Тр. международной конференции "Актуальные проблемы электронного приборостроения". Новосибирск, 1998. Т.3. - С. 12-16.

ИССЛЕДОВАНИЕ ДОПРЕДЕЛЬНЫХ РАСПРЕДЕЛЕНИЙ СТАТИСТИК КРИТЕРИЕВ СОГЛАСИЯ ПРИ ПРОВЕРКЕ СЛОЖНЫХ ГИПОТЕЗ

Лемешко Б.Ю., Постовалов С.Н.

Новосибирский государственный технический университет,
Новосибирск, Россия. E-mail: post@sun.nstu.nsk.su; headrd@ nstu.nsk.su

Аннотация. Методами статистического моделирования получены распределения статистик критериев согласия при оценивании параметров ряда распределений методами максимального правдоподобия, минимума статистики Колмогорова и минимума статистики ω^2 Мизеса. Исследуется влияние объема выборки и метода оценивания параметров распределений вероятностей на распределения статистик непараметрических критериев согласия.

1. Введение.

Известно, что непараметрические критерии (Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса) при оценивании параметров распределений вероятностей (в случае проверки сложных гипотез) теряют свойство “свободы от распределения”, т.е. предельные распределения этих статистик будут зависеть от распределения, которому подчинена выборка.

В литературе известны следующие подходы к использованию критериев согласия в этом случае.

- Если выборка достаточно большая, то ее можно разбить на две равные части: по одной оценивать параметры, а по другой проверять согласие. При больших объемах выборки такой подход оправдан [1]. Но если объем выборки невелик, то оценки параметров будут зависеть от способа разбиения выборки на две части, и, следовательно, результаты проверки согласия также будут неоднозначны.
- Для случая нормального распределения предельные распределения статистики критерия ω^2 Мизеса при оценивании одного или обоих параметров по методу максимального правдоподобия получены аналитически [2] и табулированы.
- В некоторых частных случаях проверки сложных гипотез таблицы процентных точек для предельных распределений непараметрических статистик получены методом статистического моделирования. Например при оценивании параметров распределений экспоненциального, нормального, экстремальных значений, Вейбулла и некоторых других законов [3-6].
- В работах [7-10] для статистик типа Колмогорова-Смирнова в некоторых частных случаях получены формулы для *приближенного* вычисления вероятностей согласия вида $P\{S > S^*\}$, где S^* - вычисленное по выборке значение соответствующей статистики S. Эти формулы дают достаточно хорошие приближения при малых значениях соответствующих вероятностей. С помощью таких формул вычисляется вероятность согласия в пакете STADIA [11].
- Нами в результате исследований методами статистического моделирования предельных законов распределения статистик непараметрических критериев найдены такие законы распределения вероятностей, которые хорошо аппроксимируют предельные распределения статистик непараметрических критериев согласия в случае проверки сложных гипотез, когда при оценивании по выборке параметров используется метод максимального правдоподобия [12, 13].

Тем не менее, полученные за 40 лет исследований таблицы процентных точек и предельные распределения пригодны лишь в относительно небольшом числе случаев. В самом деле, распределения статистик (или их процентные точки) получены для 10-15 законов, в то время как множество законов, используемых на практике в качестве вероятностных моделей реальных случайных величин, существенно шире. Кроме того, для многих исследователей очевиден факт, что распределения статистик существенно зависят от *объёма выборки*, а также от *метода оценивания параметров*.

2. Влияние объёма выборки на распределение статистики Колмогорова при проверке простых и сложных гипотез.

Распределение статистики

$$D_n = \sup_x \left| F(x) - F_n(x) \right|,$$

где $F_n(x)$ – эмпирическая функция распределения, $F(x)$ – теоретическая функция распределения, n – объём выборки, было получено Колмогоровым в [14]. Очевидно, что это распределение зависит от объёма выборки, но при $n \to \infty$ распределение статистики $\sqrt{n}D_n$ сходится равномерно к распределению Колмогорова.

