Измерительная техника. 2013. № 9. – С.14-21

519.24

Применение непараметрических критериев согласия Купера и Ватсона при проверке сложных гипотез

Б. Ю. ЛЕМЕШКО, А. А. ГОРБУНОВА

Новосибирский государственный технический университет, Новосибирск, Россия, e-mail: lemeshko@fpm.ami.nstu.ru

Рассмотрены модели распределений статистик и таблицы процентных точек, позволяющие применять критерии согласия Купера и Ватсона при проверке сложных гипотез относительно принадлежности выборок различным параметрическим моделям законов распределения вероятностей. Приведена методика интерактивного моделирования, дающая возможность построить и использовать распределение статистики критерия в ходе решения задачи статистического анализа, связанного с проверкой гипотезы.

Ключевые слова: непараметрические критерии согласия, критерии Купера и Ватсона, простые и сложные гипотезы.

The statistics distributions models and tables of percentage points allowing to use the Kuiper and Watson goodness-of-fit tests at testing the composite hypotheses on belonging of samples to different parametric models of statistical laws of probabilities distribution are considered. The procedure of interactive simulation allowing to construct and use the statistics distribution of test during the solution of problem of statistical analysis inherent to hypothesis testing is presented.

Key words: nonparametric goodness-of-fit test, Kuiper`s and Watson`s tests, simple and composite hypotheses.

В [1] при исследовании распределения статистик и мощности непараметрических критериев согласия Купера [2] и Ватсона [3] было показано, что в случае проверки простых гипотез они имеют некоторое преимущество в мощности перед критериями Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга. Это позволяет рекомендовать их применение в различных приложениях. При сложных проверяемых гипотезах такого заметного преимущества не наблюдается, однако ясно, что критерии Купера и Ватсона целесообразно использовать наряду с другими упомянутыми выше критериями согласия.

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 теряют свойство свободы от распределения [4]. Условные распределения $G(S|H_0)$ статистик критериев при проверке сложных гипотез зависят от ряда факторов [5]: вида наблюдаемого закона $F(x,\theta)$, соответствующего справедливой проверяемой гипотезе H_0; типа оцениваемого параметра и их числа; в некоторых случаях от конкретного значения параметра (например, для семейств гамма- и бета-распределений); метода оценивания параметров. Различия в распределениях той же самой статистики при проверке простой и сложных гипотез настолько существенны, что пренебрежение этим фактом, как правило, приводит к неправильному использованию критерия и, следовательно, к некорректным статистическим выводам.

К решению проблемы проверки сложных гипотез при помощи критериев Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга использовались различные подходы [6–11], а в [12, 13] применен компьютерных подход и статистическое моделирование, на базе которых разработаны рекомендации по применению непараметрических критериев согласия [14, 15]. Позднее эти результаты были уточнены и расширены [16–24] и в настоящее время наиболее полно представлены в [5].

В данной статье приведены таблицы процентных точек и модели распределений статистик критериев Купера и Ватсона, которые рекомендуются при проверке сложных гипотез относительно некоторых часто применяемых в приложениях параметрических законов распределения случайных величин.

Для проверки гипотезы о том, что случайная выборка принадлежит закону с непрерывной функцией распределения $F\left(x,\theta\right)$ в критерии Купера [2] используется статистика вида

$$V_n = \sup_{-\infty<x<\infty}\left\{F_n\left(x\right)-F\left(x,\theta\right)\right\} - \inf_{-\infty<x<\infty}\left\{F_n\left(x\right)-F\left(x,\theta\right)\right\},$$

которая вычисляется в соответствии с выражением

$$V_n = D_n^+ + D_n^-, \tag{1}$$

где $F_n(x)$ – эмпирическая функция распределения; $D_n^+ = \max\left\{\frac{i}{n} - F\left(x_i,\theta\right)\right\}$, $D_n^- = \max\left\{F\left(x_i,\theta\right) - \frac{i-1}{n}\right\}$; x_i – элементы вариационного ряда, построенного по выборке. В качестве предельного распределения $G\left(\sqrt{n}V_n\middle|H_0\right)$ статистики $\sqrt{n}V_n$ Купером [2] дана функция распределения $G\left(s\middle|H_0\right)=1-\sum_{m=1}^{\infty}2(4m^2s^2-1)e^{-2m^2s^2}$.

