, 1952. V.23. – P.193-212.
6. Anderson T.W., and Darling D.A. A test of goodness of fit // *J. Amer. Stist. Assoc.*, 1954. V.29. – P.765-769.
7. Кас М., Kiefer J., Wolfowitz J. On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods // *Ann. Math. Stat.*, 1955. V.26. – P.189-211.
8. Darling D.A. The Cramer-Smirnov test in the parametric case // *Ann. Math. Statist.*, 1955. V.26. – P.1-20.
9. Darling D.A. The Cramer-Smirnov test in the parametric case // *Ann. Math. Statist.*, 1957. V.28. – P.823-838.
10. Lilliefors H.W. On the Kolmogorov-Smirnov test for normality with mean and variance unknown // *J. Am. Statist. Assoc.*, 1967. V.62. – P.399-402.
11. Lilliefors H.W. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown // *J. Am. Statist. Assoc.*, 1969. V.64. – P.387-389.
12. Durbin J. Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.*, 1973. 1: 279-290.

13. Durbin J. Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests of spacings // *Biometrika*, 1975. V.62. – P.5-22.
14. Durbin J. Kolmogorov–Smirnov Test when Parameters are Estimated // *Lect. Notes Math.*, 1976. V.566. – P.33–44.
15. Гихман И.И. Некоторые замечания о состоятельности критерия А.Н. Колмогорова // *Докл. Акад. Наук СССР*, 1953. Т.91 (4). – С. 715-718.
16. Gihman I.I. On the empirical distribution function in the case of grouping data / In *Selected Translations in Mathematical Statistics and Probability*, 1961. V.1. American Math. Soc., Providence, RI. – P.77-81.
17. Мартынов Г.В. Критерии омега–квадрат. – М.: Наука, 1978. – 80 с.
18. Dzhaparidze K.O., Nikulin M.S. Probability distribution of the Kolmogorov and omega-square statistics for continuous distributions with shift and scale parameters // *J. Soviet Math.*, 1982. V.20. – P.2147-2163.
19. Nikulin M.S. Gihman and goodness-of-fit tests for grouped data. *Mathematical Reports of the academy of Science of the Royal Society of Canada*, 1992. 14 (4): 151-156.
20. Nikulin M.S. A variant of the generalized omega-square statistic // *J. Soviet Math.*, 1992. V.61 (4). – P.1896-1900, (translation from *Zapiski nauchnikh seminarov LOMI*, (1989), 177. – P.108-113.
21. Pearson E.S., Hartley H.O. *Biometrika Tables for Statistics*. 1972. V.2. Cambridge: University Press.
22. Stephens M.A. Use of Kolmogorov–Smirnov, Cramer – von Mises and Related Statistics – Without Extensive Table // *J. R. Stat. Soc.*, 1970. V.32. – P.115-122.
23. Stephens M.A. EDF Statistics for Goodness of Fit and Some Comparisons // *J. Am. Statist. Assoc.*, 1974. V.69. – P.730-737.
24. Chandra M., Singpurwalla N.D., Stephens M.A. *Statistics for Test of Fit for the*

- Extrem-Value and Weibull Distribution // J. Am. Statist. Assoc., 1981. V.76(375). – P.729-731.
- 25.Тюрин Ю.Н. О предельном распределении статистик Колмогорова–Смирнова для сложной гипотезы // Изв. АН СССР. Сер. Матем. 1984. – Т.48. – № 6. – С. 1314-1343.
- 26.Тюрин Ю.Н., Саввушкина Н.Е. Критерии согласия для распределения Вейбулла–Гнеденко // Изв. АН СССР. Сер. Техн. Кибернетика. 1984. – № 3. – С.109-112.
- 27.Саввушкина Н.Е. Критерий Колмогорова–Смирнова для логистического и гамма–распределения // Сб. тр. ВНИИ систем. исслед. – 1990, № 8. – С.50-56.
- 28.Лемешко Б.Ю., Постовалов С.Н. Прикладные аспекты использования критериев согласия в случае проверки сложных гипотез // Надежность и контроль качества. – 1997. – № 11. – С. 3-17.
- 29.Лемешко Б.Ю., Постовалов С.Н. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов // Заводская лаборатория. 1998. Т.64. - № 3. – С.61-72. [Lemeshko B.Yu., Postovalov S.N. Statistical distributions of nonparametric goodness-of-fit tests as estimated by the sample parameters of experimentally observed laws // Industrial laboratory (Ind. lab.). 1998, V. 64, n°3, - P.197-208 (Consultants Bureau, New York)]
- 30.Лемешко Б.Ю., Постовалов С.Н. О правилах проверки согласия опытного распределения с теоретическим // Методы менеджмента качества. Надежность и контроль качества. - 1999. № 11. – С. 34-43.
- 31.Лемешко Б.Ю., Постовалов С.Н. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Часть II. Непараметрические критерии. – Новосибирск: Изд-

- во НГТУ, 1999. – 85 с.
32. Лемешко Б.Ю., Постовалов С.Н. Применение непараметрических критериев согласия при проверке сложных гипотез // Автометрия. 2001. – № 2. – С. 88-102. [Lemeshko B.Yu., Postovalov S.N. Application of the nonparametric goodness-of-fit Tests in testing composite hypotheses // Optoelectronics, Instrumentation and Data Processing. 2001. – № 2. – P. 76-88]
33. Лемешко Б.Ю., Постовалов С.Н. Непараметрические критерии при проверке сложных гипотез о согласии с распределениями Джонсона // Доклады СО АН ВШ. 2002. – № 1(5). – С.65-74.
34. Лемешко Б.Ю., Маклаков А.А. Непараметрические критерии при проверке сложных гипотез о согласии с распределениями экспоненциального семейства // Автометрия. 2004. №3. – С. 3-20. [Lemeshko B.Yu., Maklakov A.A. Nonparametric Test in Testing Composite Hypotheses on Goodness of Fit Exponential Family Distributions. Optoelectronics, Instrumentation and Data Processing, V. 40, 3. - P.3-18, 2004]
35. Design of experiments and statistical analysis for grouped observations: Monograph / V.I. Denisov, K.-H. Eger, B.Yu. Lemeshko, E.B. Tsoy. – Novosibirsk: NSTU Publishing house, 2004. – 464 p.
36. Р 50.1.037-2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. – М.: Изд-во стандартов. 2002. – 64 с.

Statistics distributions models of the nonparametric goodness-of-fit tests in case composite hypothesis testing with using of maximum likelihood estimations. Part I

Lemeshko B. Yu., Lemeshko S.B.

Novosibirsk state technical university

E-mail: Lemeshko@fpm.ami.nstu.ru

In composite hypotheses testing, when the estimate of the scalar or vector parameter of the probabilities distribution laws is calculated by the same sample, the nonparametric goodness-of-fit Kolmogorov, Cramer-Mises-Smirnov, Anderson-Darling tests lose the free distribution property. In testing of composite hypotheses, the conditional distribution law of the statistic is affected by a number of factors.

In this paper we present more precise results (tables of percentage points and statistic distribution models) for the nonparametric goodness-of-fit tests in testing composite hypotheses using the maximum likelihood estimate (MLE) for some probabilities distribution laws.

Keywords: goodness-of-fit test, composite hypotheses testing, Kolmogorov test, Cramer-Mises-Smirnov test, Anderson-Darling test.