На рис.1 показано как меняется распределение статистики Колмогорова при увеличении объёма выборки (*n=1, 2, 3, 5, 10*; *K(x)* – распределение Колмогорова). При малых n распределение существенно отличается от предельного, но при $n \geq 10$ ошибка при вычислении вероятности согласия не превосходит 2 %. Та же самая картина наблюдается в случае проверки сложных гипотез о согласии. Так, например, на рис. 2 показаны распределения статистики Колмогорова при оценивании параметра сдвига нормального распределения по методу максимального правдоподобия (*n=1, 2, 3, 5, 10, 1000*). Наибольшее отклонение от предельного распределения наблюдается на “хвостах”. Все распределения получены методом Монте-Карло в результате генерирования 5000 псевдослучайных выборок, подчиненных стандартному нормальному закону.

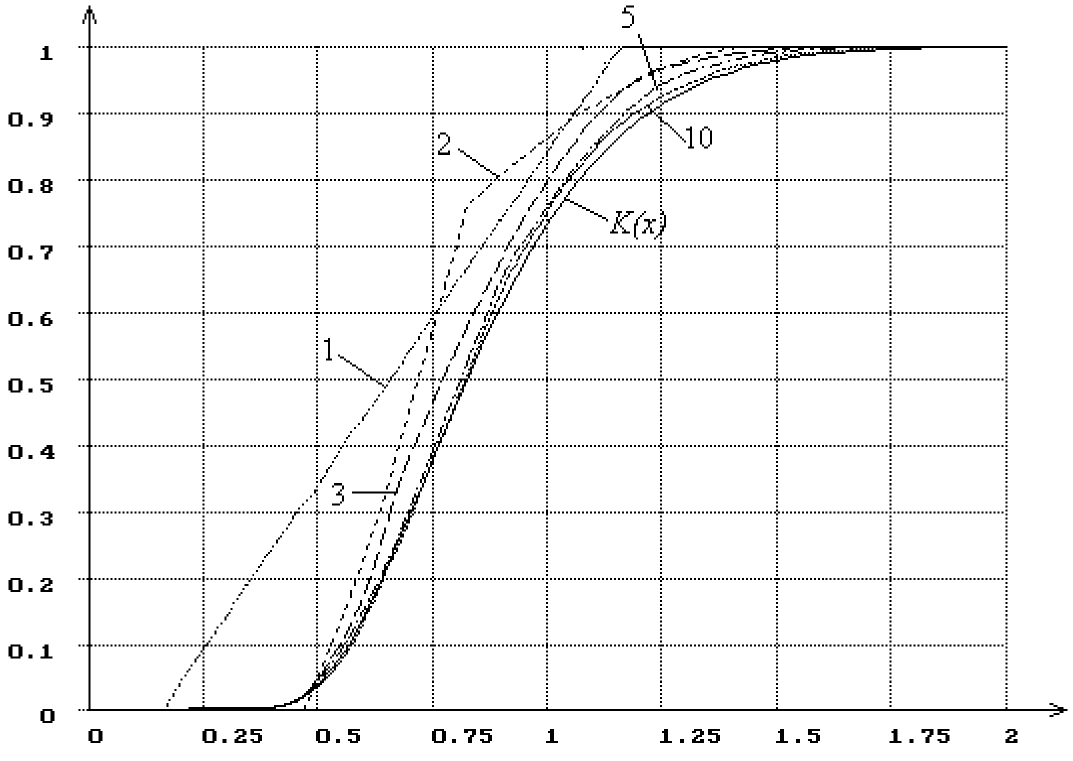

Рис. 1.