Чтобы снизить зависимость распределения $G(V_n|H_0)$ статистики (1) от объема выборки n, в [25] предложена ее модификация

$$V = V_n\left(\sqrt{n} + 0{,}155 + 0{,}24/\sqrt{n}\right). \qquad (2)$$

В [1], следуя [26], в критерии Купера предложено применять статистику

$$V_n^{\text{mod}} = \sqrt{n}(D_n^+ + D_n^-) + (3\sqrt{n})^{-1}. \qquad (3)$$

При проверке простых гипотез процентные точки и распределения статистик (2), (3) практически совпадают, а в качестве модели предельного распределения статистики (3) можно использовать бета-распределение 3-го рода с плотностью [1]:

$$f(x) = \frac{\theta_2^{\theta_0}}{\theta_3 \mathrm{B}(\theta_0, \theta_1)} \frac{\left(X\right)^{\theta_0 - 1}\left(1 - X\right)^{\theta_1 - 1}}{\left[1 + (\theta_2 - 1)X\right]^{\theta_0 + \theta_1}}\ , \qquad (4)$$

где $X = (x - \theta_4)/\theta_3$, и вектором параметров θ = (7,8985; 7,6865; 2,6852; 2,6373; 0,493)$^{\text{т}}$.

Критерий Ватсона [3] используется со статистикой вида

$$U_n^2 = \sum_{i=1}^{n}\left(F\left(x_i, \theta\right) - \frac{i - 1/2}{n}\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} F\left(x_i, \theta\right) - \frac{1}{2}\right)^2 + \frac{1}{12n}. \qquad (5)$$

При проверке простых гипотез предельное распределение $G\left(U_n^2 \middle| H_0\right)$ статистики U_n^2 (5) имеет вид [3] $G\left(s \middle| H_0\right) = 1 - 2\sum_{m=1}^{\infty}(-1)^{m-1} e^{-2m^2\pi^2 s}$ и по всей области определения хорошо приближается моделью обратного гауссова закона с плотностью [1]:

$$f(x) = \sqrt{\frac{\theta_0}{2\pi Y^3}} \exp\left(-\frac{\theta_0\left(Y - \theta_1\right)^2}{2\theta_1^2 Y}\right), \qquad (6)$$

где $Y = (x - \theta_3)/\theta_2$, и вектором параметров $\boldsymbol{\theta} = \left(0{,}5555;\ 0{,}2385;\ 0{,}3437;\ 0{,}0015\right)^{\text{т}}$.

Ниже представлены модели предельных распределений и таблицы процентных точек, которые позволяют применять критерии Купера и Ватсона для проверки сложных гипотез относительно принадлежности анализируемых выборок различным параметрическим моделям законов распределения. Предполагается, что используются оценки максимального правдоподобия.

В табл. 1 приведены законы распределения, относительно которых можно проверять сложные гипотезы, используя построенные в данной работе приближения для предельных распределений статистик непараметрических критериев согласия.

