

УДК 519.2

О РАСПРЕДЕЛЕНИЯХ СТАТИСТИК НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ СОГЛАСИЯ ПРИ ОЦЕНИВАНИИ ПО ВЫБОРКАМ ПАРАМЕТРОВ НАБЛЮДАЕМЫХ ЗАКОНОВ

Б.Ю. Лемешко¹, С.Н. Постовалов

Приведены результаты моделирования распределений статистик непараметрических критериев согласия при справедливости проверяемой гипотезы в зависимости от числа и вида оцениваемых параметров наблюдаемых законов. В этом случае предельные распределения статистик непараметрических критериев согласия сильно отличаются от распределений классических статистик.

1. Введение и постановка задачи

Наиболее часто в практике статистического анализа с необходимостью использования критериев согласия приходится сталкиваться после оценивания по этой же выборке параметров предполагаемого закона распределения. К сожалению, в этом случае предельные распределения статистик таких непараметрических критериев, как Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, при справедливости нулевой гипотезы вида $H_0: f(x, \theta_0) = f(x, \hat{\theta})$, где $f(\cdot)$ - плотность распределения наблюдаемого закона, θ_0 - истинное значение параметра, $\hat{\theta}$ - оценка параметра, вычисленная по выборке, отличаются от предельных распределений классических статистик (когда по выборке не оцениваются параметры). При оценивании по выборке параметров рассматриваемые критерии теряют свойство "свободы от распределения", и предельные распределения статистик на самом деле зависят как от числа оцененных параметров, так и от вида исследуемого закона распределения $f(x, \theta)$. Широкою известность этот факт приобрел, начиная с работы [1]. Распространенная ошибка, связанная с пренебрежением существующей проблемы, чаще всего приводит к необоснованному принятию нулевой гипотезы, что подчеркивается в работах [2,3], из-за сильно завышенных значений вероятностей "согласия" вида $P\{S > S^*\}$, где S^* - значение статистики, вычисленное по выборке. И как отмечено в [4], это не всегда учитывается в популярных пакетах статистического анализа.

Если объем выборки достаточно велик, можно, опираясь на результаты, полученные в [5], оценивать параметры распределения по одной половине выборки, а проверять согласие по другой половине. В такой ситуации применение предельных распределений рассматриваемых критериев вполне обосновано. Но в этом случае как при оценивании, так и при проверке согласия мы используем только половину имеющейся информации, что, естественно, сказывается

¹ Новосибирский государственный технический университет, Россия

на качестве статистических выводов. К тому же, объемы выборок, зачастую имеющиеся в распоряжении исследователя, не настолько велики, чтобы можно было смириться с потерей части информации при оценивании параметров.

Таким образом, с одной стороны некорректность использования классических непараметрических критериев в задачах проверки сложных гипотез из-за слишком больших вероятностей ошибок, с другой стороны свойства асимптотической оптимальности критериев типа Колмогорова-Смирнова и ω^2 (по Ходжесу-Леману) [6] предопределяют необходимость поиска предельных законов распределения непараметрических статистик в зависимости от того, сколько параметров оценивалось по выборке и с каким законом проверяется согласие.

Для случая нормального закона предельные распределения статистики критерия ω^2 Мизеса при оценивании одного из двух или обоих параметров подробно исследованы в [7], где приведены их таблицы.

Но очевидно, что теоретически найти решение задачи определения предельных распределений непараметрических статистик для множества законов, используемых для описания реальных величин, очень сложно. Именно поэтому большинство существующих таблиц и результатов, связанных с предельными распределениями, получены методом статистического моделирования [8-16]. Таким образом получены, например, таблицы процентных точек для модифицированных статистик критериев Колмогорова (типа Колмогорова) при проверке нормальности и оценивании одного или двух параметров нормального распределения и процентные точки при проверке экспоненциальности и оценивании его масштабного параметра, представленные в [12].

В работах [17-19] разработан аналитический метод расчета процентных точек для статистик Колмогорова-Смирнова для регулярных семейств распределений и больших выборок, получены формулы для приближенного вычисления вероятностей вида $P\{S > S^*\}$. С помощью этих приближенных формул, учитывающих отличие предельных распределений непараметрических статистик при оценивании параметров законов от классических, вычисляют вероятности вида $P\{S > S^*\}$ в пакете STADIA [20].

Нам наиболее реальный выход видится в моделировании предельных законов распределения статистик непараметрических критериев и в последующей идентификации полученных эмпирических законов. Под идентификацией понимается выбор такого теоретического закона распределения, который наиболее хорошо согласуется с эмпирической функцией распределения. Результаты такого моделирования позволяют, с одной стороны, количественно оценить возможные ошибки в случае пренебрежения фактом оценивания по выборке параметров наблюдаемого закона, с другой стороны, дают возможность с достаточной для практики точностью использовать эти результаты при проверке гипотез.

Реализованный алгоритм численного моделирования выглядит следующим образом.

1. Моделируется (имитируется) N выборок заданного объема n в соответствии с заданным законом распределения $f(x, \theta)$.
2. По каждой выборке оцениваются параметры этого закона распределения (один параметр, или другой, или два одновременно).
3. Вычисляются значения соответствующих непараметрических статистик.
4. По накопленной выборке значений статистики объема N анализируется полученное эмпирическое распределение статистики и осуществляется его идентификация.

Конечной целью настоящего исследования явилось стремление по возможности найти такие законы распределения вероятностей, которые с практической точки зрения хорошо аппроксимируют предельные распределения статистик непараметрических критериев в тех случаях, когда по выборке оцениваются параметры.

При постановке задачи исследования планировалось осуществить моделирование для различных объемов N и n и, в случае устойчивости полученных результатов, использовать их в программном обеспечении статистического анализа [21], а также протабулировать полученные эмпирические законы и критические значения статистик, если не удастся подобрать подходящего теоретического закона распределения. Подобную процедуру планировалось осуществить для каждого закона распределения, включенного в программную систему статистического анализа [21].

В данной статье мы приводим некоторые результаты моделирования и анализа, которые, с нашей точки зрения, могут с успехом применяться при решении практических задач проверки гипотез о согласии с использованием непараметрических критериев после вычисления оценок параметров распределения по той же выборке.

Отметим, что результаты моделирования классических непараметрических статистик (без оценивания параметров) показали хорошее согласие получившихся эмпирических распределений с предельными теоретическими законами. Это позволяет судить о точности описания предельных законов полученными эмпирическими распределениями статистик для ситуации с оцениванием параметров, а также служит косвенным подтверждением возможности использования полученных распределений соответствующих статистик при решении практических задач статистического анализа.

Статистика Колмогорова (Колмогорова-Смирнова) определяется выражением [22]

$$S_k = \frac{(6nD_n + 1)^2}{18n},$$

статистика Смирнова [22]

$$S_m = \frac{(6nD_n^+ + 1)^2}{9n},$$

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

n - объем выборки, x_1, x_2, \dots, x_n - упорядоченные по возрастанию выборочные значения, $F(x)$ - функция распределения, согласие с которой проверяется.

Распределение величины $\sqrt{S_k / 2}$, если по выборке не оценивались параметры, в пределе подчиняется закону Колмогорова с функцией распределения $K(x)$ [22].

Гипотеза о согласии не отвергается, если

$$P\{S_k > S_k^*\} = 1 - K\left(\sqrt{\frac{S_k^*}{2}}\right) > \alpha,$$

где α - задаваемый уровень значимости.

В аналогичной ситуации статистика Смирнова S_m подчиняется в пределе распределению χ^2 с числом степеней свободы, равным 2. Гипотеза о согласии не отвергается, если

$$P\{S_m > S_m^*\} = \int_{S_m^*}^{\infty} \frac{1}{2} e^{-x/2} dx = e^{-S_m^*/2} > \alpha.$$

Статистики Мизеса имеют вид [22]

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

которую чаще называют статистикой Крамера-Мизеса-Смирнова, и

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

называемую еще статистикой Андерсона-Дарлингга. Для этих статистик также известны предельные распределения вероятностей [22]

$$\lim_{n \rightarrow \infty} P\{n\omega_n^2 < x\} = a1(x),$$

$$\lim_{n \rightarrow \infty} P\{n\Omega_n^2 < x\} = a2(x).$$

Гипотезы о согласии не отвергаются, если выполняются неравенства

$$P\{S_{\omega} > S_{\omega}^*\} = 1 - a1(S_{\omega}^*) > \alpha \quad \text{и} \quad P\{S_{\Omega} > S_{\Omega}^*\} = 1 - a2(S_{\Omega}^*) > \alpha.$$

В соответствии с изложенным алгоритмом моделировались выборки значений статистик $\sqrt{S_k / 2}$, S_m , S_{ω} , S_{Ω} .

По каждому закону распределения $f(x, \theta)$ моделировалась серия из $N = 2000$ выборок объемом $n = 200$. Псевдослучайная величина, принадлежащая закону с функцией распределения $F(x, \theta)$, имитировалась по методу обратных функций, при котором случайная величина X , подчиняющаяся закону с функцией распределения $F(\theta, x)$, получается в соответствии с соотношением $X = F^{-1}(\theta, Y)$, где $F^{-1}(\cdot)$ - функция, обратная к $F(\cdot)$, а Y - случайная величина, равномерно распределенная на интервале $[0, 1]$. В качестве датчика равномерно распределенных псевдослучайных чисел использовался стандартный датчик,

реализованный в C++. Оценки параметров находились по методу максимального правдоподобия по негруппированным данным.

В данной работе исследовались распределения непараметрических статистик, когда наблюдаемые случайные величины распределены в соответствии с законами, представленными в табл. 1.

