

Отзыв
научного консультанта

о диссертации Постовалова Сергея Николаевича
“Применение компьютерного моделирования для расширения прикладных возможностей классических методов проверки статистических гипотез”,
представленной к защите на соискание ученой степени доктора технических наук
по специальности 05.13.17 – «Теоретические основы информатики»

Статистические методы анализа данных, методы прикладной математической статистики в силу объективных условий все шире используются в различных приложениях. Известно множество программных систем, предоставляющих необходимый инструментарий для решения задач статистического анализа, использование которых не снимает проблем формирования корректных статистических выводов, осуществляемых на основании обработки статистических данных в конкретном приложении.

Применяемые системы включают богатый арсенал классических методов прикладной математической статистики, однако область корректного применения этого арсенала в приложениях оказывается сильно ограниченной.

Во-первых, в задачах, возникающих на практике, предпосылки применения многих методов зачастую не выполняются. Например, как правило, не выполняются предположения о принадлежности ошибок измерений нормальному закону. Во-вторых, вследствие ограниченности объемов выборок использование асимптотических результатов может вносить существенные погрешности и приводить к формированию неверного вывода. В-третьих, форма регистрации наблюдений (группированные, цензурированные, многократно цензурированные, интервальные данные) не позволяет использовать классические методы оценивания и проверки статистических гипотез. В-четвертых, практика подкидывает множество постановок задач, на которые математический аппарат прикладной статистики пока не дает ответа (новых методов, критериев и т. п.), а в рамках аналитического подхода решение этих задач оказывается очень сложным или может просто отсутствовать.

Не приходится надеяться на то, что поставляемые практикой задачи статистического анализа данных в условиях различных нарушений стандартных предположений будут своевременно разрешаться за счет человеческого интеллекта и только в рамках аналитических методов. Выход видится в интенсивном и эффективном использовании компьютерных технологий.

В последние полтора десятка лет увеличилось количество примеров использования методов статистического моделирования (методов Монте-Карло) для исследования статистических и вероятностных закономерностей, например, для нахождения процентных точек (квантилей) распределений статистик некоторых критериев. Хотя наиболее часто методы статистического моделирования в области прикладной математической статистики используют для проверки аналитических выводов. Становится правилом хорошего тона, когда доказательство теоремы в области математической статистики подтверждается результатами имитационного моделирования. Однако эти разрозненные примеры еще не являются подтверждением системного подхода к развитию компьютерных технологий, направленных на исследование вероятностных закономерностей, на совершенство-

вание аппарата прикладной математической статистики. Парадоксально, но развитие (прикладной) математической статистики сдерживает предубеждение к численным результатам, получаемым в этой области.

За последние сто с небольшим лет предложена масса статистических методов и критериев, в обилии которых невозможно ориентироваться даже специалисту. Многие из них реализованы в программных системах статистического анализа. Во многих случаях для проверки гипотезы одного и того же вида предложен целый набор критериев. По существу, имеется набор инструментов, предназначенных для косвенного измерения одной и той же величины. И тут возникает проблема выбора: какой критерий предпочтительней? Какой критерий при заданной вероятности ошибки первого рода гарантирует меньшую вероятность ошибки второго рода? Можно ли вообще при таком объеме выборки различить эти две интересующие нас конкурирующие гипотезы? Можно ли при таких объемах выборок использовать асимптотические результаты для данного критерия? Как правило, получить ответ на эти вопросы можно только при использовании компьютерных технологий исследования. Только при использовании методов компьютерного моделирования «неожиданно» проявляются недостатки статистических критериев, ограничивающие область их корректного применения.

Отличительной особенностью исследований школы, к которой принадлежит соискатель, является развиваемый численный подход к исследованию закономерностей. В тех случаях, когда это оказывается невозможно или неоправданно сложно сделать с использованием аналитического аппарата, математические модели исследуемых закономерностей строятся на основе компьютерного моделирования. Это позволяет добиваться хороших результатов там, где этого не удастся достичь одними аналитическими методами. При таком подходе аппарат прикладной математической статистики находится в процессе постоянной эволюции: найденные фундаментальные закономерности (модели, их описывающие), совершенствующие аппарат, встраиваются в программное обеспечение, что, в свою очередь, расширяет его возможности для исследования вероятностных закономерностей.

С.Н. Постоваловым многое вложено в развитие технологий компьютерного моделирования, используемых при исследовании статистических закономерностей, в исследования свойств различных критериев проверки статистических гипотез, в сравнительный анализ мощности групп критериев, предназначенных для проверки гипотез определенного вида. Эти исследования способствовали пониманию возможностей статистических методов, совершенствованию аппарата прикладной статистики, расширению возможностей классического аппарата при его использовании в условиях нарушения классических предположений.

Основные результаты диссертационной работы Постовалова С.Н. связаны: с исследованием распределений статистик непараметрических критериев согласия при проверке сложных гипотез; с разработкой алгоритмов оценивания точных критических границ для последовательных критериев Вальда, Айвазяна и Лордена; с разработкой алгоритма оптимального планирования эксперимента для различения двух простых гипотез по критерию отношения правдоподобия; с исследованием распределений статистик и мощности критериев однородности Лемана-Розенблатта, Андерсона-Дарлинга-Петита и Багдонавичуса-Никулина; с исследованием распределений статистик и мощности критериев проверки многомерной нормальности; с исследованием некоторых критериев, используемых при проведе-

нии полногеномного анализа; с алгоритмами оптимального планирования экспериментов при проведении полногеномного анализа ассоциаций.

Проводимые автором диссертации исследования поддерживались Российским фондом фундаментальных исследований, Министерством образования и Министерством образования и науки Российской Федерации. Результаты исследований достаточно хорошо были представлены и оценены на различных, в том числе, международных конференциях.

Результаты исследований и разработанное программное обеспечение используются в курсе «Компьютерные технологии анализа данных и исследования статистических закономерностей». Этот курс в течение ряда последних лет читается на факультете прикладной математики и информатики НГТУ. Развиваемое математическое и программное обеспечение позволяет слушателям курса проводить самостоятельные исследования по различным направлениям.

В целом результаты диссертационных исследований Постовалова Сергея Николаевича можно оценить как крупное научное достижение, способствующее развитию технологий компьютерного моделирования при исследовании вероятностных и статистических закономерностей, способствующее расширению возможностей классического аппарата прикладной математической статистики, в том числе, в условиях нарушения стандартных предположений.

Научный консультант,
д.т.н., профессор, г.н.с. кафедры
прикладной математики НГТУ

 Б.Ю. Лемешко

Подпись профессора Лемешко Б.Ю. заверяю
начальник ОК НГТУ





О.К. Пустовалова