Т а б л и ц а 1

Законы распределения случайных величин

Наименование закона	Функция плотности $f(x,\theta)$
Экспоненциальный	$\theta_0^{-1}\exp[-x/\theta_0]$
Полунормальный	$2(\theta_0\sqrt{2\pi})^{-1}\exp[-x^2/(2\theta_0^2)]$
Рэлея	$(x/\theta_0^2)\exp[-x^2/(2\theta_0^2)]$
Максвелла	$[2x^2/(\theta_0^3\sqrt{2\pi})]\exp[-x^2/(2\theta_0^2)]$
Лапласа	$(2\theta_0)^{-1}\exp[-\lvert x-\theta_1\rvert/\theta_0]$
Нормальный (Гаусса)	$(\theta_0\sqrt{2\pi})^{-1}\exp[-(x-\theta_1)^2/(2\theta_0^2)]$
Логарифмически нормальный	$(x\theta_0\sqrt{2\pi})^{-1}\exp[-(\ln x-\theta_1)^2/(2\theta_0^2)]$
Коши	$\theta_0/\{\pi[\theta_0^2+(x-\theta_1)^2]\}$
Логистический	$\dfrac{\pi}{\theta_0\sqrt{3}}\exp\left\{-\dfrac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\Bigg/\left[1+\exp\left\{-\dfrac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2$
Экстремального значения (maximum)	$\dfrac{1}{\theta_0}\exp\left\{-\dfrac{x-\theta_1}{\theta_0}-\exp\left(-\dfrac{x-\theta_1}{\theta_0}\right)\right\}$
Экстремального значения (minimum)	$\dfrac{1}{\theta_0}\exp\left\{\dfrac{x-\theta_1}{\theta_0}-\exp\left(\dfrac{x-\theta_1}{\theta_0}\right)\right\}$
Вейбулла	$\theta_0 x^{\theta_0-1}\theta_1^{-\theta_0}\exp\left\{-\left(x/\theta_1\right)^{\theta_0}\right\}$
Sb-Джонсона $Sb(\theta_0,\theta_1,\theta_2,\theta_3)$	$\dfrac{\theta_1\theta_2}{(x-\theta_3)(\theta_2+\theta_3-x)}\exp\left\{-\dfrac{1}{2}\left[\theta_0-\theta_1\ln\dfrac{x-\theta_3}{\theta_2+\theta_3-x}\right]^2\right\}$
Sl-Джонсона $Sl(\theta_0,\theta_1,\theta_2,\theta_3)$	$\dfrac{\theta_1}{(x-\theta_3)\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2}\left[\theta_0+\theta_1\ln\dfrac{x-\theta_3}{\theta_2}\right]^2\right\}$
Su-Джонсона $Su(\theta_0,\theta_1,\theta_2,\theta_3)$	$\dfrac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2+\theta_2^2}}\exp\left\{-\dfrac{1}{2}\left[\theta_0+\theta_1\ln\left\{\dfrac{x-\theta_3}{\theta_2}+\sqrt{\left(\dfrac{x-\theta_3}{\theta_2}\right)^2+1}\right\}\right]^2\right\}$

Таблицы процентных точек и модели распределений статистик критериев строили по смоделированным выборкам статистик объемом

$N = 1{,}7 \cdot 10^6$. При таких N разность между истинным законом $G(S|H_0)$ распределения статистики и смоделированным эмпирическим $G_N(S|H_0)$ по модулю не превышает 10^{-3}. Значения статистик критериев вычислялись по выборкам псевдослучайных величин объемом $n = 10^3$, генерируемых в соответствии с наблюдаемым законом $F(x,\theta)$. В такой ситуации распределение $G(S_n|H_0)$ практически совпадает с предельным $G(S|H_0)$. В задачах статистического анализа можно пользоваться представленными в статье моделями, начиная с объемов выборок $n > 25$.

Распределения $G(S|H_0)$ статистик критериев Купера и Ватсона наилучшим образом аппроксимируются семейством бета-распределений III рода с функцией плотности (4), т.е. $\mathrm{B}_3(\theta_0,\theta_1,\theta_2,\theta_3,\theta_4) = f(x)$, и семейством распределений *Sl*-Джонсона $Sl(\theta_0,\theta_1,\theta_2,\theta_3)$ (см. табл. 1).

Верхние процентные точки и построенные модели для предельных распределений статистики критерия Купера в случае использования оценки максимального правдоподобия (ОМП) представлены в табл. 2 для 12 законов (кроме трех последних в табл. 1). Для тех же законов распределения верхние процентные точки и построенные модели для распределений статистики критерия Ватсона приведены в табл. 3.

В табл. 4 представлены 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 табл. 5 – законов *Sl*-Джонсона, в табл. 6 – законов *Su*-Джонсона. Во всех перечисленных случаях распределения $G(S|H_0)$ статистик критериев согласия не зависят от конкретных значений неизвестных параметров законов $F(x,\theta)$.

Когда распределения статистик непараметрических критериев согласия зависят от значений параметра или параметров закона, с которым проверяется согласие, задача может решаться следующим образом. Так как оценки параметров становятся известными в процессе анализа, требуемое для проверки гипотезы распределение статистики нельзя найти заранее.