Таблица 1

№ п/п	Распределение случайной величины	Функция плотности
1.	Экспоненциальное	$\theta_0 e^{-\theta_0 x}$
2.	Полунормальное	$\frac{2}{\theta_0 \sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
3.	Рэля	$\frac{x}{\theta_0^2} e^{-x^2/2\theta_0^2}$
4.	Максвелла	$\frac{2x^2}{\theta_0^3 \sqrt{2\pi}} e^{-\frac{x^2}{2\theta_0^2}}$
5.	Лапласа	$\frac{\theta_0}{2} e^{-\theta_0 x-\theta_1 }$
6.	Нормальное	$\frac{1}{\theta_0 \sqrt{2\pi}} e^{-\frac{(x-\theta_1)^2}{2\theta_0^2}}$
7.	Логнормальное	$\frac{1}{x\theta_0 \sqrt{2\pi}} e^{-(\ln x - \theta_1)^2/2\theta_0^2}$
8.	Коши	$\frac{\theta_0}{\pi[\theta_0^2 + (x - \theta_1)^2]}$
9.	Логистическое	$\frac{\pi}{\theta_0 \sqrt{3}} \exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0 \sqrt{3}}\right\} / \left[1 + \exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0 \sqrt{3}}\right\}\right]^2$
10.	Наибольшего значения	$\frac{1}{\theta_0} \exp\left\{-\frac{x-\theta_1}{\theta_0} - \exp\left(-\frac{x-\theta_1}{\theta_0}\right)\right\}$
11.	Наименьшего значения	$\frac{1}{\theta_0} \exp\left\{\frac{x-\theta_1}{\theta_0} - \exp\left(\frac{x-\theta_1}{\theta_0}\right)\right\}$
12.	Вейбулла-Гнеденко	$\frac{\theta_0 x^{\theta_0-1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x}{\theta_1}\right)^{\theta_0}\right\}$

2. Предельные распределения статистики Колмогорова

На рис.1 приведены результаты моделирования величины $\sqrt{S_k/2}$, используемой в критерии Колмогорова, при проверке гипотез о согласии с нормальным распределением при справедливости гипотезы H_0 . На этом и последующих рисунках представлены эмпирические функции распределения статистики, когда по выборке не оценивались параметры ("1"), по выборке оценивался только масштабный параметр ("2") (в данном случае θ_0), оценивался только параметр сдвига ("3") (в данном случае θ_1), оценивались одновременно оба параметра ("4"). Здесь же приведена функция распределения Колмогорова ("0"), которому подчиняется статистика $\sqrt{S_k/2}$, если по выборке не оцениваются параметры. Результаты проверки согласия эмпирического распределения "1" с распределением Колмогорова "0" достаточно хорошее. В то же время весьма наглядно отличие эмпирических функций распределения "2", "3", "4" от распределения Колмогорова "0". Это отличие позволяет судить о величине тех ошибок, которые мы допускаем, не учитывая факта оценивания параметров конкретного распределения при использовании критерия Колмогорова. Крестиками, лежащими практически на эмпирическом распределении "4", обозначены значения, соответствующие процентным точкам при проверке нормальности, взятые из таблицы М. Стефенса [12], а кружочками - значения, полученные по приближенным формулам Ю.Н. Тюрина [17,19] для расчетов вероятностей, соответствующих распределениям "4" и "2".

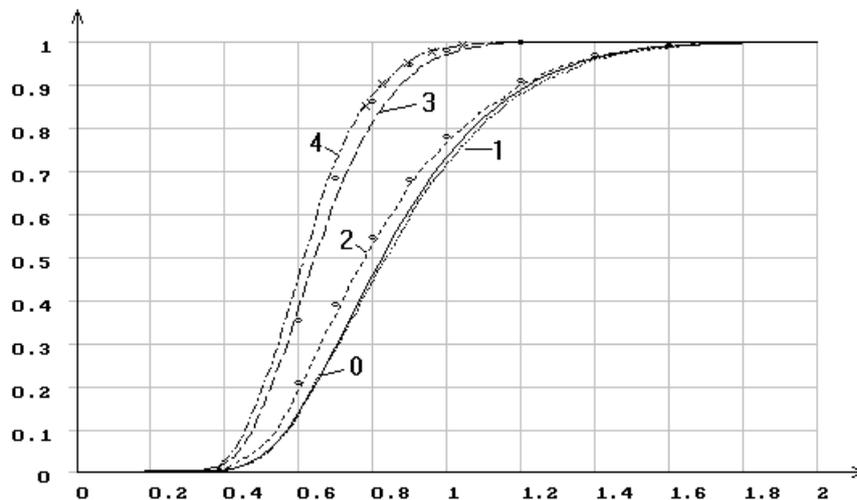


Рис.1. Эмпирические функции распределения статистики $\sqrt{S_k/2}$ Колмогорова при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения Колмогорова; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра.

Замечание: Процентные точки М.Стефенса соответствуют модифицированной статистике вида $D_n \left(\sqrt{n} - 0.001 + \frac{0.85}{\sqrt{n}} \right)$, а по формулам Ю.Н.Тюриня рассчитывают вероятности для статистики $\sqrt{n}D_n$. В то же время $\sqrt{S_k/2} = \sqrt{n}D_n + \frac{1}{6\sqrt{n}}$.

Аналогичная картина распределения статистики Колмогорова при справедливой гипотезе H_0 наблюдается для логистического распределения на рис. 2, для распределения Лапласа на рис. 3, для распределения Коши на рис. 4 и для экспоненциального распределения на рис. 5, где крестиками, обозначены значения, соответствующие процентным точкам при проверке экспоненциальности и неизвестном масштабном параметре, взятые из таблицы процентных точек [12].

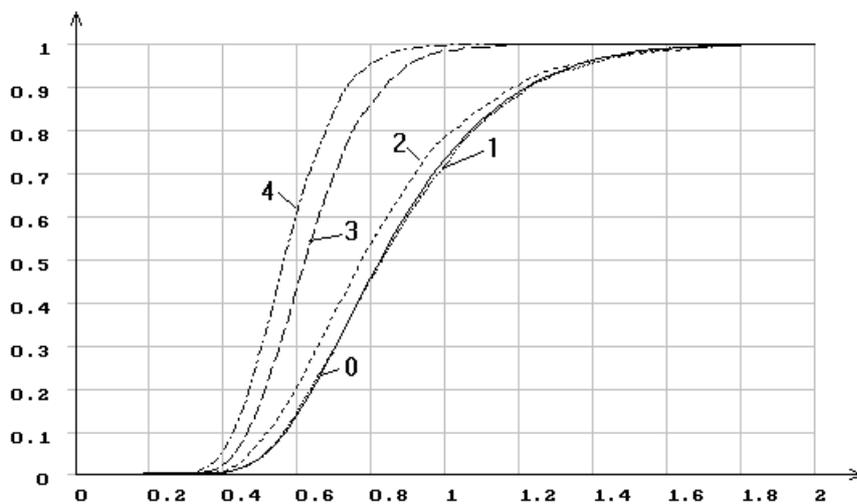


Рис.2. Эмпирические функции распределения статистики $\sqrt{S_k/2}$ Колмогорова при различном количестве оцениваемых параметров логистического распределения

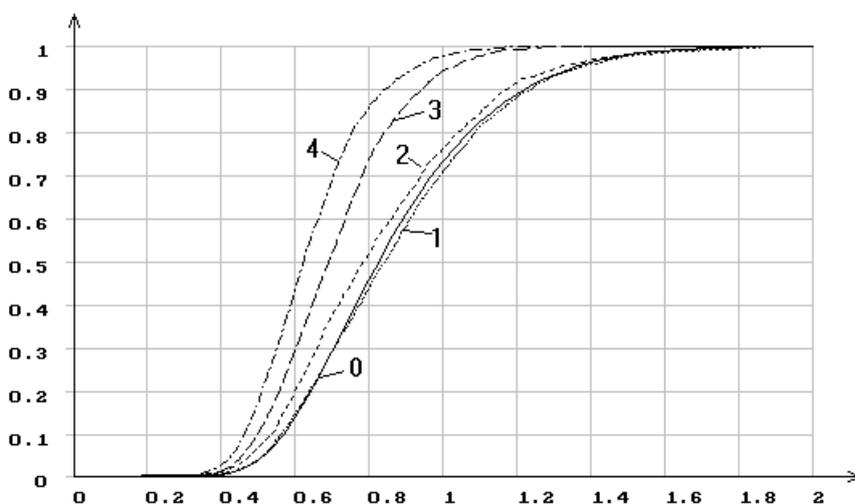


Рис.3. Эмпирические функции распределения статистики $\sqrt{S_k/2}$ Колмогорова при различном количестве оцениваемых параметров распределения Лапласа

Результаты моделирования однозначно указывают на то, что предельные распределения статистики $\sqrt{S_k}/2$ при условии оценивания параметров конкретного закона настолько сильно отличаются от распределения Колмогорова, что использование последнего никак не может быть оправдано из-за высокого риска неверных выводов.

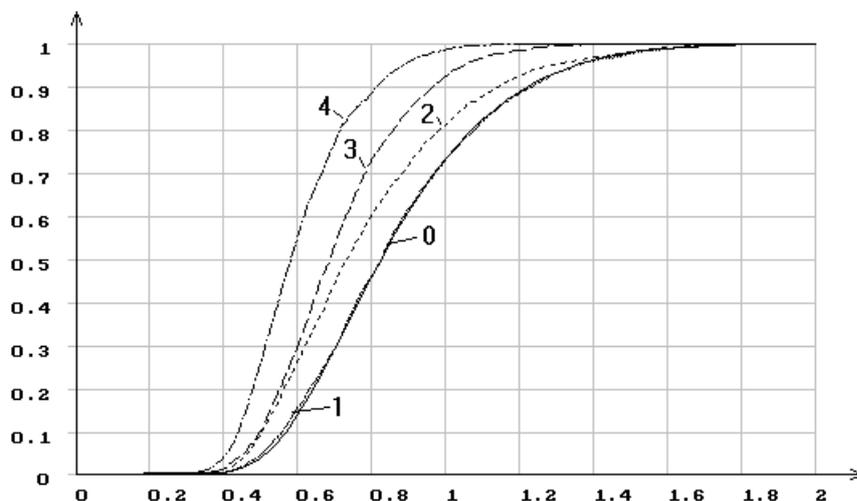


Рис.4. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при различном количестве оцениваемых параметров распределения Коши

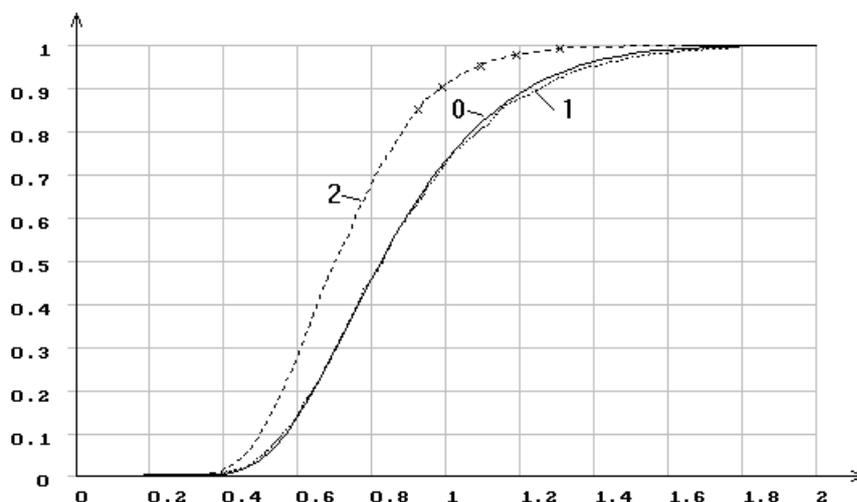


Рис.5. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при различном количестве оцениваемых параметров экспоненциального закона: 0 - функция распределения Колмогорова; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался масштабный параметр.

