РОССИЙСКАЯ АКАДЕМИЯ НАУК

СИБИРСКОЕ ОТДЕЛЕНИЕ

АВТОМЕТРИЯ

№ 2 2001

УДК 519.2

Б. Ю. Лемешко, С. Н. Постовалов

(Новосибирск)

ПРИМЕНЕНИЕ НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ СОГЛАСИЯ ПРИ ПРОВЕРКЕ СЛОЖНЫХ ГИПОТЕЗ*

Показано, что при проверке сложных гипотез на законы распределений статистик критериев согласия существенно влияет совокупность факторов: вид наблюдаемого закона; тип оцениваемого параметра и количество оцениваемых параметров; в некоторых ситуациях – конкретное значение параметра; метод оценивания параметров. Рекомендованы для применения построенные аппроксимации предельных распределений статистик непараметрических критериев согласия, расширяющие область корректного применения этих критериев при проверке сложных гипотез.

Введение. Одной из наиболее распространенных задач статистического анализа при обработке результатов экспериментальных наблюдений является проверка согласия полученного опытного распределения с теоретическим. Применяя критерии согласия, различают проверку простых и сложных гипотез. Простая проверяемая гипотеза имеет вид H_0: $F(x)=F(x,\theta)$, где $F(x,\theta)$– функция распределения вероятностей, с которой проверяется согласие наблюдаемой выборки, а θ – известное значение параметра (скалярного или векторного). Сложная проверяемая гипотеза имеет вид H_0: $F(x)\in \in\{F(x,\theta),\ \theta\in\Theta\}$. В этом случае оценка параметра распределения $\hat{\theta}$ вычисляется по той же самой выборке, по которой проверяется согласие.

В процессе проверки согласия по выборке вычисляется значение S^* статистики используемого критерия. Для того чтобы сделать вывод о принятии или отклонении гипотезы H_0, необходимо знать условное распределение $G(S\mid H_0)$ статистики S при справедливости гипотезы H_0. И если вероятность

$$P\{S>S^*\}=\int_{S^*}^{\infty} g(s\mid H_0)ds$$

достаточно большая, по крайней мере $P\{S>S^*\}>\alpha$, где $g(s\mid H_0)$– условная плотность, а α – задаваемый уровень значимости (вероятность ошибки первого рода – отклонить справедливую гипотезу H_0), то принято считать, что нет оснований для отклонения гипотезы H_0.

---

* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 00-01-00913).

К наиболее используемым критериям согласия относятся непараметрические критерии типа Колмогорова, типа ω^2 и Ω^2 Мизеса. В критерии Колмогорова в качестве расстояния между эмпирическим и теоретическим законами используется величина

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

где $F_n(x)$ – эмпирическая функция распределения, $F(x,\theta)$ – теоретическая функция распределения, n – объем выборки. При проверке гипотез обычно используется статистика вида [1]

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

где $D_n = \max(D_n^+, D_n^-), D_n^+ = \max\limits_{1 \le i \le n}\left\{\frac{i}{n} - F(x_i,\theta)\right\}, D_n^- = \max\limits_{1 \le i \le n}\left\{F(x_i,\theta) - \frac{i-1}{n}\right\}$, $x_1, x_2, \ldots, x_n$ – упорядоченные по возрастанию выборочные значения, $F(x,\theta)$ – функция закона распределения, согласие с которым проверяется. Распределение статистики S_K при проверке простой гипотезы в пределе подчиняется закону Колмогорова $K(S)$ [1].

В критериях типа ω^2 расстояние между гипотетическим и истинным распределениями рассматривается в квадратичной метрике

$$\int_{-\infty}^{\infty} \{E[F_n(x)] - F(x)\}^2 \psi(F(x)) dF(x),$$

где $E[\cdot]$ – оператор математического ожидания.

При выборе $\psi(t) \equiv 1$ в критериях типа ω^2 Мизеса пользуются статистикой (статистика Крамера – Мизеса – Смирнова) вида

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

которая при проверке простой гипотезы подчиняется распределению $a_1(S)$ [1].

При выборе $\psi(t) \equiv 1/t(1-t)$ в критериях типа Ω^2 Мизеса статистика (статистика Андерсона – Дарлинга) имеет вид

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

В пределе эта статистика подчиняется распределению $a_2(S)$ [1].

В случае простых гипотез предельные распределения статистик непараметрических критериев типа Колмогорова, ω^2 и Ω^2 Мизеса давно известны и не зависят от вида наблюдаемого закона распределения и его параметров. Говорят, что эти критерии «свободны от распределения». Это достоинство предопределяет широкое использование данных критериев в приложениях.

1. **Потеря «свободы от распределения» при проверке сложных гипотез.** При проверке сложных гипотез, когда по той же самой выборке оцениваются параметры наблюдаемого закона $F(x,\theta)$, непараметрические критерии согласия теряют свойство «свободы от распределения». Однако мощность непараметрических критериев при проверке сложных гипотез при тех же объемах выборок всегда существенно выше, чем при проверке простых. И если при проверке простых гипотез непараметрические критерии типа Колмогорова, ω^2 и Ω^2 Мизеса уступают по мощности критериям типа χ^2 при условии, что в последних используется асимптотически оптимальное группирование [2–5], то при проверке сложных гипотез непараметрические критерии оказываются более мощными. Для того чтобы воспользоваться их преимуществами, надо только знать распределение $G(S \mid H_0)$ при проверяемой сложной гипотезе.