Т а б л и ц а 2

Верхние процентные точки и модели предельных распределений статистики критерия Купера при использовании ОМП

Наименование закона	Оцениваемые параметры	Процентные точки			Модель
		0,1	0,05	0,01	
Экспоненциальное, Рэлея, Максвелла	Масштаб	1,540	1,661	1,905	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
Полунормальное	То же	1,543	1,664	1,907	$B_3(11{,}4707;40{,}7237;7{,}020;20{,}3675;0{,}3989)$
Лапласа	″–″	1,469	1,587	1,825	$B_3(7{,}8324;8{,}3778;2{,}6906;2{,}4820;0{,}4830)$
	Сдвиг	1,473	1,597	1,850	$B_3(9{,}1630;6{,}6097;4{,}0210;2{,}4081;0{,}4900)$
	Оба параметра	1,278	1,365	1,541	$B_3(10{,}0376;7{,}8452;3{,}4694;1{,}9586;0{,}4756)$
Нормальное, логарифмически нормальное	Масштаб	1,494	1,611	1,847	$B_3(6{,}3057;8{,}1797;2{,}3279;2{,}4413;0{,}5370)$
	Сдвиг	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
	Оба параметра	1,402	1,505	1,709	$B_3(7{,}4917;8{,}0016;2{,}4595;2{,}1431;0{,}4937)$
Коши	Масштаб или сдвиг	1,435	1,560	1,815	$B_3(3{,}8425;5{,}9345;2{,}4284;2{,}1927;0{,}6150)$
	Оба параметра	1,126	1,197	1,337	$B_3(9{,}4267;7{,}5349;3{,}2515;1{,}5491;0{,}4700)$
Логистическое	Масштаб	1,470	1,588	1,826	$B_3(9{,}7224;7{,}8186;3{,}2399;2{,}4541;0{,}4370)$
	Сдвиг	1,511	1,633	1,880	$B_3(9{,}1363;6{,}9693;3{,}4630;2{,}3985;0{,}4790)$
	Оба параметра	1,337	1,432	1,622	$B_3(14{,}3460;18{,}6137;3{,}6366;3{,}9560;0{,}3525)$
Экстремальных значений & Вейбулла	Масштаб[1]	1,504	1,622	1,861	$Sl(1{,}2459;4{,}0123;1{,}3063;0{,}1873)$
	Сдвиг[2]	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
	Оба параметра	1,411	1,516	1,726	$Sl(1{,}4012;5{,}0846;1{,}4465;–0{,}0070)$

Примечание: [1], [2] – при оценивании параметра соответственно формы и масштаба распределения Вейбулла.

Таблица 3

Верхние процентные точки и модели предельных распределений статистики критерия Ватсона при использовании ОМП

Наименование закона	Оцениваемые параметры	Процентные точки			Модель
		0,1	0,05	0,01	
Экспоненциальное, Рэлея, Максвелла	Масштаб	0,129	0,159	0,230	B_3(4,0419;2,9119;10,5931;0,5000;0,0096)
Полунормальное	То же	0,131	0,161	0,232	B_3(4,9988;3,8721;15,1781;0,6900;0,0059)
Лапласа	″–″	0,115	0,144	0,214	B_3(9,2136;3,8610;30,5491;0,7010;0,0015)
	Сдвиг	0,111	0,139	0,209	B_3(7,4479;3,2650;30,7784;0,6227;0,0063)
	Оба параметра	0,071	0,084	0,114	B_3(9,0116;5,3554;17,3201;0,3908;0,0038)
Нормальное, логарифмически нормальное	Масштаб	0,122	0,151	0,221	B_3(8,8122;3,7536;29,8074;0,7171;0,0019)
	Сдвиг	0,127	0,157	0,228	B_3(3,6769;4,4438;9,8994;0,6805;0,0082)
	Оба параметра	0,096	0,116	0,164	B_3(3,5230;4,4077;9,2281;0,4785;0,0104)
Коши	Масштаб или сдвиг	0,105	0,133	0,203	Sl(2,7778;1,5065;0,2690;0,0049)
	Оба параметра	0,052	0,061	0,081	B_3(8,3558;4,8650;12,0768;0,1930;0,0049)
Логистическое	Масштаб	0,115	0,144	0,214	B_3(9,2136;3,8610;30,5491;0,7010;0,0015)
	Сдвиг	0,119	0,148	0,218	B_3(3,9730;3,9414;13,2655;0,6637;0,0090)
	Оба параметра	0,081	0,098	0,135	B_3(4,2608;4,6784;9,3054;0,3810;0,0084)
Экстремальных значений, Вейбулла	Масштаб[1]	0,122	0,151	0,221	B_3(8,8122;3,7536;29,8074;0,7171;0,0019)
	Сдвиг[2]	0,129	0,159	0,230	B_3(4,9988;3,8721;15,1781;0,6792;0,0061)
	Оба параметра	0,097	0,118	0,165	Sl(1,2863;1,6736;0,0927;0,0052)

