Материалы VII международной конференции “Актуальные проблемы электронного приборостроения” АПЭП-2004. Новосибирск, 2004. – Т. 6. – С. 349-352.

ВОПРОСЫ ПРИМЕНЕНИЯ КРИТЕРИЕВ СОГЛАСИЯ В СЛУЧАЕ ИСПОЛЬЗОВАНИЯ НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК

А.В. Французов, Б.Ю. Лемешко

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

Аннотация. Методами статистического моделирования исследованы распределения статистик критериев согласия типа Колмогорова, типа ω^2 и Ω^2 Мизеса, χ^2 Пирсона при проверке адекватности непараметрических моделей законов распределений. Показано, что получаемые распределения статистик критериев согласия зависят от достаточно большого числа факторов и существенно отличаются от предельных законов, имеющих место в параметрическом случае.

Постановка задачи

В последнее время в качестве моделей законов распределения наблюдаемых случайных величин популярно использование различных непараметрических оценок. Это обусловлено тем, что на практике далеко не всегда располагают информацией о том, какая параметрическая модель закона предпочтительней для описания наблюдаемой случайной величины. В то же время, не смотря на многочисленность исследований, посвященных различным аспектам работы с непараметрическими оценками, практически не уделяется внимания задаче проверке адекватности построенных непараметрических моделей. Целью данной работы явилось исследование возможности проверки адекватности непараметрических моделей с применением различных критериев согласия, в том числе типа Колмогорова, типа ω^2 и Ω^2 Мизеса, типа χ^2 Пирсона.

Используемые непараметрические оценки плотности

В качестве непараметрических моделей в работе рассматриваются оценки плотности Розенблата–Парзена, имеющие вид [1]

$$p_n(x) = \frac{1}{n\lambda_n}\sum_{i=1}^{n}\varphi\left(\frac{x-x_i}{\lambda_n}\right), \tag{1}$$

где x_i, $i=1,\dots,n$ – выборка наблюдений одномерной непрерывной случайной величины (с.в.), λ_n – параметр размытости, а $\varphi(u)$ – колоколообразная (ядерная) функция, удовлетворяющая следующим условиям регулярности:

$$\varphi(u)=\varphi(-u)\,;\quad 0\le\varphi(u)\le\infty\,;\quad \int\varphi(u)du=1;\quad \int u^2\varphi(u)du=1;$$
$$\int u^m\varphi(u)du<\infty\,;\quad 0\le m<\infty\,. \tag{2}$$

В ходе настоящих исследований использовались ядерные функции двух видов:

1) квадратичная ядерная функция [2], обладающая наилучшими свойствами при минимизации среднеквадратичной ошибки аппроксимации

$$\varphi_1(u)=\begin{cases}\dfrac{3}{4\sqrt{5}}-\dfrac{3}{20\sqrt{5}}u^2, & \text{если } |u|\le\sqrt{5};\\ 0, & \text{если } |u|>\sqrt{5};\end{cases} \tag{3}$$

2) функция плотности стандартного нормального закона

$$\varphi_2(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,. \tag{4}$$

Распределения статистик непараметрических критериев согласия при непараметрическом оценивании

При проверке простых гипотез не имеет значения, каким образом задан рассматриваемый закон: в виде некоторого параметрического распределения $F(x,\theta)$ или в виде непараметрической оценки (1). И в том и в другом случае, предельными распределениями рассматриваемых статистик являются их классические распределения. Для статистики Колмогорова – это распределение Колмогорова $K(S)$, для статистики ω^2 Мизеса – распределение $a1(S)$, для статистики Ω^2 Мизеса – распределение $a2(S)$, для статистики χ^2 Пирсона – χ^2_{k-1}-распределения с $k-1$ степенями свободы. При проверке адекватности непараметрической модели (1) мы оказываемся в ситуации простой проверяемой гипотезы, если данная модель построена по некоторой другой выборке.

