Мат. международной НТК "Информатика и проблемы телекоммуникаций". - Новосибирск, 1997. - С. 117-120.

О РАСПРЕДЕЛЕНИЯХ СТАТИСТИК НЕПАРАМЕТРИЧЕСКИХ КРИТЕРИЕВ ПРИ ПОТЕРЕ СВОЙСТВА “СВОБОДЫ ОТ РАСПРЕДЕЛЕНИЯ”

Б.Ю.Лемешко, С.Н.Постовалов
Новосибирский государственный технический университет

Наиболее часто в практике статистического анализа с необходимостью использования критериев согласия приходится сталкиваться после оценивания по этой же выборке параметров предполагаемого закона распределения. В этом случае непараметрические критерии Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса 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 $f(x,\theta)$.

На практике при использовании непараметрических критериев согласия факт зависимости предельного распределения от оценивания параметров по разным причинам обычно не учитывается: берется предельное распределение статистики, как будто параметры и не оценивались. Это приводит к сильно завышенным значениям вероятностей “согласия” вида $P\{S > S^*\}$, где S^* - значение статистики, вычисленное по выборке.

Очевидно, что теоретически найти предельные распределения этих статистик в зависимости от того, сколько параметров оценивалось по выборке и с каким законом проверяется согласие, для множества законов, используемых для описания реальных величин, очень сложно. Один из возможных выходов заключается в моделировании эмпирических законов распределения статистик непараметрических критериев и в последующей идентификации этих законов.

Целью настоящего исследования явилось стремление 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. В данной работе мы приводим часть результатов моделирования и анализа, которые, с нашей точки зрения, могут с успехом применяться при решении практических задач проверки гипотез о согласии с использованием непараметрических критериев.

Статистика Колмогорова определяется выражением [1]

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

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

n - объем выборки, $x_1, x_2, \ldots, x_n$ - 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 $K(x)$ [1].

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 рисунке представлены эмпирические функции распределения статистики, когда по выборке не оценивались параметры (*а*), по выборке оценивался только масштабный параметр σ (*б*), оценивался только параметр сдвига μ (*в*), оценивались одновременно оба параметра (*г*). Здесь же приведена функция распределения Колмогорова (*д*), которому подчиняется статистика $\sqrt{S_k / 2}$, если по выборке не оцениваются параметры. Результаты проверки согласия эмпирического распределения (*а*) с распределением Колмогорова (*д*) очень хорошее. В то же время весьма наглядно отличие эмпирических функций распределения *б, в, г* от распределения Колмогорова (*д*). Это отличие позволяет судить о величине тех ошибок, которые мы допускаем, не учитывая факта оценивания параметров конкретного распределения при использовании критерия Колмогорова.

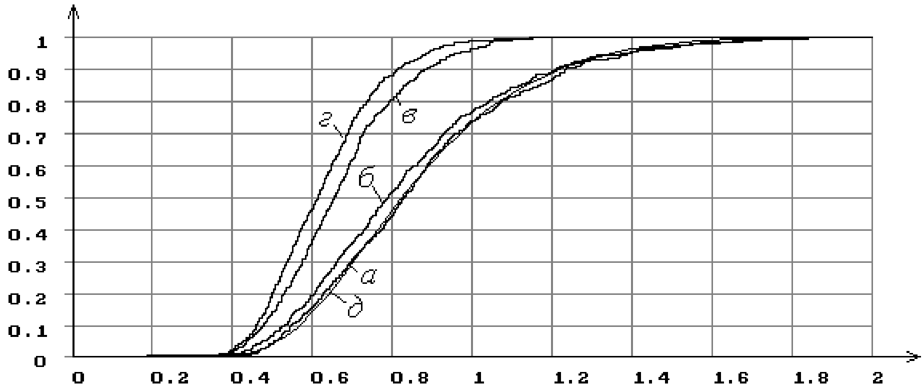

Рис.1. Эмпирические функции распределения статистики Колмогорова при оценивании параметров нормального закона

При идентификации типов предельных законов распределения непараметрических статистик в зависимости от вида закона наблюдаемой случайной величины и количества оцениваемых по наблюденной выборке параметров использовалось множество законов и семейств распределений, включенных в программную систему [2]. Оказалось, что почти всегда с достаточно высокой степенью точности эмпирические законы распределения статистик непараметрических критериев описываются одним из двух законов распределения: *логарифмически нормальным* или *гамма-распределением*.

В табл. 1 приведена часть результатов идентификации законов распределения только статистики критерия Колмогорова. Указание в соответствующей клетке на конкретное распределение означает, что выборка соответствующей статистики хорошо описывается данным законом (согласуется с законом). Если в клетке таблицы содержится указание более, чем на один закон, то на первом месте стоит распределение, согласие с которым наилучшее. В таблицах через $\ln N(\mu, \sigma)$ обозначено логарифмически нормальное распределение с параметрами сдвига μ и масштаба σ, через $\gamma(\theta_0, \theta_1, \theta_2)$ - гамма-распределение с параметрами формы θ_0, масштаба θ_1 и сдвига θ_2 [2].

Таблица 1.

Предельные распределения статистики Колмогорова				
Распредел-е случайной величины	Параметры по выборке не оценивались	Оценивался только масштабный параметр	Оценивался только параметр сдвига	Оценивалось два параметра
Нормальное	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	lnN(-0.2245,0.3157) γ(3.1875,6.2605,0.3312)	lnN(-0.4248,0.2350) γ(3.5392,11.051,0.3520)	lnN(-0.4879,0.2235) γ(5.4165,16.024,0.2917)
Коши	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	γ(2.6218,5.7693,0.3318) lnN(-0.2953,0.3272)	γ(4.4365,11.101,0.3100) lnN(-0.3768,0.2583)	γ(3.7771,12.507,0.2909) lnN(-0.5523,0.2415)
Лапласа	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	γ(3.4375,6.5969,0.3056) lnN(-0.2416,0.3183)	γ(3.3433,9.3957,0.3532) lnN(-0.3757,0.2513)	γ(4.3667,12.585,0.2945) lnN(-0.4725,0.2393)
Экспоненциальное	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	lnN(-0.3324,0.2545)		
Полунормальное	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	lnN(-0.2956,0.2684)		
Рэлея	$K(x)$, lnN(-0.1808,0.3060) γ(2.7181,5.6522,0.3942)	lnN(-0.3324,0.2545)		

Аналогичные результаты получены для критериев Смирнова, ω^2 и Ω^2 Мизеса, которые в совокупности позволяют сделать следующие выводы:

1. 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 соответственно от законов $K(s)$, χ_2^2, $a1(s)$ и $a2(s)$, что последние ни в коем случае не должны использоваться в такой ситуации.
2. На предельные распределения всех непараметрических статистик наиболее значительное влияние оказывает оценивание параметра сдвига, в существенно меньшей степени - оценивание масштабного параметра.
3. Хорошая аппроксимация для реальных распределений статистик непараметрических критериев обычно может быть получена с использованием логарифмически нормального и/или гамма-распределения.

4. Для ряда законов распределения случайных величин идентифицированы законы распределения статистик непараметричеких критериев при различном количестве оцененных параметров, что позволит делать более надежные статистические выводы на практике.

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
2. Лемешко Б.Ю. Статистический анализ одномерных наблюдений случайных величин: Программная система. - Новосибирск: Изд-во НГТУ. - 1995. - 125 с.