Примечание: [1], [2] – те же, что и в табл. 2.

Т а б л и ц а 4

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
	0,1	0,05	0,01	
Критерий Купера				
θ_0	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
θ_1	1,494	1,611	1,847	$B_3(6{,}3057;8{,}1797;2{,}3279;2{,}4413;0{,}5370)$
θ_0,θ_1	1,402	1,505	1,709	$B_3(7{,}4917;8{,}0016;2{,}4595;2{,}1431;0{,}4937)$
Критерий Ватсона				
θ_0	0,127	0,157	0,228	$B_3(3{,}6769;4{,}4438;9{,}8994;0{,}6805;0{,}0082)$
θ_1	0,122	0,151	0,221	$B_3(8{,}8122;3{,}7536;29{,}8074;0{,}7171;0{,}0019)$
θ_0,θ_1	0,096	0,116	0,164	$B_3(3{,}5230;4{,}4077;9{,}2281;0{,}4785;0{,}0104)$

Таблица 5

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
	0,1	0,05	0,01	
Критерий Купера				
θ_0	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
θ_1	1,512	1,631	1,872	$B_3(6{,}7423;8{,}0549;2{,}4935;2{,}4976;0{,}5250)$
θ_2	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
θ_0,θ_1	1,402	1,505	1,709	$B_3(7{,}4917;8{,}0016;2{,}4595;2{,}1431;0{,}4937)$
θ_0,θ_2	1,540	1,662	1,908	$B_3(5{,}5932;7{,}6149;2{,}1484;2{,}3961;0{,}5630)$
θ_1,θ_2	1,402	1,505	1,709	$B_3(7{,}4917;8{,}0016;2{,}4595;2{,}1431;0{,}4937)$
$\theta_0,\theta_1,\theta_2$	1,402	1,505	1,709	$B_3(7{,}4917;8{,}0016;2{,}4595;2{,}1431;0{,}4937)$
Критерий Ватсона				
θ_0	0,127	0,157	0,228	$B_3(3{,}6769;4{,}4438;9{,}8994;0{,}6805;0{,}0082)$
θ_1	0,124	0,153	0,223	$B_3(3{,}4122;4{,}9262;9{,}6902;0{,}7643;0{,}0087)$
θ_2	0,127	0,157	0,228	$B_3(3{,}6769;4{,}4438;9{,}8994;0{,}6805;0{,}0082)$

θ_0,θ_1	0,096	0,116	0,164	$B_3(3,5230;4,4077;9,2281;0,4785;0,0104)$
θ_0,θ_2	0,127	0,157	0,228	$B_3(3,6769;4,4438;9,8994;0,6805;0,0082)$
θ_1,θ_2	0,096	0,116	0,164	$B_3(3,5230;4,4077;9,2281;0,4785;0,0104)$
$\theta_0,\theta_1,\theta_2$	0,096	0,116	0,164	$B_3(3,5230;4,4077;9,2281;0,4785;0,0104)$