Как и в параметрическом случае, при проверке сложных гипотез распределения статистик критериев типа Колмогорова, типа ω^2 и Ω^2 Мизеса зависят от истинного закона распределения, соответствующего гипотезе H_0. Рис. 1 демонстрирует эту зависимость на примере статистики Колмогорова. На рисунке приведены условные распределения статистики типа Колмогорова $G(S_k|H_0)$ в случае различных законов наблюдаемых случайных величин (различных гипотезах H_0). В данном случае выборки с.в. объемом $n = 50$ моделировались по законам: экспоненциальному с плотностью $Exp(\mu,\sigma)=\frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\}$, логистическому – $\log(\mu,\sigma)=\frac{\pi}{\sigma\sqrt{3}}\exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\}\Big/\left[1+\exp\left\{-\frac{\pi(x-\mu)}{\sigma\sqrt{3}}\right\}\right]^2$, Коши – $Cauchy(\mu,\sigma)=\frac{\sigma}{\pi[\sigma^2+(x-\mu)^2]}$. Использовались ядерные функции (3). Для формирования выборки статистик эксперимент повторялся $N = 5000$ раз. Полученные результаты подтвердили, что при проверке сложных гипотез распределения статистик существенно зависят от истинного закона, соответствующего проверяемой гипотезе H_0. Подчеркнем, что аналогичные тенденции прослеживаются для статистик всех критериев.

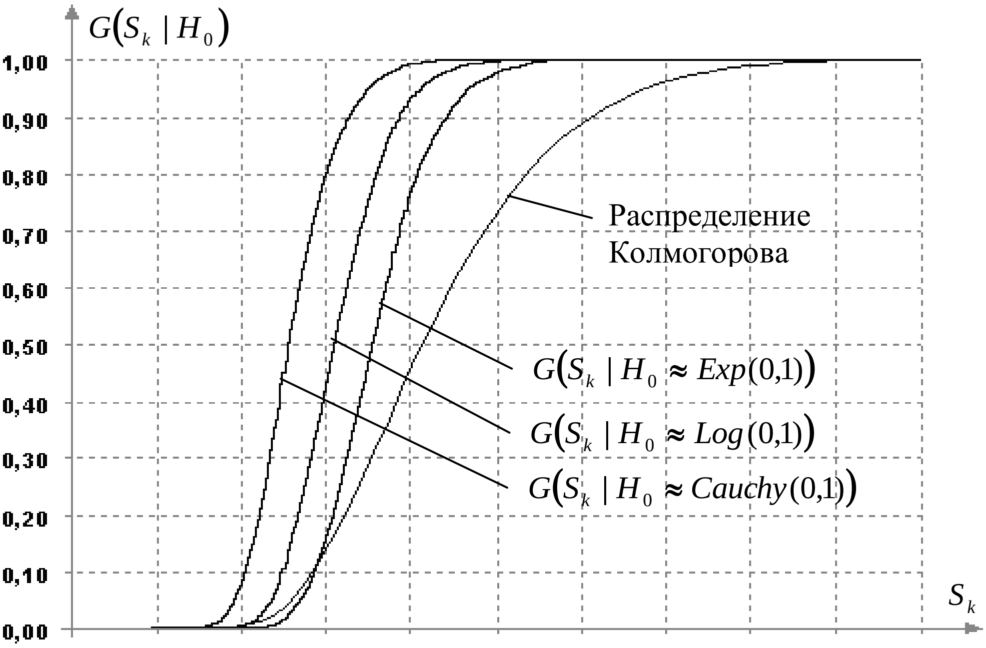

Рис. 1. Распределения статистики типа Колмогорова при сложной гипотезе в зависимости от истинного закона, соответствующего гипотезе H_0

Проведенные исследования показали, что для непараметрических моделей при проверке сложных гипотез прослеживается существенно большая зависимость от n, по сравнению с параметрическим случаем, и более медленная сходимость распределений статистик к предельным.

Результаты исследований показали, что тип используемых ядерных функций оказывает влияние на распределения статистик рассматриваемых критериев. Причем, с ростом объема выборки различие становится более существенным [3]. Кроме того, распределения статистик $G(S|H_0)$ зависят также и от метода оценивания параметра размытости λ_n.

