

О ВЛИЯНИИ СПОСОБОВ ГРУППИРОВАНИЯ НА МОЩНОСТЬ КРИТЕРИЕВ ТИПА χ^2 ПИРСОНА И ДЖАПАРИДЗЕ-НИКУЛИНА

Бушакова А.Д., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: kaniala@yandex.ru

В прикладных задачах статистического анализа в качестве моделей законов распределения вероятностей наблюдаемых случайных величин реально используется несколько десятков законов распределения. Естественно, что это не покрывает всего многообразия случайных величин. Вопрос о правомерности применения того или иного закона распределения в качестве модели для наблюдаемой величины принимается на основании критериев согласия.

Для проверки гипотез о согласии эмпирического распределения с теоретическим применяют ряд критериев типа χ^2 . В данной работе рассмотрены критерии χ^2 Пирсона и Джапаридзе-Никулина [1, 2].

Статистика критерия χ^2 Пирсона задается выражением:

$$X_n^2(\theta) = n \sum_{i=1}^k \frac{n_i/n - p_i(\theta)}{p_i(\theta)}, \quad (1)$$

где k - число интервалов группирования, n - количество наблюдений, n_i - количество наблюдений, попавших в i -й интервал, $p_i(\theta)$ - вероятность попадания в этот интервал.

Статистика модифицированного критерия типа χ^2 Джапаридзе-Никулина отличается от статистики критерия Пирсона при проверке сложных гипотез [1]:

$$U_n^2(\theta) = X_n^2(\theta) - nL^T(\theta)J^{-1}(\theta)L(\theta), \quad (2)$$

$$L(\theta) = (l_1(\theta), \dots, l_s(\theta)),$$

$$l_j(\theta) = \sum_{i=1}^k \frac{n_i}{np_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_j},$$

где m - число параметров закона, оцениваемых методом максимального правдоподобия по негруппированным данным, $J \theta$ - информационная матрица Фишера.

При любом способе группирования происходит потеря информации об исходном законе распределения. Потеря информации отражается на мощностях критериев типа χ^2 . Поэтому необходимо минимизировать потери, связанные с группированием.

В случае скалярного параметра задача асимптотически оптимального группирования, связанная с минимизацией потерь в информации Фишера, решается однозначно [3]. В случае векторного параметра мы имеем дело с информационной матрицей. Тогда критерии минимизации потерь могут быть различными. В [3] решение задачи асимптотически оптимального группирования приводится для случая D-оптимальности, когда минимизировался определитель информационной матрицы по группированным данным.

Однако не очевидно, что в случае D -оптимального группирования мы имеем вариант, обеспечивающий максимальную мощность относительно близких конкурирующих гипотез [4]. Отсюда возникла необходимость, во-первых, решения задач асимптотически оптимального группирования, обеспечивающих оптимум другим функционалам от информационной матрицы Фишера. Во-вторых, необходимость исследования мощности критериев при полученных вариантах асимптотически оптимального группирования.

В этой связи рассмотрены две постановки задачи асимптотически оптимального группирования. В задачах A - и E - оптимального группирования максимизировался след и наименьшее собственное число матрицы Фишера по группированным данным, Решение для ряда законов распределения (нормального, логистического, Вейбулла, Коши, экстремальных значений) получено в виде, инвариантном относительно параметров закона.

Мощность критерия χ^2 Пирсона исследовалась для различных пар конкурирующих гипотез и критериев оптимальности группирования.

Исследование распределений статистики критерия Джапаридзе-Никулина при справедливости проверяемой гипотезы H_0 показало, что оно очень быстро сходится к предельному χ^2_{k-m-1} -распределению [1]. Например, при проверке гипотез относительно законов логистического и нормального смоделированные распределения статистики Джапаридзе-Никулина хорошо согласуются с предельным уже при $n = 200$.

Мощность критерия Джапаридзе-Никулина исследовалась при D -оптимальном и равновероятном группировании.