Таблица 6

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
	0,1	0,05	0,01	
Критерий Купера				
θ_0	1,540	1,662	1,908	$B_3(5,5932;7,6149;2,1484;2,3961;0,5630)$
θ_1	1,512	1,631	1,872	$B_3(6,7676;8,3605;2,3501;2,4976;0,5142)$
θ_2	1,491	1,612	1,857	$B_3(7,5884;8,1397;2,6781;2,4982;0,4882)$
θ_3	1,517	1,638	1,885	$B_3(8,1449;7,2651;3,0338;2,4418;0,4880)$
θ_0,θ_1	1,402	1,505	1,709	$B_3(8,1449;7,2650;3,0338;2,1431;0,5015)$
θ_0,θ_2	1,393	1,496	1,703	$B_3(7,5234;7,3134;2,7694;2,1076;0,5035)$
θ_0,θ_3	1,390	1,496	1,713	$B_3(8,0187;7,7542;2,7862;2,1751;0,4800)$
θ_1,θ_2	1,414	1,525	1,749	$B_3(8,6702;7,5387;2,9284;2,2036;0,4600)$
θ_1,θ_3	1,375	1,475	1,675	$B_3(8,6702;7,5387;2,9284;2,0887;0,4740)$
θ_2,θ_3	1,350	1,447	1,640	$B_3(9,0132;7,9999;2,8585;2,0644;0,4635)$
$\theta_0,\theta_1,\theta_2$	1,324	1,422	1,621	$B_3(10,7806;8,4043;3,2432;2,1461;0,4150)$
$\theta_0,\theta_1,\theta_3$	1,333	1,431	1,629	$B_3(10,3455;8,0495;3,5687;2,1993;0,4463)$
$\theta_0,\theta_2,\theta_3$	1,296	1,388	1,575	$B_3(10,3223;7,7893;3,3393;2,0021;0,4358)$
$\theta_1,\theta_2,\theta_3$	1,299	1,394	1,584	$B_3(10,5957;8,2600;3,2334;2,0676;0,4194)$
$\theta_0,\theta_1,\theta_2,\theta_3$	1,235	1,321	1,494	$B_3(9,9689;7,3418;3,4037;1,8225;0,4438)$
Критерий Ватсона				
θ_0	0,127	0,157	0,228	$B_3(3,6769;4,4438;9,8994;0,6805;0,0082)$
θ_1	0,124	0,153	0,223	$B_3(3,4122;4,9262;9,6902;0,7643;0,0087)$
θ_2	0,117	0,146	0,215	$B_3(6,0296;3,7175;22,6978;0,7115;0,0057)$
θ_3	0,121	0,150	0,220	$B_3(7,4154;3,9208;22,4649;0,6800;0,0022)$
θ_0,θ_1	0,096	0,116	0,164	$B_3(3,5230;4,4077;9,2281;0,4785;0,0104)$
θ_0,θ_2	0,093	0,114	0,161	$B_3(4,0651;4,8643;9,5614;0,4903;0,0078)$

θ_0, θ_3	0,092	0,113	0,162	$B_3(4,4170;4,9456;10,4292;0,5005;0,0067)$
θ_1, θ_2	0,099	0,123	0,181	$B_3(5,5181;4,1815;16,0852;0,5478;0,0055)$
θ_1, θ_3	0,089	0,108	0,151	$B_3(5,7461;4,4051;13,9768;0,4528;0,0060)$
θ_2, θ_3	0,084	0,101	0,141	$B_3(5,9952;4,3409;13,8757;0,4020;0,0060)$
$\theta_0, \theta_1, \theta_2$	0,077	0,093	0,131	$B_3(5,5809;4,9570;14,1052;0,4540;0,0060)$
$\theta_0, \theta_1, \theta_3$	0,080	0,097	0,137	$B_3(5,8959;4,4478;14,5923;0,4132;0,0060)$
$\theta_0, \theta_2, \theta_3$	0,072	0,087	0,121	$B_3(6,1780;4,6712;14,5568;0,3791;0,0060)$
$\theta_1, \theta_2, \theta_3$	0,072	0,087	0,121	$B_3(6,1780;4,6712;14,5568;0,3791;0,0060)$
$\theta_0, \theta_1, \theta_2, \theta_3$	0,062	0,074	0,101	$B_3(7,3816;4,4215;14,1896;0,2616;0,0053)$

Отсюда следует, что распределения статистик применяемых критериев должны находиться в интерактивном режиме в ходе проводимого статистического анализа. Конечно, это требует развитого программного обеспечения, позволяющего (как в нашем случае) в целях ускорения распараллеливать процессы моделирования и использовать доступные вычислительные ресурсы. В таких условиях время построения (с требуемой точностью) необходимого для проверки гипотезы распределения $G_N(S_n|H_0)$ статистики критерия и определения по нему достигнутого уровня значимости $P\{S_n \geq S^*\}$, где S^* – вычисленное по выборке значение статистики, оказывается не очень заметным на фоне полного решения задачи статистического анализа.