При идентификации типов предельных законов распределения непараметрических статистик в зависимости от вида закона наблюдаемой случайной величины и количества оцениваемых по выборке параметров использовалось множество из 26 законов и семейств распределений, включенных в программную систему [21]. Оказалось, что почти всегда эмпирические законы распределения статистик непараметрических критериев наиболее хорошо описываются

одним из двух законов распределения: *логарифмически нормальным* или *гамма-распределением*. То есть, эмпирическая функция статистики хорошо согласуется с одним из этих законов по всем используемым в системе критериям согласия: отношения правдоподобия, χ^2 Пирсона, Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса.

На рис. 6 представлены результаты выравнивания распределения статистики Колмогорова при оценивании одновременно двух параметров нормального распределения. На рисунке приведены эмпирическая функция распределения статистики и функция распределения логарифмически нормального распределения с параметрами $\theta_1 = -0.4849$, $\theta_0 = 0.2254$ (они практически совпали). Здесь же отражены значения статистик всех используемых при проверке согласия критериев и соответствующие вероятности вида $P\{S > S^*\}$. Факт оценивания параметров логарифмически нормального распределения в критериях отношения правдоподобия и χ^2 Пирсона учитывается уменьшением числа степеней свободы χ^2 -распределения. Причем в критериях использовано асимптотически оптимальное группирование наблюдений при 15 интервалах, максимизирующее мощность при близких альтернативах. Приведенные на рис. 6 значения вероятностей $P\{S > S^*\}$ для непараметрических статистик соответствуют классическим критериям (без учета факта оценивания параметров по данной выборке) и являются существенно завышенными. Если использовать представленные ниже в табл. 2, 5, 7, 9 результаты по аппроксимации смоделированных предельных распределений статистик, то вероятность вида $P\{S > S^*\}$ для критерия Колмогорова составит 0.6067, а не 0.8905, для критерия Смирнова - 0.4416, а не 0.5111, для критерия ω^2 Мизеса - 0.6277, а не 0.8897, для критерия Ω^2 Мизеса - 0.4737, а не 0.8397.

В табл. 2 сведены результаты идентификации законов для статистики критерия Колмогорова. Указание в клетке на конкретное распределение означает, что выборка соответствующей статистики хорошо описывается данным законом (согласуется с законом). В случае если согласие с каким-то законом не очень хорошее (гипотеза о согласии принимается с уровнем значимости $\alpha < 0.1$), то соответствующий закон указан на сером фоне. В таблицах через $\ln N(\theta_1, \theta_0)$ 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

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

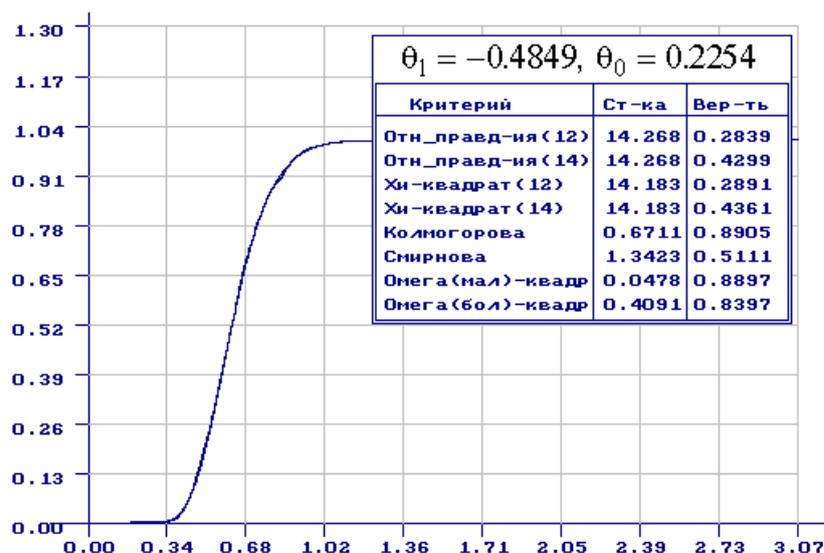


Рис.6. Эмпирическая функция распределения статистики Колмогорова при справедливой гипотезе H_0 и оценивании 2-х параметров нормального распределения и выравнивающая её функция распределения логарифмически нормального распределения

Таблица 2

<i>Предельные распределения статистики Колмогорова</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-0.3477, 0.2638)$		
2.	Полунормальное	$\gamma(3.4090, 8.2385, 0.3443)$		
3.	Рэлея	$\ln N(-0.3366, 0.2579)$		
4.	Максвелла	$\gamma(3.4809, 9.0801, 0.3417)$		
5.	Лапласа	$\gamma(3.2121, 6.5137, 0.3400)$	$\ln N(-0.3721, 0.2426)$	$\ln N(-0.4679, 0.2329)$
6.	Нормальное	$\gamma(3.6448, 7.0208, 0.3164)$	$\ln N(-0.4349, 0.2337)$	$\ln N(-0.4849, 0.2254)$
7.	Логнормальное	$\gamma(3.7326, 7.4146, 0.3265)$	$\gamma(3.0857, 8.4464, 0.3532)$	$\gamma(3.7311, 10.045, 0.3062)$
8.	Коши	$\gamma(3.1388, 6.7500, 0.3261)$	$\ln N(-0.3691, 0.2542)$	$\gamma(4.2049, 13.595, 0.2983)$
9.	Логистическое	$\gamma(3.3283, 6.6563, 0.3280)$	$\ln N(-0.4681, 0.2248)$	$\ln N(-0.5684, 0.2111)$
10.	Наибольшего значения	$\gamma(3.5340, 6.9273, 0.3302)$	$\ln N(-0.3588, 0.2555)$	$\ln N(-0.4950, 0.2231)$
11.	Наименьшего значения	$\gamma(3.4746, 6.9386, 0.3362)$	$\gamma(3.7737, 9.3627, 0.3293)$	$\ln N(-0.5009, 0.2221)$
12.	Вейбулла	$\gamma(3.6550, 9.0285, 0.3282)$	$\ln N(-0.2325, 0.2974)$ **	$\ln N(-0.5026, 0.2202)$

** - оценивался параметр формы распределения Вейбулла.

Отметим, что предельные распределения статистики Колмогорова для выборок из распределения Вейбулла-Гнеденко, как это и доказано в [18], не зависят от параметров. В случае распределений экстремальных значений и распределения Вейбулла-Гнеденко могут использоваться одни и те же предельные распределения.

Посмотрим, что будет получаться, если, например, мы будем использовать распределение Колмогорова для вычисления вероятности вида $P\{S > S^*\}$ в случае, когда по выборке предварительно вычисляются оценки параметров нормального распределения. Распределение статистики $\sqrt{S_k}/2$ в этом случае хорошо описывается логарифмически нормальным $\ln N(-0.4849, 0.2254)$. Для сравнения в табл. 3 для некоторых значений статистик S^* представлены соответствующие значения функции распределения Колмогорова, данного логарифмически нормального распределения и эмпирической функции распределения $F(\sqrt{S/2})$, полученной при моделировании распределения статистики. Так, для распределения Колмогорова величина $P\{\sqrt{S^2/2} > 0.9\} = 0.392731$, а для логарифмически нормального $P\{\sqrt{S^2/2} > 0.9\} = 0.04612$. Это означает, что при значении статистики $\sqrt{S/2} = 0.9$, используя распределение Колмогорова мы безоговорочно примем гипотезу H_0 , когда на самом деле даже при уровне значимости $\alpha = 0.05$ она должна быть отклонена.

Таблица 3

$\sqrt{S/2}$	0.5	0.9	1.0	1.22	1.36
$K(\sqrt{S/2})$	0.036055	0.607269	0.730000	0.898102	0.990154
$\ln N(\sqrt{S/2})$	0.177708	0.953880	0.984268	0.998791	0.999780
$F(\sqrt{S/2})$	0.19	0.9533	0.98425	0.999	1

Распределения, приводимые в табл. 2, удобны для реализации в программном обеспечении задач статистического анализа, как это сделано в [21], так как для вычисления вероятностей вида $P\{S > S^*\}$, соответствующих логарифмически нормальному или гамма-распределениям, не обойтись без процедур численного интегрирования или использования приближенных формул. В то же время для большинства практиков, применяющих статистические методы в различных приложениях, более привычно опираться на верхние процентные точки распределений статистик. В табл. 4 приводятся приближенные значения верхних процентных точек для статистики $\sqrt{S_k}/2$, которые для более высокой точности рассчитаны не по распределениям табл. 2, а получены сглаживанием “хвостов” эмпирических распределений. Точность этих процентных точек вполне достаточна для корректного использования критерия в задачах статистического анализа.