При малом числе интервалов критерий Джапаридзе-Никулина уступает по мощности критерию χ^2 Пирсона (при проверке сложной гипотезы), а с ростом числа интервалов имеет заметное преимущество. Поэтому критерий Джапаридзе-Никулина, целесообразно применять при большем количестве интервалов группирования. При равновероятном группировании критерий Джапаридзе-Никулина существенно уступает по мощности критерию Пирсона.

Для сравнения рассматривался еще один способ группирования, предложенный в [5], который может использоваться при заданной конкурирующей гипотезе H_1 . В случае интервалов Неймана-Пирсона [5] в качестве границ интервалов выбираются точки пересечения плотностей, соответствующих гипотезам H_0 и H_1 . Данное группирование рекомендуется к применению в случае, если нет возможности решать оптимизационную задачу.

Результаты исследований данной работы могут быть сформулированы следующим образом.

1. Решены задачи A - и E -оптимального группирования для ряда законов распределения. Построенные таблицы асимптотически оптимального группирования могут использоваться в задачах оценивания параметров по группированным данным и в критериях согласия.

2. Исследовано влияние способов группирования на мощность критерия χ^2 Пирсона и Джапаридзе-Никулина. Показано, что при A -оптимальном группи-

ровании мощность критерия χ^2 Пирсона несколько превосходит его же мощность в случае *D*-оптимального группирования. Применение *E*-оптимального группирования в критериях согласия неперспективно.

2. Исследована мощность критерия χ^2 Пирсона в случае применения интервалов Неймана-Пирсона относительно заданных пар конкурирующих гипотез. Показано, что применение интервалов Неймана-Пирсона не гарантирует максимума мощности при соответствующем числе интервалов. В то же время применение интервалов Неймана-Пирсона на практике является оправданным, когда нет возможности найти оптимальное решение.

3. Исследован критерий Джапаридзе-Никулина при различных парах конкурирующих гипотез. Проверено, насколько хорошо распределение статистики критерия при справедливости проверяемой гипотезы согласуется с предельным χ^2_{k-m-1} -распределением.

4. Исследована мощность критерия Джапаридзе-Никулина при различных способах группирования (*A*-, *D*-, *E*-оптимальном и равновероятном группировании). Показано, что с позиций мощности критерия, равновероятный способ группирования уступает всем остальным. При паре конкурирующих гипотез «логистическое - нормальное» при малом количестве интервалов предпочтительнее *A*-оптимальное группирование. При большем количестве интервалов только *D*-оптимальное группирование обеспечивает мощность, сравнимую с мощностью критерия χ^2 Пирсона. В случае пары конкурирующих гипотез «нормальное - логистическое» показано, что значения мощностей при *A*- и *D*-оптимальном группировании близки, но в случае *A*-оптимального они несколько выше.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 09-01-00056а), Федерального агентства по образованию в рамках АВЦП "Развитие научного потенциала высшей школы" и ФЦП "Научные и научно-педагогические кадры инновационной России"

Литература

1. Джапаридзе К.О., Никулин М.С. Об одном видоизменении стандартной статистики Пирсона // Теория вероятностей и ее применения, 1974. Т.19. № 4. – С.886-888.
2. Dzhaparidze K.O., Nikulin M.S. On the computation of chi-square-type statistics // Journal of Mathematical Sciences, 1995. – V.75. – № 5. – P.1910-1921.
3. Денисов В.И., Лемешко Б.Ю., Цой Е.Б. Оптимальное группирование, оценка параметров и планирование регрессионных экспериментов: Монография. В 2-х ч. – Новосибирск: Новосиб. гос. техн. ун-т, 1993. – 347 с.
4. Воинов В.Г. Об оптимальных свойствах критерия Рао-Робсон-Никулина // Заводская лаборатория. Диагностика материалов. 2006. – Т.72. – № 3. – С.65-70.
5. Greenwood P.E., Nikulin M.S. A Guide to Chi-Squared Testing. New York: John Wiley & Sons, Inc. 1996. – 280 p.