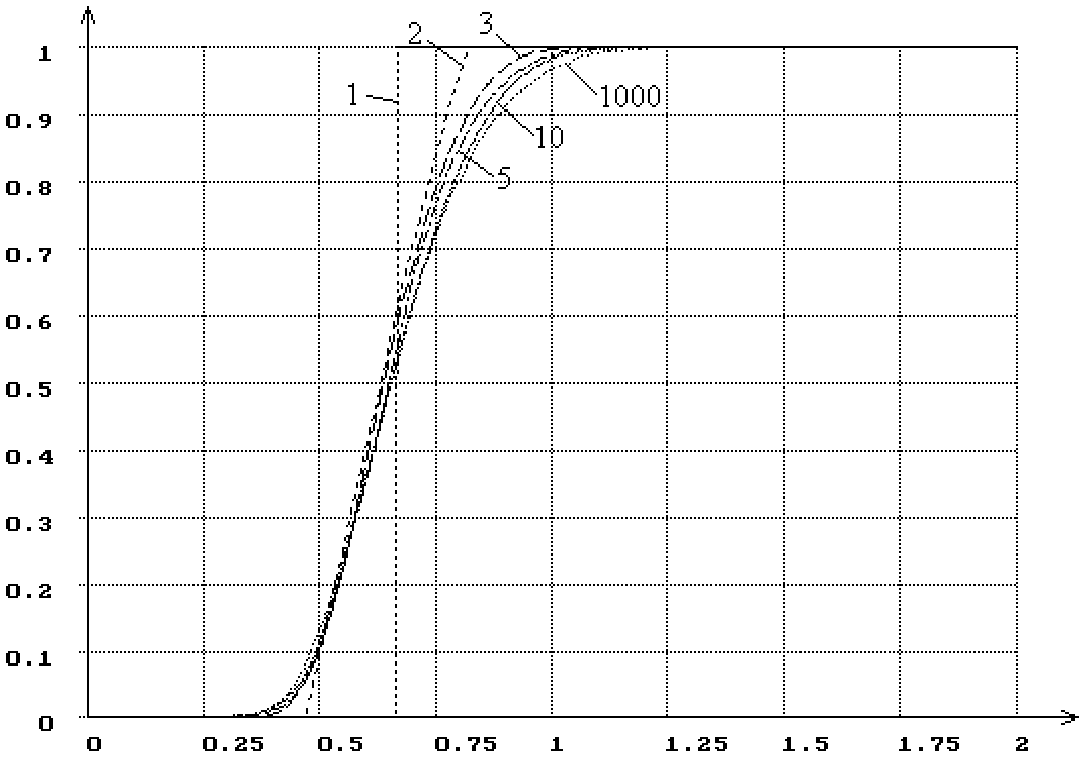

Рис. 2.

3. Влияние метода оценивания на распределения сатистик критериев согласия при проверке сложных гипотез

Распределения статистик непараметрических критериев согласия существенно зависит от метода оценивания параметров. Например, очевидно, что распределение статистики Колмогорова будет наиболее смещено влево в случае, когда оценки параметров находятся в результате минимизации этой статистики. На рис. 3 показаны распределения статистики Колмогорова при вычислении оценок параметра сдвига нормального распределения тремя различными методами: минимизацией статистики Колмогорова (график отмечен цифрой “1”), минимизацией статистики ω^2 Мизеса (“2”) и методом максимального правдоподобия (“3”). На рисунке *k(x)* – функция плотности распределения Колмогорова. Аналогичная ситуация характерна и для распределений других статистик.

4. Зависимость распределения статистик критериев согласия при проверке сложных гипотез от вида распределения

Хотя в общем случае при проверке сложных гипотез непараметрические критерии согласия теряют свойство “свободы от распределения”, тем не менее в ряде случаев распределения для разных законов оказываются одинаковыми. Немалую роль при этом играет метод оценивания параметров, что проиллюстрировано на рис. 4, где изображены распределения статистики Колмогорова при оценивании параметра сдвига нормального распределения (графики помечены цифрами “1” и “3”) и распределения Лапласа (“1” и “2”) по методу минимума статистики Колмогорова (“1”) и по методу максимального правдоподобия (“2” и “3”). Результаты моделирования показали, что при использовани метода максимального правдоподобия распределения статистики Колмогорова сильно отличаются друг от друга, а при вычислении оценок параметров минимизацией статистики Колмогорова – практически совпадают. Это является свидетельством того, что при проверке сложных гипотез более перспективно в качестве оценок выбирать те значения параметров, которые минимизируют статистику критерия.