Таблица 4

<i>Процентные точки распределения статистики Колмогорова</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.9291	0.9872	1.0861	1.1846	1.3145
2.	Полунормальное	масштабный	0.9879	1.0519	1.1539	1.2511	1.3752
3.	Рэля	масштабный	0.9402	0.9999	1.0952	1.1859	1.3017
4.	Максвелла	масштабный	0.9284	0.9890	1.0853	1.1770	1.2938
5.	Лапласа	масштабный	1.1081	1.1897	1.3222	1.4501	1.6147
		сдвиг	0.8914	0.9435	1.0240	1.0992	1.1935
		два параметра	0.7966	0.8467	0.9261	1.0016	1.0978
6.	Нормальное	масштабный	1.1208	1.2081	1.3446	1.4731	1.6356
		сдвиг	0.8330	0.8790	0.9497	1.0156	1.0982
		два параметра	0.7808	0.8255	0.8954	0.9611	1.0442
7.	Логнормальное	масштабный	1.0880	1.1736	1.3147	1.4523	1.6308
		сдвиг	0.9147	0.9875	1.1083	1.2266	1.3805
		два параметра	0.8539	0.9268	1.0500	1.1723	1.3330
8.	Коши	масштабный	1.0590	1.1497	1.2950	1.4339	1.6116
		сдвиг	0.9080	0.9659	1.0589	1.1481	1.2623
		два параметра	0.7620	0.8117	0.8868	0.9557	1.0414
9.	Логистическое	масштабный	1.1034	1.1957	1.3441	1.4864	1.6689
		сдвиг	0.7912	0.8373	0.9109	0.9813	1.0714
		два параметра	0.7060	0.7400	0.7964	0.8516	0.9234
10.	Наибольшего значения	масштабный	1.1157	1.2033	1.3402	1.4689	1.6315
		сдвиг	0.9137	0.9775	1.0794	1.1765	1.3004
		два параметра	0.7705	0.8119	0.8808	0.9485	1.0367
11.	Наименьшего значения	масштабный	1.1027	1.1897	1.3286	1.4612	1.6305
		сдвиг	0.9386	1.0048	1.1103	1.2108	1.3390
		два параметра	0.7655	0.8080	0.8758	0.9405	1.0233
12.	Вейбулла	формы	1.0824	1.1659	1.3043	1.4396	1.6157
		масштаба	0.9425	1.0064	1.1087	1.2065	1.3315
		два параметра	0.7634	0.8022	0.8658	0.9276	1.0074

3. Предельные распределения статистики Смирнова

На рис. 7 представлены результаты моделирования статистики Смирнова S_m при справедливой гипотезе H_0 , соответствующей нормальному распределению, на рис. 8, соответствующей логистическому распределению, на рис. 9, соответствующей распределению Лапласа, на рис. 10. - распределению Коши, а на рис. 11 - экспоненциальному распределению. На этих рисунках символом "0" отмечена функция распределения χ_2^2 , которому подчиняется статистика Смирнова, если по выборке не оцениваются параметры.

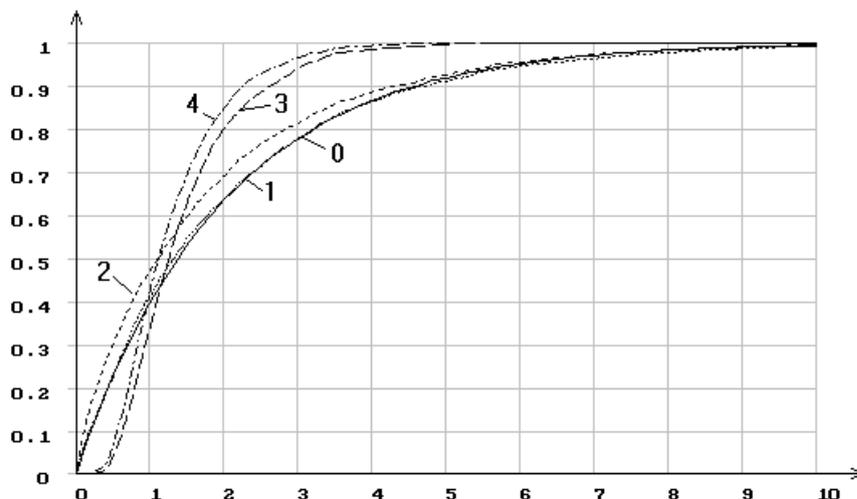


Рис.7. Эмпирические функции распределения статистики Смирнова при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения χ_2^2 (с двумя степенями свободы); 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра

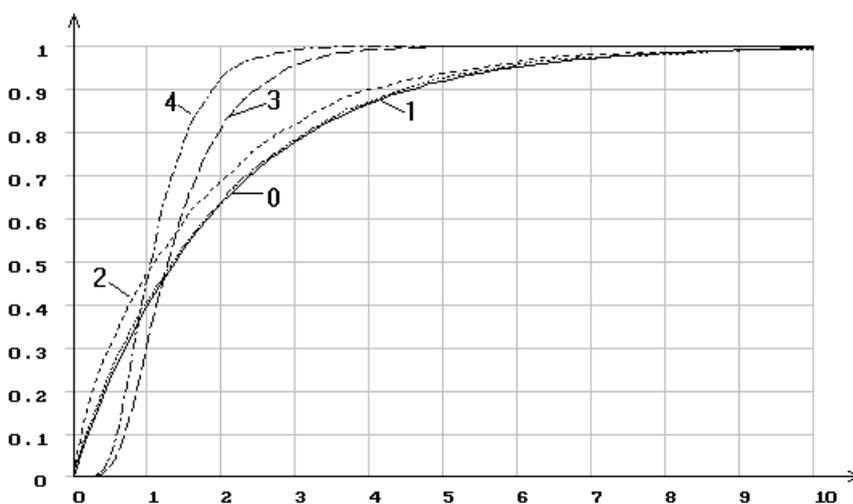


Рис.8. Эмпирические функции распределения статистики S_m Смирнова при различном количестве оцениваемых параметров логистического распределения

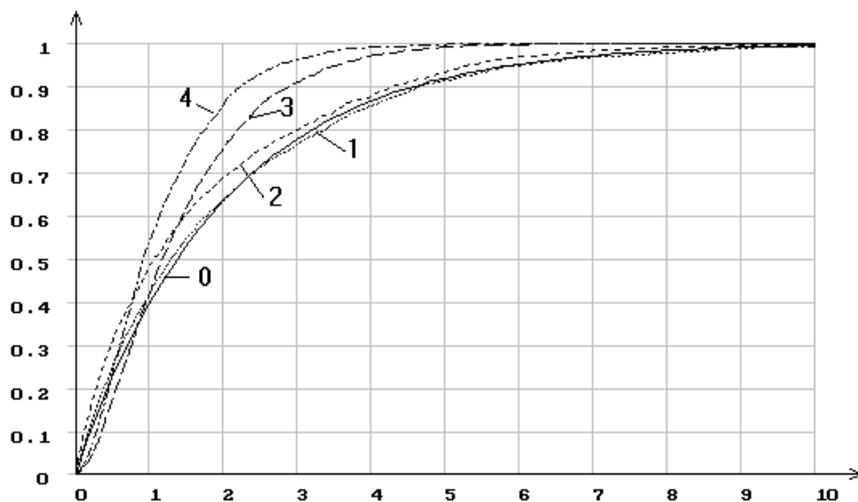


Рис.9. Эмпирические функции распределения статистики S_m Смирнова при различном количестве оцениваемых параметров распределения Лапласа

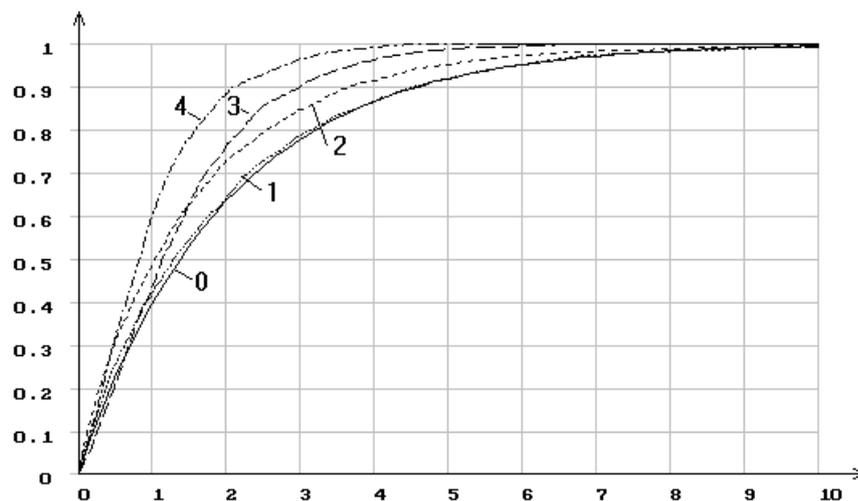


Рис.10. Эмпирические функции распределения статистики S_m Смирнова при различном количестве оцениваемых параметров распределения Коши

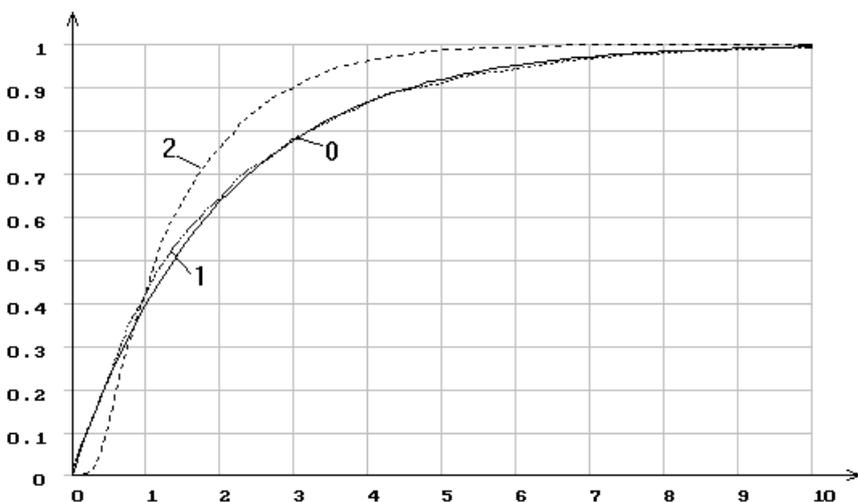


Рис.11. Эмпирические функции распределения статистики S_m Смирнова без оценивания параметров (1), при оценивании масштабного параметра (2) экспоненциального распределения

Как видим, и в данном случае предельные распределения статистики Смирнова при условии оценивания параметров конкретного закона кардинально отличаются от распределения χ^2_2 .

