

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ВАРИАНТОВ АСИМПТОТИЧЕСКОЙ ОПТИМАЛЬНОСТИ ГРУППИРОВАНИЯ НА МОЩНОСТЬ КРИТЕРИЕВ СОГЛАСИЯ ТИПА χ^2

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

В настоящее время в прикладных задачах статистического анализа в качестве моделей законов распределения вероятностей реальных случайных величин используется несколько десятков параметрических моделей. Естественно, что это не покрывает все многообразие случайных величин. Вопрос о правомерности применения того или иного закона распределения в качестве модели для описания наблюдаемых случайных величин принимается на основании выводов с использованием критериев согласия.

Существует ряд критериев типа χ^2 для проверки гипотез о согласии эмпирического распределения с теоретическим. Мы рассмотрим критерий χ^2 Пирсона и критерий Джапаридзе-Никулина [1, 2].

Статистика критерия χ^2 Пирсона задается выражением:

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

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.

Критерий Джапаридзе-Никулина отличается от критерия Пирсона только при проверке сложных гипотез. Его статистика имеет вид [1]:

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

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

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

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

Естественно, что при любом группировании происходит потеря информации об исходном законе распределения. Это в свою очередь влияет на мощность 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] задача асимптотически оптимального группирования была решена для случая D -оптимального группирования, когда в качестве минимизируемого функционала рассматривался определитель информационной матрицы по группированным данным.

Однако далеко не очевидно, что в случае D -оптимального группирования мы имеем наилучший вариант, обеспечивающий максимальную мощность при близких конкурирующих гипотезах [4].

Отсюда возникает 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.

Поэтому были рассмотрены две постановки задач асимптотически оптимального группирования. В задачах A - и E - оптимального группирования максимизировался след и наименьшее собственное число матрицы Фишера по группированным данным для ряда законов распределения (нормального, логистического, Вейбулла, Коши, экстремальных значений) 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 -распределению с $k-s-1$ степенями свободы [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 способе группирования критерий Джапаридзе-Никулина существенно уступает по мощности критерию χ^2 Пирсона.

Для сравнения рассматривался еще один способ группирования, предложенный в [5], который может использоваться при проверке гипотез при заданной конкурирующей гипотезе H_1 . В случае интервалов Неймана-Пирсона [5] в качестве границ интервалов выбираются точки пересечения соответствующих плотностей. Как правило, применение интервалов Неймана-Пирсона

обеспечивает более высокую мощность по сравнению с другими способами при том же числе интервалов. В некоторых случаях мощность оказывается максимальной. Однако решение соответствующей оптимизационной задачи чаще приводит к решению, не совпадающему с интервалами Неймана-Пирсона. Применение интервалов Неймана-Пирсона можно рекомендовать к применению в случае, если нет возможности решать оптимизационную задачу.

В данной работе были 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.

Исследовано влияние способов группирования на мощность критерия χ^2 Пирсона и Джапаридзе-Никулина. Показано, что при *A*-оптимальном группировании мощность критерия χ^2 Пирсона несколько превосходит мощность его же в случае *D*-оптимального группирования. Применение *E*-оптимального группирования в критериях согласия неперспективно.

Исследована мощность критерия Джапаридзе-Никулина в зависимости от числа интервалов. Отмечено, что при равновероятном группировании мощность критерия Джапаридзе-Никулина существенно ниже, чем при асимптотически *A*- и *D*-оптимальном группировании.

Работа выполнена при частичной финансовой поддержке РФФИ (проект № 06-01-0059а) и Федерального агентства по образованию Минобрнауки РФ в рамках Аналитической ведомственной целевой программы "Развитие научного потенциала высшей школы" (проект № 2.1.2/3970).

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.