Следующий пример демонстрирует точность определения достигнутого уровня значимости в зависимости от объема выборки N моделируемого в интерактивном режиме эмпирического распределения статистики.

Пример. Необходимо проверить сложную гипотезу о принадлежности обратному гауссову закону с плотностью (6) следующей выборки объемом $n=100$:

0,945	1,040	0,239	0,382	0,398	0,946	1,248	1,437	0,286	0,987
2,009	0,319	0,498	0,694	0,340	1,289	0,316	1,839	0,432	0,705
0,371	0,668	0,421	1,267	0,466	0,311	0,466	0,967	1,031	0,477
0,322	1,656	1,745	0,786	0,253	1,260	0,145	3,032	0,329	0,645
0,374	0,236	2,081	1,198	0,692	0,599	0,811	0,274	1,311	0,534
1,048	1,411	1,052	1,051	4,682	0,111	1,201	0,375	0,373	3,694
0,426	0,675	3,150	0,424	1,422	3,058	1,579	0,436	1,167	0,445
0,463	0,759	1,598	2,270	0,884	0,448	0,858	0,310	0,431	0,919
0,796	0,415	0,143	0,805	0,827	0,161	8,028	0,149	2,396	2,514
1,027	0,775	0,240	2,745	0,885	0,672	0,810	0,144	0,125	1,621

По выборке оцениваются параметры формы θ_0, θ_1 и масштабный параметр θ_2, параметр сдвига $\theta_3 = 0$ предполагается известным. Найденные по данной выборке ОМП параметров $\hat{\theta}_0 = 0,7481$, $\hat{\theta}_1 = 0,7806$, $\hat{\theta}_2 = 1,3202$. Вычисленные значения статистик: Купера – $V_n^{\mathrm{mod}} = 1,1113$, Ватсона – $U_n^2 = 0,05200$. Распределения статистик критериев в данном случае зависят от θ_0 и θ_1 [20, 22], не зависят от θ_2 и должны быть найдены при $\theta_0 = 0,7481$, $\theta_1 = 0,7806$.

Достигнутые уровни значимости по критериям $P\{V_n^{\mathrm{mod}} \geq 1,1113\}$ и $P\{U_n^2 \geq 0,05200\}$, полученные при различной точности моделирования распределений статистик (при различном объеме N моделируемых выборок статистик), приведены в табл. 7.

Таблица 7

Уровни значимости для статистик Купера и Ватсона при различных объемах выборок *N*

Статистика	10^3	10^4	10^5	10^6
$V_n^{\mathrm{mod}} = 1,1113$	0,479	0,492	0,493	0,492
$U_n^2 = 0,05200$	0,467	0,479	0,483	0,482

Таким образом, приведенные модели распределений статистик и таблицы процентных точек позволяют корректно применять критерии Купера и Ватсона при проверке сложных гипотез относительно ряда параметрических моделей законов распределения. Реализованная интерактивная методика дает возможность корректного применения критериев, когда распределение статистики критерия, соответствующее справедливости проверяемой гипотезы H_0 к моменту использования, неизвестно.

Работа выполнена при поддержке Минобрнауки РФ в рамках госзадания (проект 8.1274.2011) и ФЦП «Научные и научно-педагогические кадры инновационной России» (соглашение № 14.B37.21.0860).

Литература

1. **Лемешко Б. Ю., Горбунова А. А.** О применении и мощности непараметрических критериев согласия Купера, Ватсона и Жанга // Измерительная техника. 2013. № 5. С. 3–9.

2. **Kuiper N. H.** Tests concerning random points on a circle // Proc. Konikl. Nederl. Akad. Van Wettenschappen. Ser. A. 1960. V. 63. P. 38–47.

3. **Watson G. S.** Goodness-of-fit tests on a circle. I // Biometrika. 1961. V. 48. N. 1–2. P. 109–114.

4. **Kac M., 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.

5. **Лемешко Б. Ю. и др.** Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: Монография. Новосибирск: Изд-во НГТУ, 2011.