Выше уже говорилось, что о точности моделирования можно судить по результатам моделирования классических непараметрических статистик (без оценивания параметров). На рис. 12 для иллюстрации приведены результаты проверки гипотез о согласии с χ^2_2 -распределением смоделированной выборки классической статистики Смирнова в случае нормального закона (см. рис. 7). На рисунке отражены функция плотности χ^2_2 -распределения и гистограмма, построенная по выборке при асимптотически оптимальном группировании.



Рис.12. Результаты статистического анализа смоделированной выборки классической статистики Смирнова (для нормального закона без оценивания параметров)

В табл. 5 сведены результаты идентификации предельных законов для статистики критерия Смирнова.

Таблица 5

<i>Предельные распределения статистики Смирнова</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(0.1585, 0.7009)$		
2.	Полунормальное	$\ln N(0.1289, 0.7900)$		
3.	Рэля	$\ln N(0.1936, 0.7073)$		
4.	Максвелла	$\ln N(0.2221, 0.6794)$		
5.	Лапласа	$\gamma(0.8146, 0.4654, 0.0006)$	$\gamma(1.7664, 1.2256, 0.0207)$	$\gamma(1.8235, 1.5842, 0.0058)$
6.	Нормальное	$\gamma(0.8088, 0.4549, 0.0006)$	$\ln N(0.2471, 0.5321)$	$\ln N(0.1299, 0.5331)$
7.	Логнормальное	$\gamma(0.8391, 0.4641, 0.0006)$	$\ln N(0.4252, 0.6481)$	$\ln N(0.1947, 0.6783)$
8.	Коши	$\gamma(0.8570, 0.5348, 0.0006)$	$\gamma(1.4215, 0.9846, 0.0006)$	$\gamma(1.2931, 1.2542, 0.0006)$

9.	Логистическое	$\gamma(0.8164,0.4709,0.0)$	$\ln N(0.2684,0.4856)$	$\ln N(0.0569,0.4491)$
10.	Наибольшего значения	$\gamma(0.8507,0.4682,0.0006)$	$\ln N(0.2667,0.6821)$	$\ln N(0.1049,0.5058)$
11.	Наименьшего значения	$\gamma(0.8234,0.4592,0.0006)$	$\ln N(0.2022,0.7015)$	$\ln N(0.1238,0.5024)$
12.	Вейбулла	$\ln N(0.1615,0.7186)$	$\gamma(0.8641,0.4851,0.0006)$ **	$\gamma(2.3495,2.3008,0.2564)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 6 приводятся приближенные значения верхних процентных точек для статистики S_m .

Таблица 6

<i>Процентные точки распределения статистики Смирнова</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	2.5341	2.9847	3.7675	4.5593	5.6145
2.	Полунормальное	масштабный	2.6856	3.2651	4.2439	5.2147	6.4902
3.	Рэлея	масштабный	2.6590	3.1390	3.9317	4.7058	5.7115
4.	Максвелла	масштабный	2.5751	3.0644	3.9162	4.7788	5.9296
5.	Лапласа	масштабный	3.5933	4.3253	5.5464	6.7473	8.3154
		сдвиг	2.5139	2.9110	3.5728	4.2230	5.0716
		два параметра	1.9821	2.2836	2.8034	3.3261	4.0201
6.	Нормальное	масштабный	3.4951	4.3824	5.8586	7.3074	9.1966
		сдвиг	2.2762	2.6147	3.1684	3.7055	4.3998
		два параметра	2.0255	2.3001	2.7552	3.2005	3.7801
7.	Логнормальное	масштабный	3.4308	4.2549	5.6948	7.1566	9.1105
		сдвиг	3.0417	3.6974	4.8495	6.0236	7.5972
		два параметра	2.3621	2.9020	3.9162	4.9983	6.4985
8.	Коши	масштабный	3.0648	3.7974	5.0833	6.3929	8.1473
		сдвиг	2.5211	2.9562	3.7225	4.5046	5.5542
		два параметра	1.8510	2.1830	2.7286	3.2596	3.9478
9.	Логистическое	масштабный	3.3625	4.1502	5.5238	6.9163	8.7754
		сдвиг	2.1914	2.4814	2.9781	3.4754	4.1334
		два параметра	1.6973	1.9035	2.2459	2.5814	3.0185
10.	Наибольшего значения	масштабный	3.5767	4.3936	5.7430	7.0609	8.7735
		сдвиг	2.6771	3.1331	3.9274	4.7318	5.8051
		два параметра	1.8947	2.1702	2.6283	3.0778	3.6638
11.	Наименьшего значения	масштабный	3.5733	4.3186	5.5688	6.8028	8.4186
		сдвиг	2.6298	3.1309	3.9946	4.8630	6.0157
		два параметра	1.9043	2.1615	2.6199	3.0917	3.7285
12.	Вейбулла	формы	3.4390	4.1895	5.5179	6.8786	8.7090
		масштаба	2.5623	3.0621	3.9347	4.8203	6.0035
		два параметра	1.9035	2.1477	2.5708	2.9977	3.5659

4. Предельные распределения статистики Крамера-Мизеса-Смирнова

Результаты моделирования статистики S_{ω} Мизеса (Крамера-Мизеса-Смирнова) при справедливой гипотезе H_0 для случая нормального распределения представлены на рис. 13. На этом рисунке крестиками нанесены соответствующие значения предельных распределений, полученные в работе Г.В. Мартынова [7]. Для случая логистического распределения результаты моделирования приведены на рис. 14, для случая распределения Лапласа - на рис. 15, для распределения Коши - на рис. 16, а на рис. 17 - для экспоненциального распределения, где для сравнения крестиками обозначены процентные точки при проверке экспоненциальности из таблицы М.Стефенса [12]. На этих рисунках символом "0" отмечена функция распределения $a1(s)$ [22], предельная для этой статистики, если по выборке не оценивались параметры распределения.

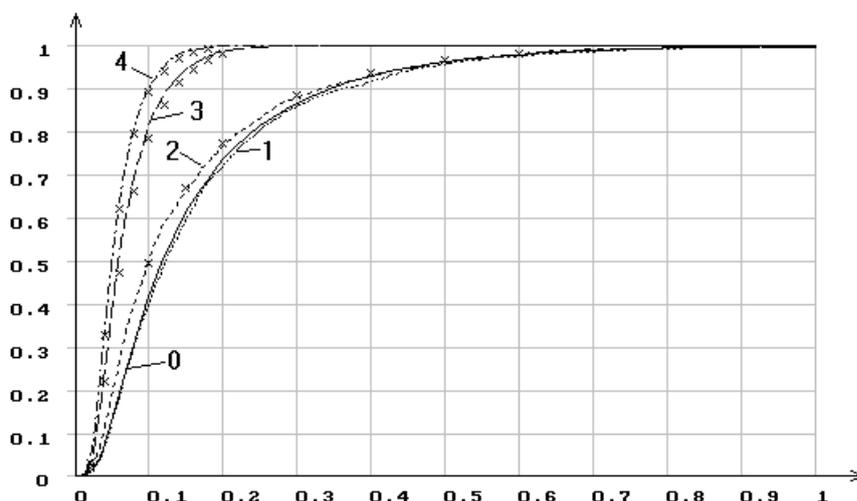


Рис.13. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения $a1(s)$; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра

В табл. 7 сведены результаты идентификации законов для статистики критерия ω^2 Мизеса. Следует отметить, что в данном случае предельные распределения хуже аппроксимируются логарифмически нормальным, которое в каждом случае оказалось наиболее подходящим из всего множества законов, включенных в систему [21].

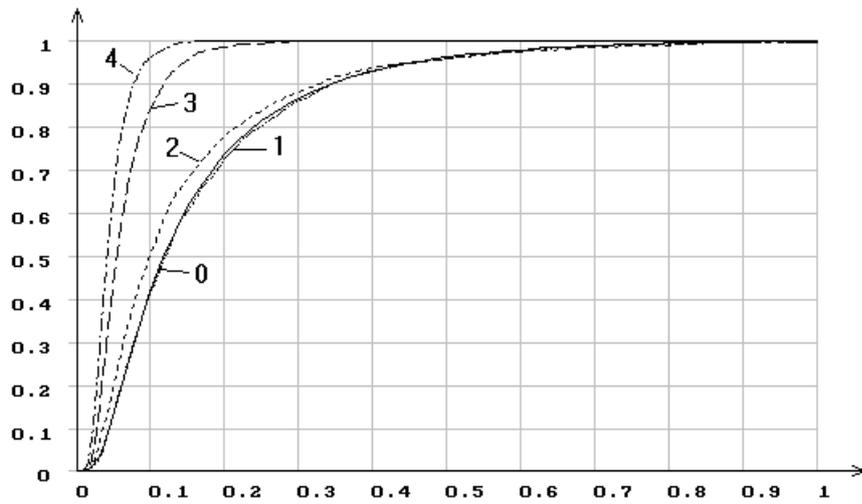


Рис.14. Эмпирические функции распределения статистики S_{\circ} Мизеса при различном количестве оцениваемых параметров логистического распределения

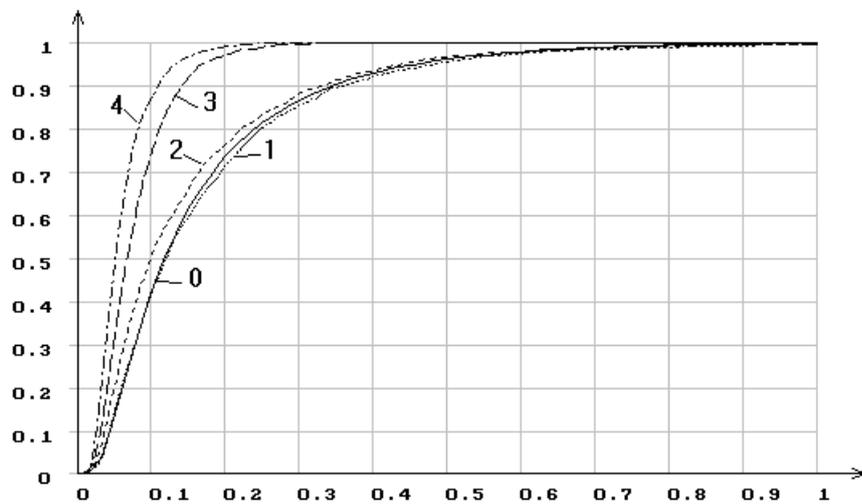


Рис.15. Эмпирические функции распределения статистики S_{\circ} Мизеса при различном количестве оцениваемых параметров распределения Лапласа