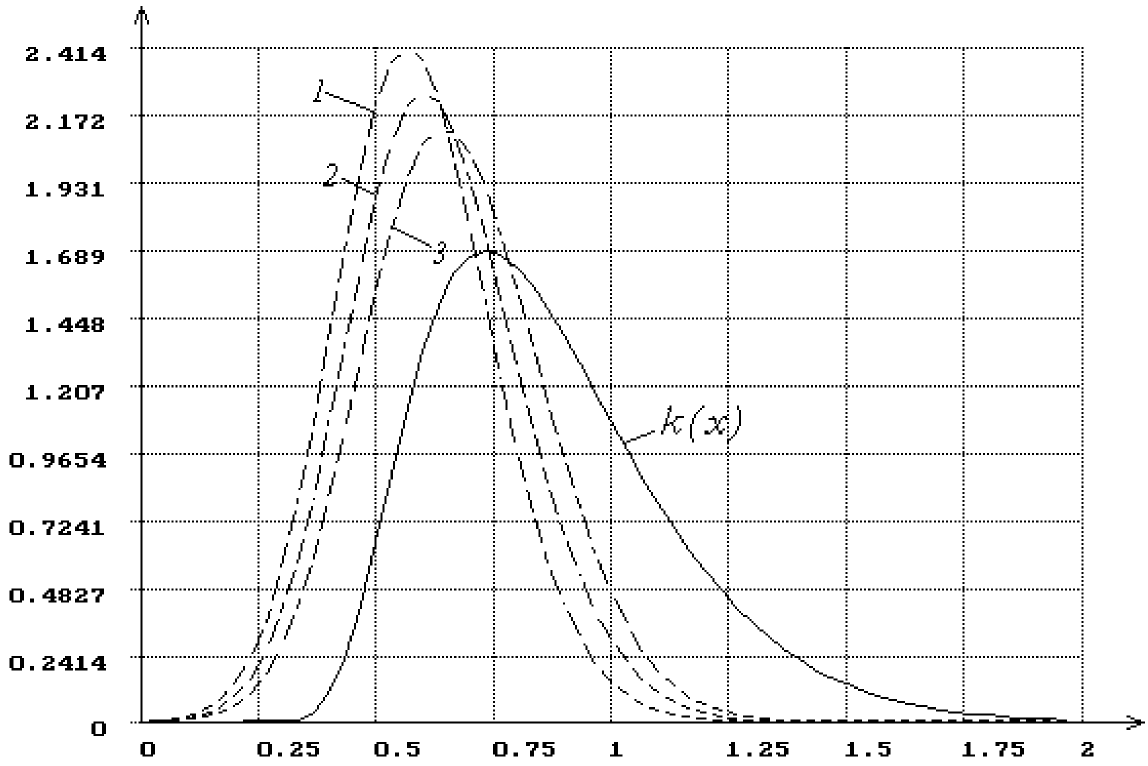
2.414
2.172
1.931
1.689
1.448
1.207
0.9654
0.7241
0.4827
0.2414
0
0
0.25
0.5
0.75
1
1.25
1.5
1.75
2
1
2
3
k(x)

Рис. 3.

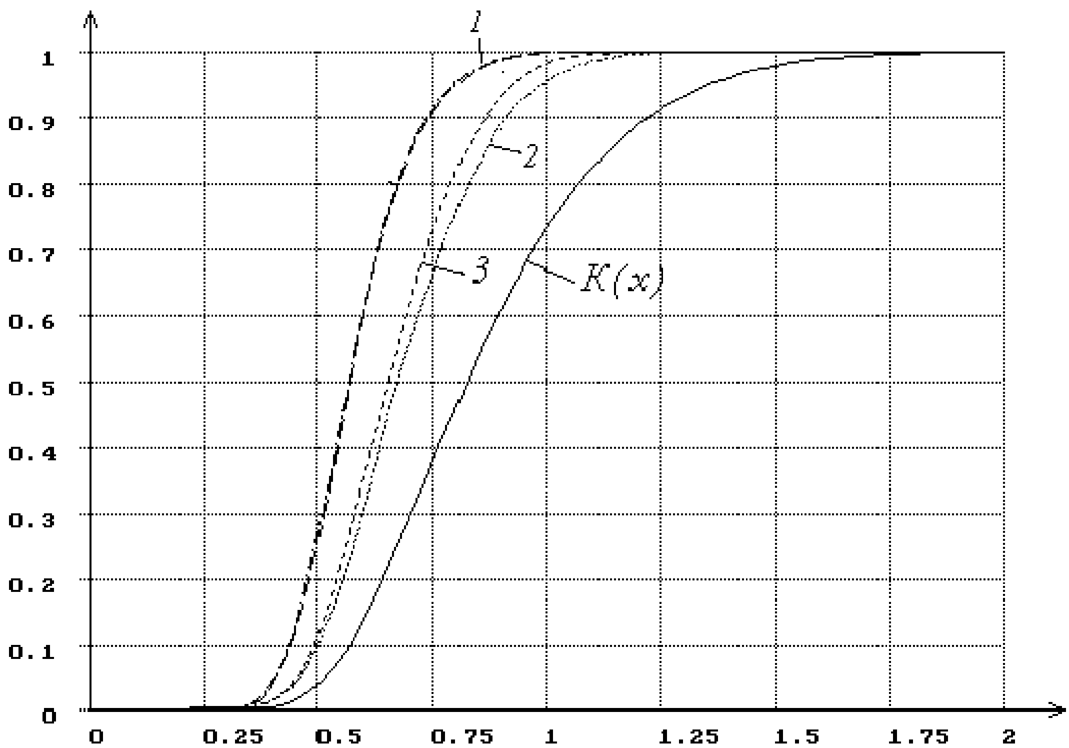
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.25
0.5
0.75
1
1.25
1.5
1.75
2
1
2
3
K(x)