Различия в предельных распределениях тех же самых статистик при проверке простых и сложных гипотез настолько существенны, что пренебрегать этим абсолютно недопустимо. Поэтому предостережения против неаккуратного применения критериев согласия при проверке сложных гипотез неоднократно высказывались на страницах печати [6–8].

Начало исследованиям предельных распределений статистик непараметрических критериев согласия при проверке сложных гипотез было положено работой [9]. В литературе изложен ряд подходов к использованию непараметрических критериев согласия в случае проверки сложных гипотез. При достаточно большом объеме выборки ее можно разбить на две части и по одной из них оценивать параметры, а по другой проверять согласие [10]. В некоторых частных случаях предельные распределения статистик исследованы аналитическими методами [11], процентные точки распределений построены методами статистического моделирования [12–15]. Для приближенного вычисления вероятностей «согласия» вида $P\{S > S^*\}$ (достигаемого уровня значимости) построены формулы, дающие достаточно хорошие приближения при малых значениях соответствующих вероятностей [16–20]. В работах [21–24] исследование распределений статистик непараметрических критериев согласия и построение моделей этих распределений осуществлялось с использованием методики компьютерного анализа статистических закономерностей.

Как выяснилось, при проверке сложных гипотез на условный закон распределения статистики $G(S \mid H_0)$ влияет целый ряд факторов, определяющих сложность гипотезы: вид наблюдаемого закона $F(x,\theta)$, соответствующего истинной гипотезе H_0; тип оцениваемого параметра и количество оцениваемых параметров; в некоторых ситуациях – конкретное значение параметра (например, в случае гамма-распределения); метод оценивания параметров.

Например, рис. 1 иллюстрирует зависимость $G(S \mid H_0)$ от вида наблюдаемого закона $F(x,\theta)$, соответствующего гипотезе H_0, для критерия типа Колмогорова. На рис. 2 показаны распределения статистики ω^2 Мизеса при проверке согласия с распределением Вейбулла с использованием различных методов оценивания: оценок максимального правдоподобия (ОМП) и *MD*-оценок, получаемых при минимизации значения статистики, используемой в критерии.

Распределения статистик критериев согласия существенно зависят от метода оценивания параметров. Строго говоря, каждому типу оценок при

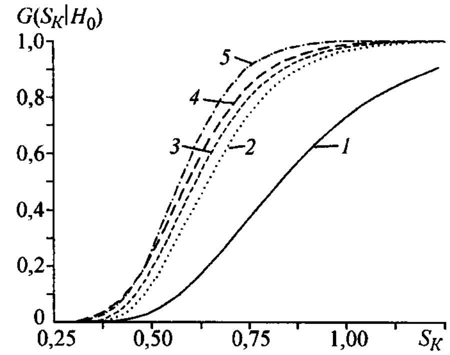

Рис. 1. Функции распределения $G(S_K \mid H_0)$ статистики S_K критерия типа Колмогорова: *1* – при проверке простой гипотезы; *2–5* – при вычислении ОМП двух параметров распределения Лапласа, нормального распределения, распределения Коши, логистического распределения соответственно

конкретной сложной проверяемой гипотезе соответствует свое предельное распределение $G(S \mid H_0)$ статистики. Применяя непараметрические критерии согласия, следует непременно учитывать используемый метод оценивания. В случае метода максимального правдоподобия распределения статистик $G(S \mid H_0)$ очень сильно зависят от закона, соответствующего гипотезе H_0. Разброс распределений $G(S \mid H_0)$ при использовании *MD*-оценок, минимизирующих статистику критерия, существенно меньше зависит от закона $F(x,\theta)$, соответствующего гипотезе H_0.

При использовании *MD*-оценок, минимизирующих статистику критерия, эмпирические распределения $G(S_n \mid H_0)$, соответствующие различным гипотезам H_0, имеют минимальный разброс, что позволяет говорить об определенной «свободе от распределения» для рассматриваемых критериев. Если опираться только на этот факт, то, казалось бы, только такие методы оценивания и следует применять при проверке сложных гипотез. Однако исследование мощности рассматриваемых критериев при различных методах оценивания показало, что наибольшую мощность данные критерии при близких альтернативах имеют в случае использования ОМП.