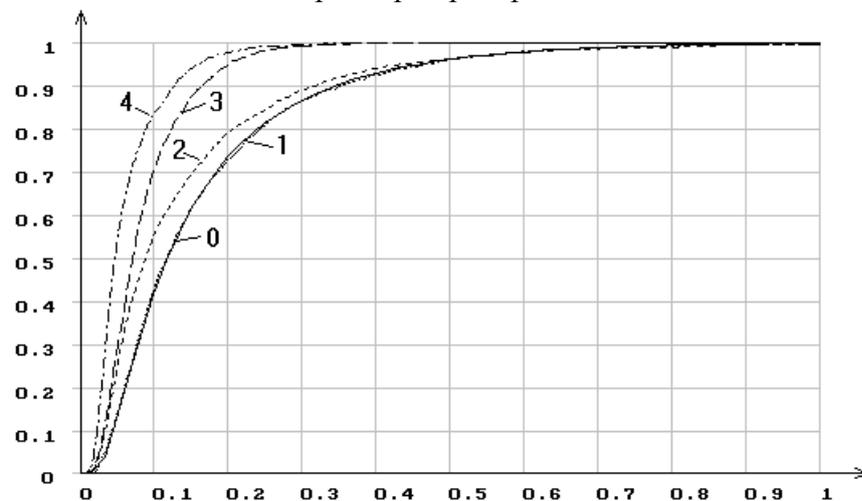


Рис.16. Эмпирические функции распределения статистики S_{\circ} Мизеса при различном количестве оцениваемых параметров распределения Коши

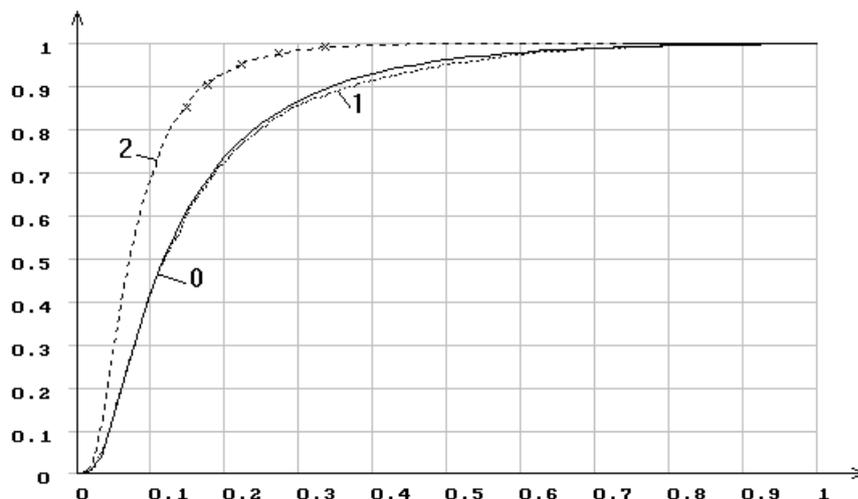


Рис.17. Эмпирические функции распределения статистики S_{ω} Мизеса при различном количестве оцениваемых параметров экспоненциального распределения

Таблица 7

<i>Предельные распределения статистики ω^2 Мизеса</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-2.6028, 0.6453)$		
2.	Полунормальное	$\ln N(-2.5046, 0.6814)$		
3.	Рэлея	$\ln N(-2.5743, 0.6345)$		
4.	Максвелла	$\ln N(-2.6147, 0.6361)$		
5.	Лапласа	$\ln N(-2.2328, 0.8302)$	$\ln N(-2.6890, 0.5802)$	$\ln N(-2.9386, 0.5500)$
6.	Нормальное	$\ln N(-2.2290, 0.8284)$	$\ln N(-2.8102, 0.5625)$	$\ln N(-2.9685, 0.5187)$
7.	Логнормальное	$\ln N(-2.2334, 0.7951)$	$\ln N(-2.6226, 0.6972)$	$\ln N(-2.8233, 0.6673)$
8.	Коши	$\ln N(-2.3228, 0.8554)$	$\ln N(-2.6269, 0.6202)$	$\ln N(-2.9746, 0.6493)$
9.	Логистическое	$\ln N(-2.2460, 0.8329)$	$\ln N(-2.8755, 0.5612)$	$\ln N(-3.1713, 0.4841)$
10.	Наибольшего значения	$\ln N(-2.2043, 0.8174)$	$\ln N(-2.6070, 0.6178)$	$\ln N(-3.0093, 0.5078)$
11.	Наименьшего значения	$\ln N(-2.2246, 0.8149)$	$\ln N(-2.5863, 0.6469)$	$\ln N(-3.0200, 0.5173)$
12.	Вейбулла	$\ln N(-2.5875, 0.6669)$	$\ln N(-2.2404, 0.7910)$ **	$\ln N(-3.0258, 0.5058)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 8 представлены полученные приближенные значения верхних процентных точек для статистики S_{ω} .

Таблица 8

Процентные точки распределения статистики ω^2 Мизеса							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.1463	0.1752	0.2263	0.2785	0.3486
2.	Полунормальное	масштабный	0.1730	0.2073	0.2673	0.3283	0.4098
3.	Рэлея	масштабный	0.1518	0.1790	0.2253	0.2713	0.3319
4.	Максвелла	масштабный	0.1463	0.1738	0.2208	0.2681	0.3306
5.	Лапласа	масштабный	0.2702	0.3394	0.4594	0.5809	0.7427
		сдвиг	0.1249	0.1440	0.1770	0.2103	0.2546
		два параметра	0.0942	0.1103	0.1377	0.1651	0.2014
6.	Нормальное	масштабный	0.2711	0.3414	0.4636	0.5871	0.7515
		сдвиг	0.1096	0.1270	0.1561	0.1847	0.2222
		два параметра	0.0874	0.0998	0.1208	0.1417	0.1691
7.	Логнормальное	масштабный	0.2599	0.3235	0.4347	0.5476	0.6987
		сдвиг	0.1481	0.1902	0.2709	0.3580	0.4799
		два параметра	0.1134	0.1474	0.2154	0.2912	0.3996
8.	Коши	масштабный	0.2565	0.3278	0.4545	0.5848	0.7604
		сдвиг	0.1416	0.1656	0.2060	0.2461	0.2987
		два параметра	0.1062	0.1285	0.1660	0.2031	0.2518
9.	Логистическое	масштабный	0.2648	0.3369	0.4667	0.6012	0.7838
		сдвиг	0.1027	0.1200	0.1507	0.1822	0.2245
		два параметра	0.0692	0.0784	0.0943	0.1161	0.1354
10.	Наибольшего значения	масштабный	0.2752	0.3413	0.4601	0.5829	0.7494
		сдвиг	0.1435	0.1697	0.2133	0.2560	0.3117
		два параметра	0.0840	0.0963	0.1172	0.1380	0.1654
11.	Наименьшего значения	масштабный	0.2649	0.3324	0.4541	0.5804	0.7518
		сдвиг	0.1500	0.1774	0.2269	0.2785	0.3488
		два параметра	0.0848	0.0974	0.1182	0.1383	0.1644
12.	Вейбулла	формы	0.2521	0.3165	0.4292	0.5438	0.6971
		масштаба	0.1535	0.1855	0.2405	0.2956	0.3686
		два параметра	0.0833	0.0940	0.1125	0.1311	0.1557

5. Предельные распределения статистики Андерсона-Дарлингга

На рис. 18-22 отражены результаты моделирования предельных законов распределения статистики S_{Ω} Мизеса (Андерсона-Дарлингга). На рисунках символом “0” отмечена функция распределения $a_2(s)$ [18]. Этому распределению подчиняется статистика S_{Ω} , если по выборке не оцениваются параметры наблюдаемого закона.

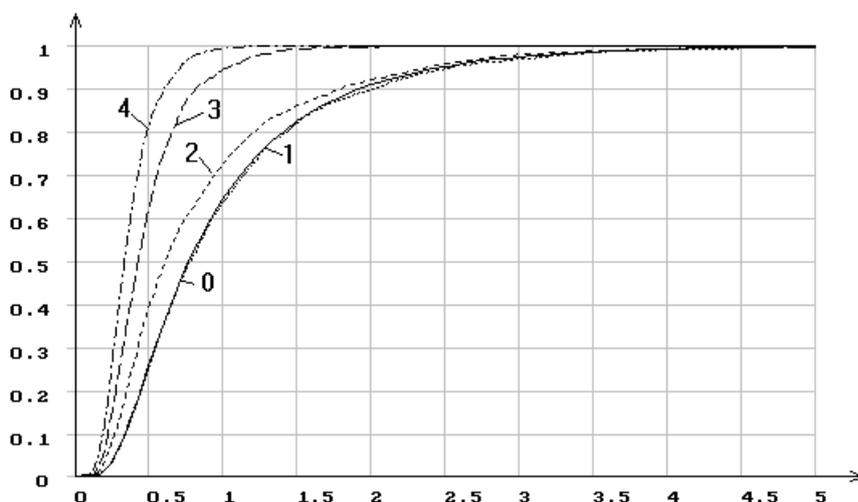


Рис. 18. Эмпирические функции распределения статистики S_{Ω} Мизеса при различном количестве оцениваемых параметров нормального закона: 0 - функция распределения $a2(s)$; 1 - по выборке не оценивались параметры; 2 - по выборке оценивался только масштабный параметр θ_0 ; 3 - оценивался только параметр сдвига θ_1 ; 4 - оценивались одновременно оба параметра.