Распределения статистик критериев согласия типа χ^2 Пирсона при непараметрическом оценивании

Проведенные исследования показали, что при проверке сложных гипотез распределения статистик критериев согласия типа χ^2 зависят от тех же факторов, что и распределения статистик непараметрических критериев, а также от способа группирования. При этом, получаемые распределения статистики существенно отличаются от предельных χ^2_{k-r-1}-распределений, которые имеют место при проверке согласия с параметрическими моделями и оценивании параметров по группированным данным.

Рис. 2 демонстрирует зависимость распределения статистики χ^2 Пирсона от объема выборки с.в., по которой строилась непараметрическая модель. Выборки моделировались по нормальному закону, использовалось равноинтервальное группирование с 5 интервалами и ядерные функции (3). Видно, что, начиная с выборок объемом 500 с.в., распределение статистики стремится к своему предельному закону. Было установлено, что в данном случае хорошей аппроксимацией распределения статистики является распределение Г(1.855, 1.362) с масштабом 1.089 и сдвигом 0.016.

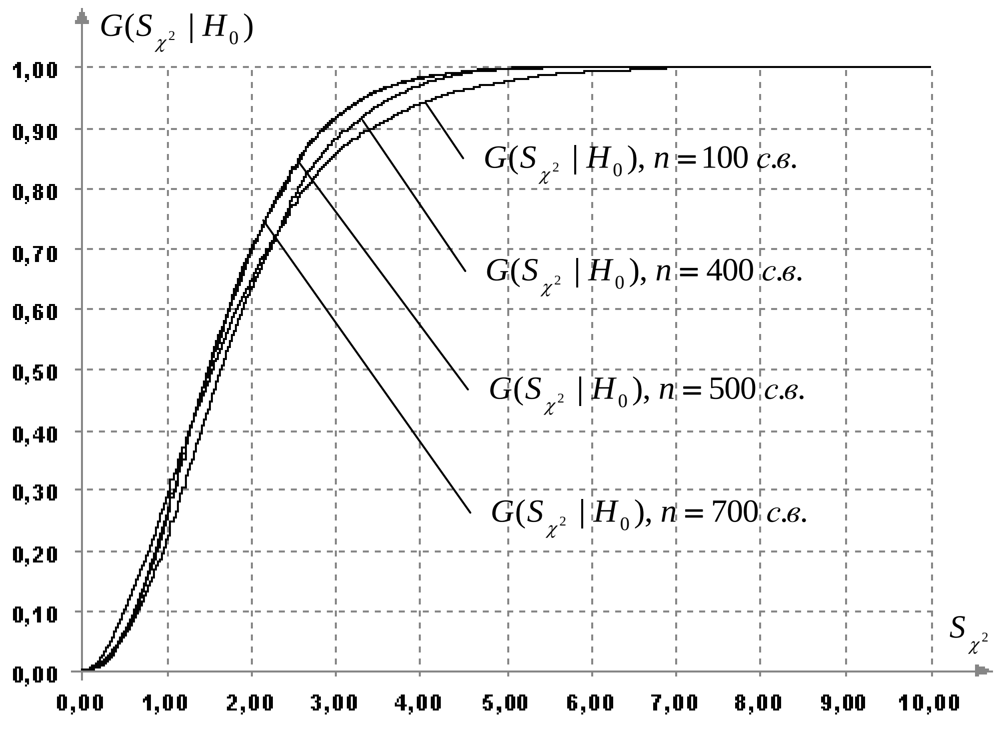

Рис. 2. Зависимость статистики χ^2 при проверке сложной гипотезы от объема выборки с.в.

Влияние способа группирования на распределение статистики χ^2 Пирсона отражено на рис. 3. Выборки объемом 100 элементов моделировались по нормальному закону, использовалось 5 интервалов группирования и ядерные функции (3).