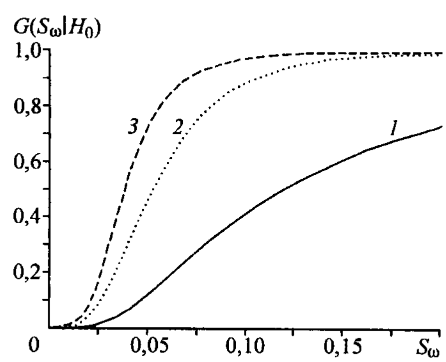

Рис. 2. Функции распределения $G(S_\omega \mid H_0)$ статистики критерия типа ω^2 Мизеса при проверке согласия с распределением Вейбулла: *1* – при проверке простой гипотезы; *2* – при вычислении ОМП двух параметров распределения; *3* – при вычислении *MD*-оценок двух параметров

При малых объемах выборки распределения $G(S_n | H_0)$ зависят от n. Однако существенная зависимость распределения статистик от n наблюдается только при небольших объемах выборки. Как показали исследования, при $n \geq 15-20$ распределения $G(S_n | H_0)$ достаточно близки к предельным $G(S | H_0)$ и зависимостью от n можно пренебречь.

2. **Построение приближений для предельных распределений статистик.** Построенные на настоящий момент таблицы процентных точек и предельные распределения статистик непараметрических критериев ограничены относительно узким кругом сложных гипотез.

Бесконечное множество случайных величин, с которым мы можем столкнуться на практике, нельзя описать ограниченным подмножеством моделей законов распределений, наиболее часто используемых для описания реальных наблюдений. Любой исследователь для конкретной наблюдаемой величины может предложить (построить) свою параметрическую модель закона, которая, с его точки зрения, наиболее адекватно описывает эту случайную величину. После оценки по данной выборке параметров модели возникает необходимость проверки гипотезы об адекватности выборочных наблюдений и построенного закона с использованием критериев согласия. Далее вопрос упирается в знание предельного распределения статистики, соответствующего данной сложной гипотезе.

Построение предельного распределения аналитическими методами – чрезвычайно сложная задача. Наиболее целесообразно воспользоваться методикой компьютерного анализа статистических закономерностей, хорошо зарекомендовавшей себя при моделировании распределений статистик критериев [21–24].

Для этого следует в соответствии с законом $F(x, \hat{\theta})$ смоделировать N выборок того же объема n, что и выборка, для которой необходимо проверить гипотезу $H_0: F(x) \in \{F(x,\theta),\ \theta \in \Theta\}$, а затем для каждой из N выборок вычислить оценки тех же параметров закона и значение статистики S соответствующего критерия согласия. В результате будет получена выборка значений статистики $S_1, S_2, ..., S_N$ с законом распределения $G(S_n | H_0)$ для проверяемой гипотезы H_0. По этой выборке при значительном числе N можно построить достаточно гладкую эмпирическую функцию распределения $G_N(S_n | H_0)$, которой можно непосредственно воспользоваться, чтобы сделать вывод, следует ли принимать гипотезу H_0. При необходимости можно по $G_N(S_n | H_0)$ построить приближенную аналитическую модель, аппроксимирующую $G_N(S_n | H_0)$, и тогда уже, опираясь на эту модель, принимать решение относительно проверяемой гипотезы.

Как показали исследования, хорошей аналитической моделью для $G_N(S_n | H_0)$ часто оказывается один из следующих законов: логарифмически нормальный, гамма-распределение, распределение *Su*-Джонсона, распределение *Sl*-Джонсона [23, 24]. В крайнем случае, всегда можно, опираясь на ограниченное множество законов распределения, построить модель в виде смеси законов.

Реализация такой процедуры компьютерного анализа распределений статистики в настоящий момент не содержит ни принципиальных, ни практических трудностей. Уровень вычислительной техники позволяет быстро получить результаты моделирования, а реализация алгоритма под силу инженеру, владеющему навыками программирования. В данной работе постро-

Таблица 1

Перечень распределений, соответствующих проверяемой гипотезе H_0

Распределение случайной величины	Функция плотности
Экспоненциальное	$\frac{1}{\theta_0} e^{-x/\theta_0}$
Максвелла	$\frac{2x^2}{\theta_0^3\sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
Лапласа	$\frac{1}{2\theta_0} e^{-\lvert x-\theta_1\rvert/\theta_0}$
Нормальное	$\frac{1}{\theta_0\sqrt{2\pi}} e^{-(x-\theta_1)^2/2\theta_0^2}$
Логарифмически нормальное	$\frac{1}{x\theta_0\sqrt{2\pi}} e^{-(\ln x-\theta_1)^2/2\theta_0^2}$
Коши	$\frac{\theta_0}{\pi[\theta_0^2+(x-\theta_1)^2]}$
Логистическое	$\frac{\pi}{\theta_0\sqrt{3}} \exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\} \Big/ \left[1+\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2$
Наибольшего значения	$\frac{1}{\theta_0} \exp\left\{-\frac{x-\theta_1}{\theta_0} - \exp\left(-\frac{x-\theta_1}{\theta_0}\right)\right\}$
Наименьшего значения	$\frac{1}{\theta_0} \exp\left\{\frac{x-\theta_1}{\theta_0} - \exp\left(\frac{x-\theta_1}{\theta_0}\right)\right\}$
Вейбулла	$\frac{\theta_0 x^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x}{\theta_1}\right)^{\theta_0}\right\}$
Гамма-распределение	$\frac{1}{\theta_1^{\theta_0}\Gamma(\theta_0)} (x-\theta_2)^{\theta_0-1} e^{-(x-\theta_2)/\theta_1}$

ены модели, аппроксимирующие предельные распределения статистик для ряда сложных гипотез при использовании ОМП и *MD*-оценок.