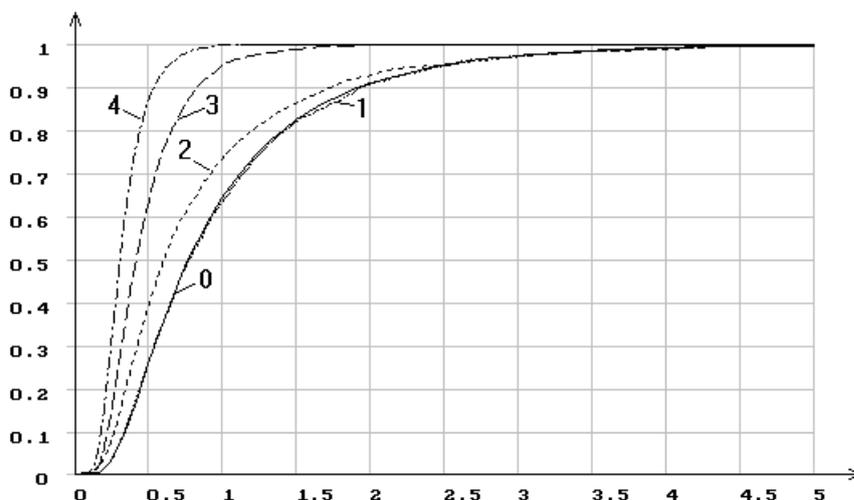


Рис.19. Эмпирические функции распределения статистики S_{Ω} Мизеса при различном количестве оцениваемых параметров логистического распределения

В табл. 9 сведены результаты идентификации законов для статистики критерия Ω^2 Мизеса. Как и в предыдущем случае здесь предельные распределения также хуже аппроксимируются логарифмически нормальным распределением.

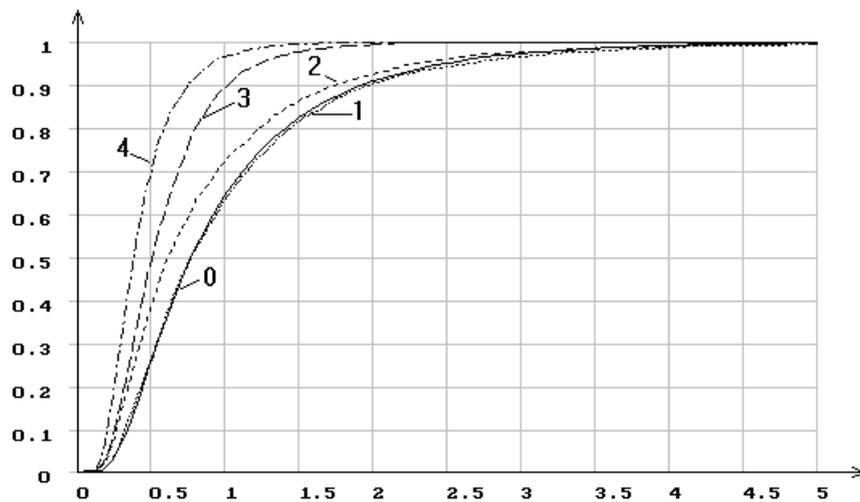


Рис.20. Эмпирические функции распределения статистики S_{Ω} Мизеса при различном количестве оцениваемых параметров распределения Лапласа

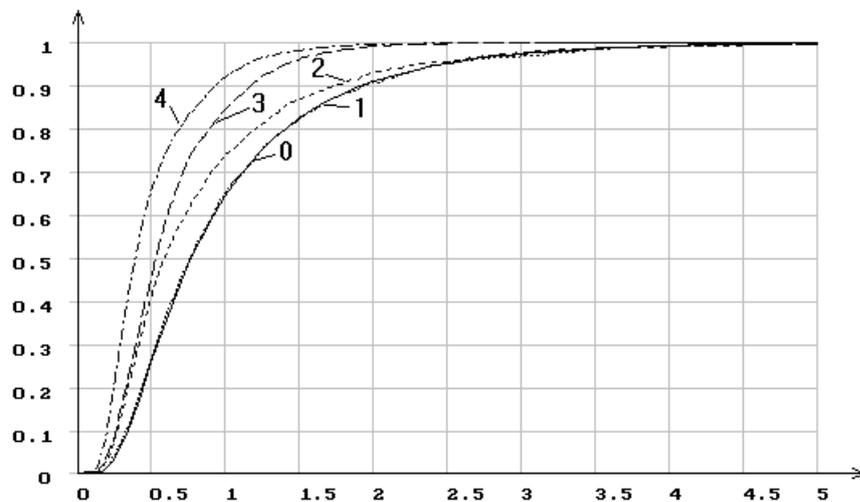


Рис.21. Эмпирические функции распределения статистики S_{Ω} Мизеса при различном количестве оцениваемых параметров распределения Коши

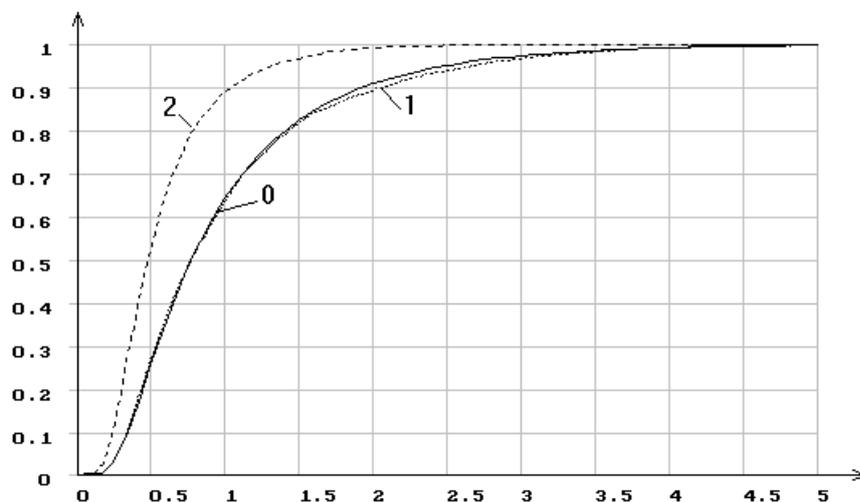


Рис.22. Эмпирические функции распределения статистики S_{Ω} Мизеса при различном количестве оцениваемых параметров экспоненциального распределения

Таблица 9

<i>Предельные распределения статистики Ω^2 Мизеса</i>				
№ п/п	Распределение случайной величины	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
1.	Экспоненциальное	$\ln N(-0.7055, 0.5690)$		
2.	Полунормальное	$\ln N(-0.6931, 0.5900)$		
3.	Рэля	$\ln N(-0.6850, 0.5541)$		
4.	Максвелла	$\ln N(-0.7051, 0.5658)$		
5.	Лапласа	$\ln N(-0.4107, 0.7170)$	$\ln N(-0.6654, 0.5345)$	$\ln N(-0.9378, 0.4999)$
6.	Нормальное	$\ln N(-0.4121, 0.7206)$	$\ln N(-0.8363, 0.5096)$	$\ln N(-1.0840, 0.4509)$
7.	Логнормальное	$\ln N(-0.4092, 0.6938)$	$\ln N(-0.6827, 0.6146)$	$\ln N(-0.9322, 0.5819)$
8.	Коши	$\ln N(-0.4326, 0.7164)$	$\ln N(-0.6102, 0.5737)$	$\ln N(-0.8815, 0.5905)$
9.	Логистическое	$\ln N(-0.4243, 0.7224)$	$\ln N(-0.8465, 0.5109)$	$\ln N(-1.1685, 0.4239)$
10.	Наибольшего значения	$\ln N(-0.3863, 0.7097)$	$\ln N(-0.6903, 0.5534)$	$\ln N(-1.0809, 0.4447)$
11.	Наименьшего значения	$\ln N(-0.4007, 0.7108)$	$\ln N(-0.6866, 0.5667)$	$\ln N(-1.0986, 0.4474)$
12.	Вейбулла	$\ln N(-0.6908, 0.5866)$	$\ln N(-0.4177, 0.6846)$ **	$\ln N(-1.1044, 0.4405)$

** - оценивался параметр формы распределения Вейбулла.

В табл. 10 приведены приближенные значения верхних процентных точек для статистики S_{Ω} .

Таблица 10

<i>Процентные точки распределения статистики Ω^2 Мизеса</i>							
№ п/п	Распределение случайной величины	Оцениваемые параметры	Верхние процентные точки				
			0.15	0.1	0.05	0.025	0.01
1.	Экспоненциальное	масштабный	0.8975	1.0599	1.3407	1.6235	1.9995
2.	Полунормальное	масштабный	0.9912	1.1672	1.4791	1.7987	2.2289
3.	Рэля	масштабный	0.9168	1.0629	1.3128	1.5629	1.8938
4.	Максвелла	масштабный	0.8975	1.0569	1.3342	1.6150	1.9896
5.	Лапласа	масштабный	1.4505	1.7830	2.3789	2.9943	3.8275
		сдвиг	0.8979	1.0326	1.2598	1.4851	1.7810
		два параметра	0.6604	0.7593	0.9312	1.1051	1.3369
6.	Нормальное	масштабный	1.4538	1.8026	2.4220	3.0579	3.9146
		сдвиг	0.7300	0.8371	1.0340	1.2410	1.5249
		два параметра	0.5409	0.6095	0.7239	0.8366	0.9837
7.	Логнормальное	масштабный	1.4278	1.7433	2.2896	2.8407	3.5739
		сдвиг	0.9489	1.1734	1.5913	2.0342	2.6453
		два параметра	0.6893	0.8589	1.2006	1.5829	2.1318
8.	Коши	масштабный	1.4282	1.7664	2.3817	3.0238	3.8996

		сдвиг	1.0213	1.1849	1.4549	1.7185	2.0607
		два параметра	0.8019	0.9480	1.1928	1.4345	1.7508
9.	Логистическое	масштабный	1.4373	1.7879	2.4150	3.0617	3.9361
		сдвиг	0.7344	0.8489	1.0486	1.2510	1.5213
		два параметра	0.4849	0.5423	0.6364	0.7277	0.8460
10.	Наибольшего значения	масштабный	1.4936	1.8259	2.4120	3.0108	3.8148
		сдвиг	0.9061	1.0523	1.3054	1.5607	1.9004
		два параметра	0.5443	0.6107	0.7248	0.8392	0.9907
11.	Наименьшего значения	масштабный	1.4538	1.7934	2.3948	3.0111	3.8404
		сдвиг	0.9088	1.0660	1.3426	1.6250	2.0039
		два параметра	0.5398	0.6092	0.7198	0.8251	0.9594
12.	Вейбулла	формы	1.3909	1.7030	2.2511	2.8096	3.5579
		масштаба	0.9355	1.1049	1.4018	1.7039	2.1081
		два параметра	0.5253	0.5908	0.6991	0.8052	0.9432

6. Предельные распределения статистик для выборок из гамма-распределения

Как и следовало ожидать, предельные распределения рассматриваемых статистик для выборок из гамма-распределения с плотностью

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

зависят от параметра формы θ_0 . Выборки статистик моделировались при различных значениях параметра θ_0 . Для иллюстрации мы приведем лишь результаты моделирования статистики Колмогорова $\sqrt{S_k/2}$. На рис. 23 представлены эмпирические распределения статистики при оценивании по выборке параметра формы, на рис. 24 - масштабного параметра, на рис. 25 - двух параметров распределения. На этих рисунках "1" помечена эмпирическая функция распределения статистики при $\theta_0=0.5$; "2" - при $\theta_0=1.0$; "3" - при $\theta_0=2.0$; "4" - при $\theta_0=5.0$; "5" - при $\theta_0=10.0$. "0" помечена функция распределения Колмогорова. Отметим, что с ростом θ_0 предельные распределения статистик стремятся к предельным распределениям статистик для выборок из нормального закона. При значениях $\theta_0 > 5$ эмпирические распределения статистик при оценивании двух параметров практически совпадают и хорошо согласуются с распределением соответствующей статистики для нормального закона.