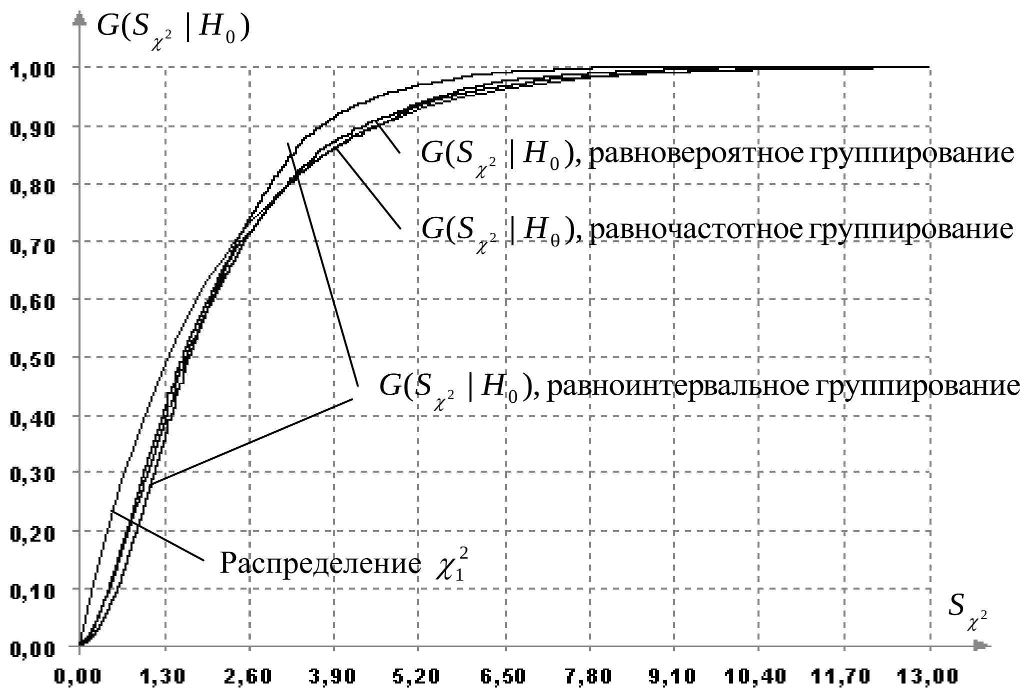

Рис. 3. Зависимость статистики χ^2 при проверке сложной гипотезы от типа группирования

Заключение

При проверке адекватности непараметрической модели мы имеем дело с простой проверяемой гипотезой только в том случае, если построение оценки и проверка согласия осуществляются по разным выборкам (или по разным частям одной выборки). В таких ситуациях процедура проверки должна опираться на классические предельные распределения статистик критериев независимо от вида наблюдаемого закона.

Проверка сложных гипотез о согласии при использовании непараметрических оценок (по сравнению с применением параметрических моделей) отличается еще большим многообразием факторов, определяющих “сложность” гипотезы (и вид распределения статистики). На распределения статистик критериев согласия существенно влияют: истинный закон распределения наблюдаемой случайной величины, соответствующий проверяемой гипотезе H_0; вид используемой ядерной функции; объем выборки; метод оценивания (вид оценки) параметра или параметров размытости. Распределения статистик критериев типа χ^2 Пирсона при заданном числе интервалов зависят также и от способа группирования. Причем распределения статистик достаточно медленно сходятся к некоторым предельным, которые зависят от вида истинного закона.

Проведенные исследования показали возможность использования критериев согласия для проверки адекватности непараметрических моделей законов распределений при проверке как простых, так и сложных гипотез. На основе компьютерного моделирования и анализа полученных эмпирических распределений можно строить модели предельных распределений статистик критериев согласия для различных проверяемых (сложных) гипотез.

Литература

1. Parzen E. On the estimation of probability density function and the mode // Ann. Math. Stat., 1962. – Vol. 33. – P.1065-1076.
2. Епанечников В.А. Непараметрическая оценка многомерной плотности вероятности. Теория вероятностей и ее применения, 1969. – Т.14. – № 1. – с. 156-161.
3. Лемешко Б.Ю., Постовалов С.Н., Французов А.В. К применению непараметрических критериев согласия для проверки адекватности непараметрических моделей // Автометрия. 2002. – № 2. – С. 3-15.