В табл. 1 приведен список распределений, относительно которых могут проверяться сложные гипотезы о согласии с использованием построенных приближений предельных законов статистик. Модели распределений статистик, построенные в результате применения методики компьютерного анализа статистических закономерностей, представлены в табл. 2–7. Перечень таблиц с моделями предельных распределений статистик, в том числе таблицы процентных точек, для более широкого класса проверяемых сложных гипотез представлен на WEB-сайте [25].

Таблица 2

Аппроксимация предельных распределений статистики Колмогорова при использовании метода максимального правдоподобия

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	двух параметров
Экспоненциальное	$\ln N(-0{,}3422,\ 0{,}2545)$	–	–
Полунормальное	$\gamma(4{,}1332,\ 0{,}1076,\ 0{,}3205)$	–	–
Рэлея	$\ln N(-0{,}3388,\ 0{,}2621)$	–	–
Максвелла	$\ln N(-0{,}3461,\ 0{,}2579)$	–	–
Лапласа	$\gamma(3{,}7580,\ 0{,}1365,\ 0{,}3163)$	$\gamma(4{,}6474,\ 0{,}0870,\ 0{,}3091)$ $\ln N(-0{,}3690,\ 0{,}2499)$	$\gamma(4{,}4525,\ 0{,}0761,\ 0{,}3252)$ $\ln N(-0{,}4358,\ 0{,}2276)$
Нормальное	$\gamma(3{,}7460,\ 0{,}1385,\ 0{,}3142)$	$\ln N(-0{,}4172,\ 0{,}2272)$	$\gamma(4{,}9014,\ 0{,}0691,\ 0{,}2951)$ $\ln N(-0{,}4825,\ 0{,}2296)$
Логарифмически нормальное	$\gamma(3{,}0622,\ 0{,}1577,\ 0{,}3547)$	$Su(-2{,}0328,\ 2{,}3642,\ 0{,}2622,\ 0{,}4072)$	$Su(-1{,}8093,\ 1{,}9041,\ 0{,}1861,\ 0{,}4174)$
Коши	$Su(-3{,}3278,\ 2{,}2529,\ 0{,}2185,\ 0{,}2858)$	$\gamma(4{,}8247,\ 0{,}0874,\ 0{,}2935)$	$\ln N(-0{,}5302,\ 0{,}2427)$
Логистическое	$\gamma(3{,}2167,\ 0{,}1476,\ 0{,}3538)$	$Su(-2{,}8534,\ 3{,}0657,\ 0{,}2872,\ 0{,}3199)$	$\ln N(-0{,}5611,\ 0{,}2082)$
Наибольшего значения	$\gamma(3{,}3841,\ 0{,}1439,\ 0{,}3509)$	$\gamma(4{,}1008,\ 0{,}0997,\ 0{,}3269)$	$\gamma(4{,}9738,\ 0{,}0660,\ 0{,}3049)$
Наименьшего значения	$\gamma(3{,}3841,\ 0{,}1439,\ 0{,}3509)$	$\gamma(4{,}1008,\ 0{,}0997,\ 0{,}3269)$	$\gamma(4{,}9738,\ 0{,}0660,\ 0{,}3049)$
Вейбулла	$\gamma(3{,}3841,\ 0{,}1439,\ 0{,}3509)$**	$\gamma(4{,}1008,\ 0{,}0997,\ 0{,}3269)$*	$\gamma(4{,}9738,\ 0{,}0660,\ 0{,}3049)$

Примечание. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распределения Вейбулла.

Таблица 3

Аппроксимация предельных распределений минимума статистики Колмогорова при использовании *MD*-оценок, минимизирующих статистику S_K

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	двух параметров
Экспоненциальное	γ(4,4983, 0,0621, 0,2891)	–	–
Полунормальное	γ(4,2884, 0,0705, 0,3072)	–	–
Рэлея	γ(4,8579, 0,0639, 0,2900)	–	–
Максвелла	γ(5,3106, 0,0581, 0,2865)	–	–
Лапласа	γ(3,0431, 0,1355, 0,3182)	γ(5,0103, 0,0602, 0,2968) ln*N*(–0,5358, 0,2122)	*Su*(–2,1079, 2,4629, 0,1661, 0,3340) ln*N*(–0,6970, 0,1952)
Нормальное	γ(3,2458, 0,1343, 0,3072)	ln*N*(–0,5469, 0,2152)	ln*N*(–0,7236, 0,1837)
Логарифмически нормальное	γ(3,2458, 0,1343, 0,3072)	ln*N*(–0,5469, 0,2152)	ln*N*(–0,7236, 0,1837)
Коши	γ(3,4398, 0,1255, 0,3022)	ln*N*(–0,5182, 0,2268)	*Su*(–1,6929, 2,5234, 0,1892, 0,3607) ln*N*(–0,6946, 0,1938)
Логистическое	*Su*(–2,6522, 1,8288, 0,1738, 0,3384) γ(3,6342, 0,1284, 0,2772)	*Su*(–3,8497, 3,2770, 0,2136, 0,2607) ln*N*(–0,5511, 0,2045)	ln*N*(–0,7389, 0,1771) *Su*(–2,5093, 3,1277, 0,1932, 0,3041)
Наибольшего значения	γ(3,5424, 0,1203, 0,2975)	*Su*(–1,9028, 2,3972, 0,2227, 0,389)	*Su*(–1,3144, 2,2480, 0,1616, 0,3858) ln*N*(–0,7174, 0,1841)
Наименьшего значения	γ(3,5424, 0,1203, 0,2975)	*Su*(–1,9028, 2,3972, 0,2227, 0,389)	*Su*(–1,3144, 2,2480, 0,1616, 0,3858) ln*N*(–0,7174, 0,1841)
Вейбулла	γ(3,5424, 0,1203, 0,2975)**	*Su*(–1,9028, 2,3972, 0,2227, 0,389)*	*Su*(–1,3144, 2,2480, 0,1616, 0,3858) ln*N*(–0,7174, 0,1841)