6. **Darling D. A.** The Cramer–Smirnov test in the parametric case // Ann. Math. Statist. 1955. V. 26. P. 1–20.

7. **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.

8. **Durbin J.** Weak convergence of the sample distribution function when parameters are estimated // Ann. Statist. 1973. N. 1. P. 279–290.

9. **Мартынов Г. В.** Критерии омега-квадрат. М.: Наука, 1978.

10. **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.

11. **Тюрин Ю. Н.** О предельном распределении статистик Колмогорова–Смирнова для сложной гипотезы // Изв. АН СССР. Сер. Математическая. 1984. Т. 48. № 6. С. 1314–1343.

12. **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.

13. **Лемешко Б. Ю., Постовалов С. Н.** Применение непараметрических критериев согласия при проверке сложных гипотез // Автометрия. 2001. № 2. С. 88–102.

14. **Лемешко Б. Ю., Постовалов С. Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Ч. II. Непараметрические критерии. Новосибирск: Изд-во НГТУ, 1999.

15. **Р 50.1.037–2002.** Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. II. Непараметрические критерии.

16. **Лемешко Б. Ю., Маклаков А. А.** Непараметрические критерии при проверке сложных гипотез о согласии с распределениями экспоненциального семейства // Автометрия. 2004. № 3. С. 3–20.

17. **Лемешко С. Б., Лемешко Б. Ю.** Распределения статистик непараметрических критериев согласия при проверке гипотез относительно бета-распределений // ДАН ВШ России. 2007. № 2(9). С. 6–16.

18. **Лемешко Б. Ю., Лемешко С. Б.** Модели распределений

статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. I // Измерительная техника. 2009. № 6. С. 3–11.; **Lemeshko B. Yu., Lemeshko S. B.** Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Pt. I // Measurement Techniques. 2009. V. 52. N 6. P. 555–565.

19. **Лемешко Б. Ю., Лемешко С. Б.** Модели распределений статистик непараметрических критериев согласия при проверке сложных гипотез с использованием оценок максимального правдоподобия. Ч. II // Измерительная техника. 2009. № 8. С. 17–26; **Lemeshko B. Yu., Lemeshko S. B.** Models for statistical distributions in nonparametric fitting tests on composite hypotheses based on maximum-likelihood estimators. Pt II // Measurement Techniques. 2009. V. 52. N 8. P. 799–812.

20. **Лемешко Б. Ю., Лемешко С. Б., Никулин М. С., Саaидиа Н.** Моделирование распределений статистик непараметрических критериев согласия при проверке сложных гипотез относительно обратного гауссовского закона // Автоматика и телемеханика, 2010. № 7. С. 83–102.

21. **Lemeshko B. Yu., Lemeshko S. B., Postovalov S. N.** Statistic Distribution Models for Some Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses // Communications in Statistics–Theory and Methods, 2010. V. 39. N 3. P. 460–471.

22. **Lemeshko B. Yu. e.a.** Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis // Mathematical and Statistical Models and Methods in Reliability. Applications to Medicine, Finance, and Quality Control. Ser. Statistics for Industry and Technology /Ed. V. Rykov, N. Balakrishnan, M. Nikulin. Birkhäuser, Boston, 2011. P. 433–453.

23. **Lemeshko B. Yu., Lemeshko S. B.** Models of Statistic Distributions of Nonparametric Goodness-of-Fit Tests in Composite Hypotheses Testing for Double Exponential Law Cases // Communications in Statistics – Theory and Methods. 2011. V. 40. N. 16. P. 2879–2892.

24. **Lemeshko B. Yu., Lemeshko S. B.** Construction of Statistic Distribution Models for Nonparametric Goodness-of-Fit Tests in Testing Composite Hypotheses: The Computer Approach // Quality Technology & Quantitative Management. 2011. V. 8. N. 4. P. 359–373.

25. **Stephens M.A.** EDF Statistics for Goodness of Fit and Some Comparisons // J. Am. Statistical Association. 1974. V. 69, N. 347. P. 730–737.

26. **Большев Л.Н., Смирнов Н.В.** Таблицы математической статистики. М.: Наука, 1983.

Дата принятия 17.07.2013 г.