Общая картина принципиально сохраняется и для распределений других непараметрических статистик.

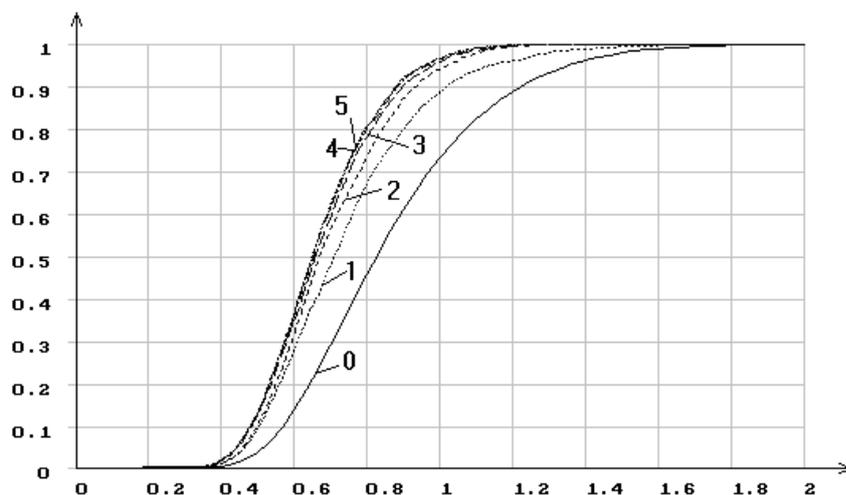


Рис.23. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при оценивании параметра формы гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0=0.5$; 2 - $\theta_0=1.0$; 3 - $\theta_0=2.0$; 4 - $\theta_0=5.0$; 5 - $\theta_0=10.0$.

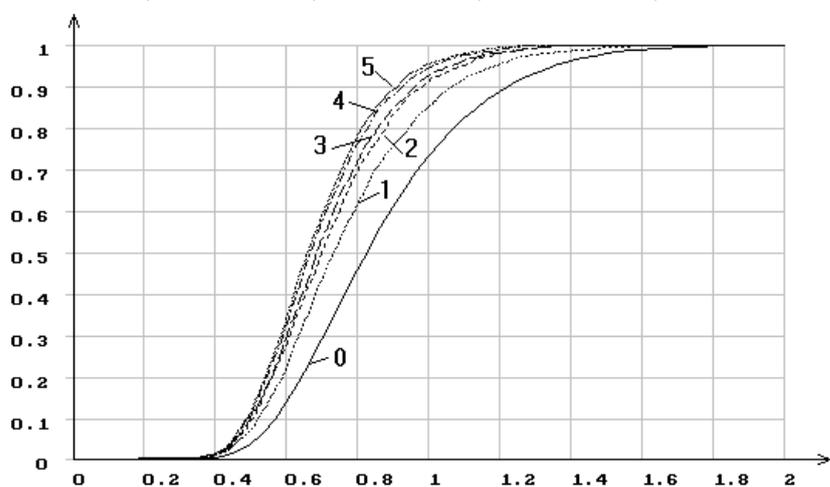


Рис.24. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при оценивании масштабного параметра гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0=0.5$; 2 - $\theta_0=1.0$; 3 - $\theta_0=2.0$; 4 - $\theta_0=5.0$; 5 - $\theta_0=10.0$.

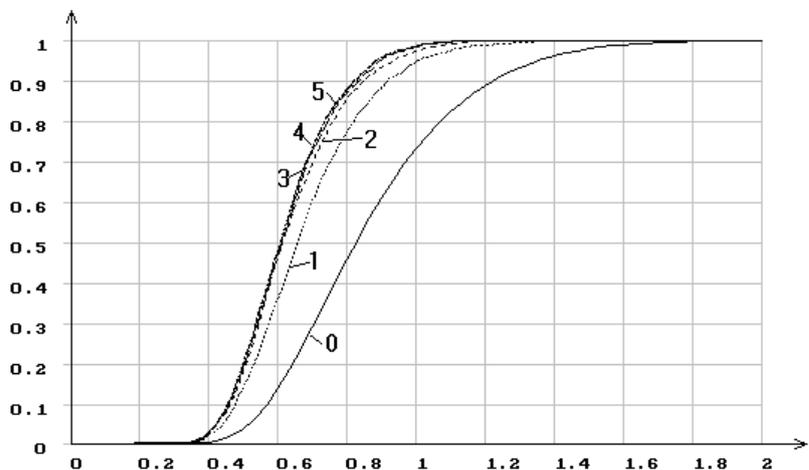


Рис.25. Эмпирические функции распределения статистики $\sqrt{S_k}/2$ Колмогорова при оценивании одновременно двух параметров гамма-распределения: 0 - функция распределения Колмогорова; 1 - $\theta_0=0.5$; 2 - $\theta_0=1.0$; 3 - $\theta_0=2.0$; 4 - $\theta_0=5.0$; 5 - $\theta_0=10.0$.

7. Заключение

Можно констатировать, что на основании результатов моделирования могут быть получены хорошие приближения предельных распределений непараметрических статистик.

Для законов, определяемых только параметрами сдвига и масштаба наиболее значительное влияние на предельные распределения всех непараметрических статистик оказывает оценивание параметра сдвига, в существенно меньшей степени - оценивание масштабного параметра.

Достаточно хорошая аппроксимация для реальных распределений статистик непараметрических критериев обычно может быть получена с использованием логарифмически нормального распределения и/или гамма-распределения. Для ряда законов распределения случайных величин идентифицированы законы распределения статистик непараметрических критериев при различном количестве оцененных параметров, построены таблицы верхних процентных точек. Полученные распределения при практическом использовании критериев согласия позволят делать более надежные статистические выводы.

В случае необходимости, вообще говоря, несложно провести аналогичное моделирование распределений статистик непараметрических критериев для любого закона распределения случайной величины.

Результаты проведенных исследований еще раз подчеркивают, что предельные распределения статистик непараметрических критериев согласия Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса при оценивании по выборке параметров в случае справедливости гипотезы H_0 настолько сильно отличаются от распределений классических статистик, соответственно законов $K(s)$, χ_2^2 , $a1(s)$ и $a2(s)$, что последние ни в коем случае не должны использоваться в такой ситуации.

Литература

1. 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.
2. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория, 1985. Т. 51. №1. - С. 60-62.
3. Бондарев Б.В. О проверке сложных статистических гипотез // Заводская лаборатория. 1986. Т. 52. № 10. - С. 62-63.
4. Кулинская Е.В., Саввушкина Н.Е. О некоторых ошибках в реализации и применении непараметрических методов в пакете для IBM PC // Заводская лаборатория, 1990. Т. 56. № 5. - С. 96-99.

5. Durbin J. Kolmogorov-Smirnov tests when parameters are estimated // Lect. Notes Math., 1976. V.566. - P.33-44.
6. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Физматлит. 1995. - 240 с.
7. Мартынов Г.В. Критерии омега-квадрат. - М.: Наука, 1978. - 80 с.
8. Birnbaum Z.W. Numerical tabulation of the distribution of Kolmogoroff's statistic for finite sample size // J. Am. Statist. Assoc., 1952. V.47. - P.425-441.
9. Anderson T.W., Darling D.A. Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes // Ann. Math. Stat., 1952. V.23. - P.193-212.
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. Pearson E.S., Hartley H.O. Biometrika tables for Statisticians. Vol. 2. - Cambridge: University Press. 1972. - 634 p.
13. Stephens M.A. Use of Kolmogorov - Smirnov, Cramer - von Mises and related statistics - without extensive tables // J. R. Stat. Soc., 1970. B.32. - P.115-122.
14. Stephens M.A. EDF statistics for goodness of fit and some comparisons // J. Am. Statist. Assoc., 1974. V.69. - P.730-737.
15. Chandra M., Singpurwalla N.D., Stephens M.A. Statistics for Test of Fit for the Extreme-Value and Weibull Distribution // J. Am. Statist. Assoc., 1981. V.76. - P.375.
16. Durbin J. Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentially and tests of spacings // Biometrika, 1975. V.62, № 1. - P.5-22.
17. Тюрин Ю.Н. О предельном распределении статистик Колмогорова-Смирнова для сложной гипотезы / Известия АН СССР. Сер. Матем. 1984. Т. 48. № 6. - С. 1314-1343.
18. Тюрин Ю.Н., Саввушкина Н.Е. Критерий согласия для распределения Вейбулла-Гнеденко / Известия АН СССР. Сер. Техн. кибернетика. 1984. № 3. - С. 109-112.
19. Тюрин Ю.Н. Исследования по непараметрической статистике (непараметрические методы и линейная модель). Автореф. дисс. докт. физ.-мат. наук. - М.: МГУ, 1985. - 33 с.
20. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. - М.: ИНФРА-М, Финансы и статистика, 1995. - 384 с.
21. Лемешко Б.Ю. Статистический анализ одномерных наблюдений случайных величин: Программная система. - Новосибирск: Изд-во НГТУ. - 1995. - 125 с.
22. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.