Примечание. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распределения Вейбулла.

Таблица 4

Аппроксимация предельных распределений статистики ω^2 Мизеса при использовании метода максимального правдоподобия

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	двух параметров
Экспоненциальное	$Su(-1{,}8734,\ 1{,}2118,\ 0{,}0223,\ 0{,}0240)$	–	–
Полунормальное	$Sl(0{,}9735,\ 1{,}1966,\ 0{,}1531,\ 0{,}0116)$	–	–
Рэлея	$Su(-1{,}5302,\ 1{,}0371,\ 0{,}0202,\ 0{,}0299)$	–	–
Максвелла	$Su(-2{,}0089,\ 1{,}2557,\ 0{,}0213,\ 0{,}0213)$	–	–
Лапласа	$Sl(0{,}9719,\ 0{,}9805,\ 0{,}2347,\ 0{,}0139)$	$Su(-2{,}0821,\ 1{,}2979,\ 0{,}0196,\ 0{,}0200)$	$Su(-1{,}6085,\ 1{,}2139,\ 0{,}0171,\ 0{,}0247)$
Нормальное	$Su(-2{,}2550,\ 0{,}9569,\ 0{,}0152,\ 0{,}0212)$	$\ln N(-2{,}7536,\ 0{,}5610)$	$\ln N(-2{,}9794,\ 0{,}5330)$
Логарифмически нормальное	$Sl(1{,}0669,\ 1{,}0010,\ 0{,}2537,\ 0{,}0144)$	$\ln N(-2{,}7271,\ 0{,}6092)$	$Su(-1{,}6292,\ 1{,}1541,\ 0{,}0144,\ 0{,}0234)$
Коши	$Sl(1{,}0086,\ 1{,}0539,\ 0{,}2282,\ 0{,}0064)$	$Sl(1{,}1230,\ 1{,}2964,\ 0{,}1383,\ 0{,}0105)$	$Sl(1{,}2420,\ 1{,}2833,\ 0{,}1135,\ 0{,}0064)$
Логистическое	$Sl(0{,}9982,\ 1{,}0287,\ 0{,}2303,\ 0{,}0126)$	$Sl(1{,}3982,\ 1{,}3804,\ 0{,}1205,\ 0{,}0102)$	$\ln N(-3{,}1416,\ 0{,}4989)$
Наибольшего значения	$Sl(1{,}0056,\ 1{,}0452,\ 0{,}2296,\ 0{,}0137)$	$\ln N(-2{,}5818,\ 0{,}6410)$	$\ln N(-2{,}9541,\ 0{,}5379)$
Наименьшего значения	$Sl(1{,}0056,\ 1{,}0452,\ 0{,}2296,\ 0{,}0137)$	$\ln N(-2{,}5818,\ 0{,}6410)$	$\ln N(-2{,}9541,\ 0{,}5379)$
Вейбулла	$Sl(1{,}0056,\ 1{,}0452,\ 0{,}2296,\ 0{,}0137)$**	$\ln N(-2{,}5818,\ 0{,}6410)$*	$\ln N(-2{,}9541,\ 0{,}5379)$

Примечание. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распределения Вейбулла.