Рис. 4.

5. Заключение

Таким образом, на основании результатов проведенных исследований можно утверждать следующее:

- Объем выборки достаточно слабо влияет на распределение статистики критерия согласия Колмогорова при $n > 10$, как при проверке простых, так и при проверке сложных гипотез.
- Существенное влияние на распределение статистики Колмогорова и статистик других непараметрических критериев согласия при проверке сложных гипотез оказывает выбранный метод оценивания.
- При использовании *MD*-оценок, минимизирующих статистику Колмогорова, распределения этой статистики при наблюдении различных законов распределений вероятностей отличаются существенно меньше, чем при использовании оценок максимального правдоподобия. Аналогично, при использовании других непараметрических критериев согласия при проверке сложных гипотез повидимому более перспективно в качестве оценок выбирать те значения параметров, которые минимизируют соответствующую статистику.

Литература

1. Durbin J. Kolmogoriv-Smirnov test when parameters are estimated // Lect. Notes Math. 1976. V. 566. P. 33-44.
2. Мартынов Г.В.Критерии омега-квадрат. – М.: Наука, 1978. – 80 с.
3. Pearson E.S., Hartley H.O. Biometrica tables for Statistics. V.2. – Cambridge: University Press, 1972. – 634 p.
4. Stephens M.A. Use of Kolmogorov-Smirnov, Cramer - von Mises and related statistics – vithout extensive table // J. R. Stat. Soc., 1970, B. 32. – P. 115-122.
5. Stephens M.A. EDF statistics for goodness of fit and some comparisons // J. Am. Statist. Assoc., 1974, v.69. – P. 730-737.
6. Chandra M., Singpurwalla N.D., Stephens M.A. Statistics for Test of Fit for the Extrem-Value and Weibull Distribution // J. Am. Statist. Assoc., 1981, v.76. – P. 375.
7. Тюрин Ю.Н. О предельном распределении статистик Колмогорова-Смирнова для сложной гипотезы // Изв. АН СССР. Сер. Матем., 1984, т. 48, № 6. – С. 1314-1343.
8. Тюрин Ю.Н., Саввушкина Н.Е. Критерии согласия для распределения Вейбулла-Гнеденко. // Изв. АН СССР. Сер. Техн. кибернетика, 1984, № 3. – С. 109-112.
9. Тюрин Ю.Н. Исследования по непараметрической статистике (непараметрические методы и линейная модель). Автореф. дисс. на соиск. учен. степени д-ра физ.-мат. наук. – М., 1985. - 33 с. – (МГУ).
10. Саввушкина Н.Е. Критерий Колмогорова-Смирнова для логистического и гамма-распределения // Сб. тр. ВНИИ систем. исслед. – 1990, № 8.
11. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. // М.: ИНФРА-М, Финансы и статистика, 1995. – 384 с.
12. Лемешко Б.Ю., Постовалов С.Н. Прикладные аспекты использования критериев согласия в случае проверки сложных гипотез // Надежность и контроль качества. – Москва, 1997. – № 11. – С. 3-17.
13. Лемешко Б.Ю., Постовалов С.Н. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов // Заводская лаборатория. – Москва, 1998. – № 3. – С. 61-72
14. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione. // G. Ist. Ital.attuar., 1933, vol. 4., № 1, p. 83-91.