

Применение робастных статистических методов для обработки результатов измерений

Новиков Ю.И., Рыбаков К.А.

МАИ, г. Москва

Цель работы состоит в исследовании возможностей применения метода складного ножа (*jackknife*) и бутстреп-метода (*bootstrap*) для обработки результатов измерений в сравнении с классическими подходами, не использующими вторичные выборки и вторичные статистики.

Основная идея метода складного ножа заключена в систематическом пересчете оцениваемого параметра распределения с учетом временного исключения одного или нескольких наблюдений из общей совокупности исходных данных. Путем перебора с исключением всех возможных комбинаций элементов выборки получается массив различных значений оцениваемого параметра (для укороченных выборок), на основе которых и решается вопрос о вычислении окончательной оценки.

Бутстреп основан на оценке некоторого параметра распределения путем вычисления этого параметра для новых выборок, порожденных из исходной. В случае, когда для имеющихся наблюдений можно предположить, что они независимы и одинаково распределены, процедура может быть реализована путем построения большого числа новых выборок (равного объема), каждая из которых получается методом случайной выборки с возвращением из исходного набора данных.

В работе получены программные реализации метода складного ножа и бутстрепа, позволяющие проводить статистический анализ входных данных и оценивать эффективность этих методов при работе с большим количеством выборок. Для обработки как исходных данных, так и массива значений оцениваемого параметра, полученных по порожденным выборкам, реализованы следующие процедуры: построение эмпирической функции распределения и гистограммы, нахождение выборочного среднего, моды, медианы и оценки Ходжеса–Лемана (*HL*-оценки).

Рассмотрены разные варианты применения метода складного ножа и бутстрепа для оценки математического ожидания и проведена апробация для выборок разных объемов, соответствующих различным законам распределения (нормальному, логарифмически нормальному, показательному и равномерному). По результатам вычислительного эксперимента была найдена оптимальная схема, при которой эти методы, как правило, превосходят в точности оценивания математического ожидания для любого из рассмотренных распределений: вычисление выборочных средних для порожденных выборок с последующей оценкой моды. Это позволяет с достаточной долей уверенности утверждать, что такая же ситуация будет прослеживаться и при работе с данными, закон распределения которых неизвестен.