Таблица 5

Аппроксимация предельных распределений минимума статистики ω^2 Мизеса при использовании *MD*-оценок, минимизирующих статистику S_ω

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	двух параметров
Экспоненциальное	*Su*(–1,9324, 1,1610, 0,0134, 0,0203)	–	–
Полунормальное	*Su*(–1,5024, 1,0991, 0,0173, 0,0256)	–	–
Рэлея	*Su*(–1,4705, 1,1006, 0,0164, 0,0259)	–	–
Максвелла	*Su*(–1,7706, 1,2978, 0,0188, 0,0220)	–	–
Лапласа	*Sl*(1,0117, 0,9485, 0,2162, 0,0137)	ln*N*(–2,8601, 0,5471)	ln*N*(–3,2853, 0,4666)
Нормальное	*Sl*(1,0477, 0,9883, 0,2356, 0,0112)	ln*N*(–2,8649, 0,5668)	ln*N*(–3,2715, 0,4645)
Логарифмически нормальное	*Sl*(1,0477, 0,9883, 0,2356, 0,0112)	ln*N*(–2,8649, 0,5668)	ln*N*(–3,2715, 0,4645)
Коши	*Sl*(1,2759, 1,0437, 0,2825, 0,0089)	ln*N*(–2,8577, 0,5739)	ln*N*(–3,2603, 0,4874)
Логистическое	*Sl*(1,0898, 1,0225, 0,2399, 0,0096)	ln*N*(–2,8831, 0,5367)	ln*N*(–3,2915, 0,4592)
Наибольшего значения	*Sl*(1,0771, 1,0388, 0,2065, 0,0109)	*Su*(–1,5348, 1,1226, 0,0166, 0,0252)	*Su*(–1,5326, 1,4446, 0,0147, 0,0188) ln*N*(–3,2627, 0,4680)
Наименьшего значения	*Sl*(1,0771, 1,0388, 0,2065, 0,0109)	*Su*(–1,5348, 1,1226, 0,0166, 0,0252)	*Su*(–1,5326, 1,4446, 0,0147, 0,0188) ln*N*(–3,2627, 0,4680)
Вейбулла	*Sl*(1,0771, 1,0388, 0,2065, 0,0109)**	*Su*(–1,5348, 1,1226, 0,0166, 0,0252)*	*Su*(–1,5326, 1,4446, 0,0147, 0,0188) ln*N*(–3,2627, 0,4680)

Примечание. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распределения Вейбулла.

Таблица 6

Аппроксимация предельных распределений статистики Ω^2 Мизеса при использовании метода максимального правдоподобия

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	двух параметров
Экспоненциальное	$Su(-2{,}8653,\ 1{,}4220,\ 0{,}1050,\ 0{,}1128)$	–	–
Полунормальное	$Su(-2{,}5603,\ 1{,}3116,\ 0{,}1147,\ 0{,}1330)$	–	–
Рэлея	$Su(-2{,}5610,\ 1{,}4003,\ 0{,}1174,\ 0{,}1337)$	–	–
Максвелла	$Su(-2{,}6064,\ 1{,}4426,\ 0{,}1190,\ 0{,}1285)$	–	–
Лапласа	$Sl(0{,}3148,\ 1{,}0999,\ 0{,}6901,\ 0{,}1093)$	$Su(-2{,}5528,\ 1{,}4006,\ 0{,}1216,\ 0{,}1358)$	$Su(-2{,}8942,\ 1{,}4897,\ 0{,}0846,\ 0{,}1131)$
Нормальное	$Su(-2{,}3507,\ 1{,}0531,\ 0{,}1012,\ 0{,}1595)$	$Su(-3{,}1202,\ 1{,}5233,\ 0{,}0874,\ 0{,}1087)$	$Su(-2{,}7057,\ 1{,}7154,\ 0{,}1043,\ 0{,}0925)$
Логарифмически нормальное	$Su(-2{,}4168,\ 1{,}1296,\ 0{,}1151,\ 0{,}1560)$	$\ln N(-0{,}8052,\ 0{,}5123)$	$Su(-2{,}3966,\ 1{,}5967,\ 0{,}1012,\ 0{,}1179)$
Коши	$Su(-2{,}4935,\ 1{,}0789,\ 0{,}0923,\ 0{,}1458)$	$Su(-2{,}8420,\ 1{,}3528,\ 0{,}1010,\ 0{,}1221)$	$Su(-2{,}3195,\ 1{,}1812,\ 0{,}0769,\ 0{,}1217)$
Логистическое	$Sl(0{,}3065,\ 1{,}1628,\ 0{,}7002,\ 0{,}0930)$	$Su(-3{,}5408,\ 1{,}6041,\ 0{,}0773,\ 0{,}0829)$	$\ln N(-1{,}1452,\ 0{,}4426)$
Наибольшего значения	$Su(-2{,}5427,\ 1{,}1057,\ 0{,}0960,\ 0{,}1569)$	$Su(-2{,}5550,\ 1{,}3714,\ 0{,}1152,\ 0{,}1289)$	$Su(-2{,}4622,\ 1{,}6473,\ 0{,}1075,\ 0{,}1149)$
Наименьшего значения	$Su(-2{,}5427,\ 1{,}1057,\ 0{,}0960,\ 0{,}1569)$	$Su(-2{,}5550,\ 1{,}3714,\ 0{,}1152,\ 0{,}1289)$	$Su(-2{,}4622,\ 1{,}6473,\ 0{,}1075,\ 0{,}1149)$
Вейбулла	$Su(-2{,}5427,\ 1{,}1057,\ 0{,}0960,\ 0{,}1569)$**	$Su(-2{,}5550,\ 1{,}3714,\ 0{,}1152,\ 0{,}1289)$*	$Su(-2{,}4622,\ 1{,}6473,\ 0{,}1075,\ 0{,}1149)$

П р и м е ч а н и е. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распределения Вейбулла.

7*

Таблица 7

Аппроксимация предельных распределений минимума статистики Ω^2 Мизеса при использовании *MD*-оценокщих статистику S_Ω

Распределение случайной величины	Оценивание		
	масштабного параметра	параметра сдвига	араметров
Экспоненциальное	*Su*(–2,6741, 1,4068, 0,0958, 0,1230)	–	–
Полунормальное	*Su*(–2,6752, 1,3763, 0,0952, 0,1280)	–	–
Рэлея	*Su*(–2,2734, 1,3473, 0,1101, 0,1496)	–	–
Максвелла	*Su*(–2,2759, 1,3988, 0,1171, 0,1514)	–	–
Лапласа	*Su*(–2,3884, 1,0811, 0,0948, 0,1548)	*Su*(–2,7267, 1,4972, 0,1044, 0,1239)	04, 0,0902, 0,1123)
Нормальное	*Su*(–2,4180, 1,0702, 0,0957, 0,1464)	*Su*(–2,7639, 1,5393, 0,1102, 0,1115)	05, 0,0979, 0,1043) 651, 0,4271)
Логарифмически нормальное	*Su*(–2,4180, 1,0702, 0,0957, 0,1464)	*Su*(–2,7639, 1,5393, 0,1102, 0,1115)	05, 0,0979, 0,1043) 651, 0,4271)
Коши	*Su*(–2,5043, 1,1355, 0,1035, 0,1384)	*Su*(–2,7029, 1,5179, 0,1188, 0,1100)	64, 0,0929, 0,1301) 043, 0,4692)
Логистическое	*Sl*(0,3223, 1,1159, 0,6836, 0,0953) *Su*(–2,3007, 1,0135, 0,0906, 0,1593)	*Su*(–2,6212, 1,4318, 0,0932, 0,1370)	51, 0,0800, 0,0898)
Наибольшего значения	*Su*(–2,4454, 1,1083, 0,0968, 0,1459)	*Su*(–2,6557, 1,4282, 0,1024, 0,1254)	46, 0,0941, 0,1279)
Наименьшего значения	*Su*(–2,4454, 1,1083, 0,0968, 0,1459)	*Su*(–2,6557, 1,4282, 0,1024, 0,1254)	46, 0,0941, 0,1279)
Вейбулла	*Su*(–2,4454, 1,1083, 0,0968, 0,1459)**	*Su*(–2,6557, 1,4282, 0,1024, 0,1254)*	46, 0,0941, 0,1279)

Примечание. ** – оценивался параметр формы распределения Вейбулла, * – параметр масштаба распреде

В табл. 2–7, содержащих рекомендуемые для использования при проверке сложных гипотез распределения $G(S \mid H_0)$, через $\ln N(\theta_1, \theta_2)$ обозначено логарифмически нормальное распределение с функцией плотности

$$f(x) = \frac{1}{x\theta_0\sqrt{2\pi}} e^{-(\ln x - \theta_1)^2/2\theta_0^2},$$

через $\gamma(\theta_0, \theta_1, \theta_2)$ – гамма-распределение с функцией плотности

$$f(x) = \frac{1}{\theta_1^{\theta_0}\Gamma(\theta_0)}(x - \theta_2)^{\theta_0 - 1} e^{-(x-\theta_2)/\theta_1},$$

через $Sl(\theta_0, \theta_1, \theta_2, \theta_3)$ – распределение Sl-Джонсона с функцией плотности

$$f(x) = \frac{\theta_1}{(x - \theta_3)} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\frac{x - \theta_3}{\theta_2}\right]^2\right\},$$

через $Su(\theta_0, \theta_1, \theta_2, \theta_3)$ – распределение Su-Джонсона с функцией плотности

$$f(x) = \frac{\theta_1}{\sqrt{2\pi}\sqrt{(x-\theta_3)^2 + \theta_2^2}} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\left\{\frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2 + 1}\right\}\right]^2\right\}.$$

В качестве примера покажем, насколько сильно меняется вероятность $P\{S > S^*\}$ при одном и том же значении статистики в случае простой и сложной гипотез. Продемонстрируем это на критерии типа ω^2 Мизеса.

П р и м е р. Пусть проверяется гипотеза о согласии с распределением Вейбулла и вычисленное значение статистики $S_\omega^* = 0{,}14$. Тогда в случае простой гипотезы на основании распределения $a_1(S)$ [1] находим, что $P\{S_\omega > 0{,}14\} = 0{,}4215$. Если по выборке были вычислены ОМП двух параметров распределения, то хорошей аппроксимацией предельного распределения (см. табл. 4) является логарифмически нормальное $\ln N(-2{,}9541, 0{,}5379)$, а соответствующая вероятность $P\{S_\omega > 0{,}14\} = 0{,}0331$. При использовании MD-оценок в аналогичной ситуации наиболее подходящей моделью (см. табл. 5) является распределение Su-Джонсона $Su(-1{,}5326,\ 1{,}4446,\ 0{,}0147,\ 0{,}0188)$, в соответствии с которым $P\{S_\omega > 0{,}14\} = 0{,}0058$.

Заключение. При проверке сложных гипотез и выборе (или построении) распределений статистик $G(S \mid H_0)$ критериев согласия необходимо учитывать все факторы, влияющие на закон распределения статистики: вид наблюдаемого закона; тип оцениваемого параметра и количество оцениваемых параметров; в некоторых ситуациях – конкретное значение параметра; метод оценивания параметров.

Построенные аппроксимации предельных распределений статистик непараметрических критериев согласия расширяют область корректного применения этих критериев и могут быть рекомендованы широкому кругу исследователей. Апробированная методика моделирования распределений статистик может быть рекомендована для построения статистических закономерностей в ситуации, когда аналитическими методами решить задачу не удается.

СПИСОК ЛИТЕРАТУРЫ

1. **Большев Л. Н., Смирнов Н. В.** Таблицы математической статистики. М.: Наука, 1983.
2. **Денисов В. И., Лемешко Б. Ю., Цой Е. Б.** Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов. Новосибирск: Изд-во НГТУ, 1993.
3. **Лемешко Б. Ю.** Асимптотически оптимальное группирование наблюдений – это обеспечение максимальной мощности критериев // Надежность и контроль качества. 1997. № 8. С. 3.
4. **Лемешко Б. Ю.** Асимптотически оптимальное группирование наблюдений в критериях согласия // Завод. лаб. 1998. **64**, № 1. С. 56.
5. **Денисов В. И., Лемешко Б. Ю., Постовалов С. Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Ч. I. Критерии типа χ^2. Новосибирск: Изд-во НГТУ, 1998.
6. **Орлов А. И.** Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Завод. лаб. 1985. **51**, № 1. С. 60.
7. **Бондарев Б. В.** О проверке сложных статистических гипотез // Завод. лаб. 1986. **52**, № 10. С. 62.
8. **Кулинская Е. В., Саввушкина Н. Е.** О некоторых ошибках в реализации и применении непараметрических методов в пакете для IBM PC // Завод. лаб. 1990. **56**, № 5. С. 96.
9. **Kac M., Kiefer J., Wolfowitz J.** On tests of normality and other tests of goodness of fit based on distance methods // Ann. Math. Statist. 1955. **26**. P. 189.
10. **Durbin J.** Kolmogorov–Smirnov test when parameters are estimated // Lect. Notes Math. 1976. **566**. P. 33.
11. **Мартынов Г. В.** Критерии омега-квадрат. М.: Наука, 1978.
12. **Pearson E. S., Hartley H. O.** Biometrica Tables for Statistics. Cambridge: University Press, 1972. V. 2.
13. **Stephens M. A.** Use of Kolmogorov–Smirnov, Cramer – von Mises and related statistics – without extensive table // Journ. Roy. Statist. Soc. 1970. **B32**. P. 115.
14. **Stephens M. A.** EDF statistics for goodness of fit and some comparisons // Journ. Amer. Statist. Assoc. 1974. **69**. P. 730.
15. **Chandra M., Singpurwalla N. D., Stephens M. A.** Statistics for test of fit for the extrem-value and weibull distribution // Journ. Amer. Statist. Assoc. 1981. **76**. P. 375.
16. **Тюрин Ю. Н.** О предельном распределении статистик Колмогорова – Смирнова для сложной гипотезы // Изв. АН СССР. Сер. Мат. 1984. **48**, № 6. С. 1314.
17. **Тюрин Ю. Н., Саввушкина Н. Е.** Критерии согласия для распределения Вейбулла – Гнеденко // Изв. АН СССР. Сер. Техн. кибернетика. 1984. № 3. С. 109.
18. **Тюрин Ю. Н.** Исследования по непараметрической статистике (непараметрические методы и линейная модель): Автореф. дис. … д-ра физ.-мат. наук /МГУ. М., 1985.
19. **Саввушкина Н. Е.** Критерий Колмогорова – Смирнова для логистического и гамма-распределения // Сб. тр. ВНИИ системных исследований. 1990. № 8.
20. **Тюрин Ю. Н., Макаров А. А.** Анализ данных на компьютере. М.: ИНФРА-М, Финансы и статистика, 1995.

21. **Лемешко Б. Ю., Постовалов С. Н.** Прикладные аспекты использования критериев согласия в случае проверки сложных гипотез // Надежность и контроль качества. 1997. № 11. С. 3.

22. **Лемешко Б. Ю., Постовалов С. Н.** Исследование допредельных распределений статистик критериев согласия при проверке сложных гипотез // Тр. IV Междунар. конф. «Актуальные проблемы электронного приборостроения». Новосибирск, 1998. Т. 3. С. 12.

23. **Лемешко Б. Ю., Постовалов С. Н.** О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов // Завод. лаб. 1998. **64**, № 3. С. 61.

24. **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.

25. **http:** www.ami.nstu.ru/~headrd/

Новосибирский государственный
технический университет,
E-mail: headrd@fpm.ami.nstu.ru

Поступила в редакцию
22 февраля 2000 г.