

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию на соискание ученой степени кандидата физико-математических наук Симушкина Дмитрия Сергеевича на тему «Статистические критерии с ограничениями на d -риски» по специальности 01.01.05 «Теория вероятностей и математическая статистика»

Диссертационная работа Симушкина Дмитрия Сергеевича посвящена задаче различения сложных гипотез в нетрадиционной (в отличие от подхода Неймана/Пирсона) постановке, в которой качество статистической процедуры оценивается условными вероятностями справедливости гипотезы/альтернативы при условии их отвержения (так называемыми d -рисками). Статистические выводы, получаемые в рамках такого « d -апостериорного» подхода, во многих практических ситуациях представляются более адекватными содержанию реальных решаемых проблем контроля качества, и, в то же время, порождает новые интересные математические задачи. Указанный подход к задачам различения сложных гипотез предложил выдающийся советский статистик Л.Н. Большев, назвавший соответствующие критерии гарантийными, а огромный вклад в развитие этого направления внес И.Н. Володин и его ученики. Рассматриваемая диссертация продолжает исследования в этом направлении, углубляя и расширяя имеющиеся результаты.

Диссертация объемом 135 страниц состоит из введения, двух глав и списка литературы, содержащего 70 источников.

Во введении дается краткая история возникновения d -апостериорного подхода, описываются возможности применения этого подхода к решению задач контроля качества. Особо подчеркивается связь этого подхода с методологией множественного тестирования, когда контролируемая характеристика процедуры тестирования совпадает с общепринятым показателем «Частота ложного обнаружения» (False Discovery Rate, FDR). Приведенный во введении аналитический обзор большого числа теоретических и прикладных работ, выполненных в мире по тематике диссертации уже в нашем веке, сви-

детельствует об актуальности тематики диссертации. В заключительной части Введения обосновываются постановки решаемых в диссертации задач и приводятся описания новых, полученных диссертантов, результатов.

Глава 1 содержит основные теоретические результаты, полученные автором. Исследуется задача различения двух сложных гипотез $H_0: \theta \leq \theta_0$ и $H_1: \theta > \theta_0$ и анализируются три, наиболее часто используемых на практике, модели правдоподобия и априорного распределения:

- «нормально-нормальная» модель, в которой наблюдения подчинены нормальному закону со средним значением θ , и предполагается, что в экспериментах значения θ получены как независимые реализации нормального закона;
- «Гамма-показательная» модель, в которой наблюдения имеют экспоненциальное распределение с параметром масштаба $1/\theta$, а в экспериментах значения θ получены как независимые реализации гамма-распределения;
- модель «бета-Бернулли», в которой требуется различить две сложные гипотезы о вероятности «успеха» θ с априорным бета-распределением.

В рамках этих моделей получены (Раздел 1.2) новые результаты относительно величины необходимого объема выборки (НОВ), обеспечивающего выполнение заданных ограничений (гарантий) на величину d -рисков. В Теореме 1.1 для «нормально-нормальной» модели и одинаковых значениях $\beta_0 = \beta_1$ ограничений на d -риски получены точные выражения для НОВ.

Далее изучается асимптотика НОВ при малых значениях β_0 и β_1 . При фиксированном отношении значений β_0 и β_1 , такая асимптотика получена в Теореме 1.3 (для «нормально-нормальной» модели) и в Теореме 1.4 (для «Гамма-показательной» модели). В диссертации приведены численные иллюстрации полученных формул, из которых следует, что точность предложенных приближений значительно выше точности асимптотических формул, полученных предшественниками (И.Н. Володиным и Ан.А. Новиковым).

В Разделе 1.3 исследуется универсальная последовательная гарантийная процедура, которая останавливает процесс наблюдений, если апостериорная вероятность одной из гипотез будет меньше выбранных ограничений. Вопрос замкнутости момента остановки такой процедуры (то есть, конечности момента остановки с вероятности единица) в общем случае не решен, однако в диссертации показано замкнутость момента остановки для рассматриваемых вероятностных моделей. При доказательстве этих формул использовалась методика, предложенная Л.Н. Большевым, опирающаяся на разложение функции распределения в ряд Эджворта. Кроме того, устанавливается, что в рамках нормально-нормальной модели математическое ожидание момента остановки универсальной процедуры конечно, если истинное значение параметра θ не попадает на границу θ_0 различаемых гипотез (теорема 1.5), и бесконечно, если значение θ совпадает с границей θ_0 (теорема 1.6).

В завершение главы 1 рассматривается задача построения эмпирических аналогов гарантийных процедур в духе идей Г. Роббинса. Здесь предполагается наличие некоего архива наблюдений, по которому можно найти оценки неизвестных (мешающих) параметров модели, и приводятся формулы для анализа их точности в зависимости от объема архива данных. В частности, для гамма-показательной модели предлагаются оценки, использующие моменты обратных величин наблюдений. Обсуждается возможность применения техники ядерных оценок априорной плотности. Показывается, что точность оценки функции d-риска пропорциональна точности (в смысле L_1) применяемой ядерной оценки.

Главу 2 можно отнести к области прикладной математической статистики. В первом параграфе этой главы рассматриваются задачи построения процедур статистического контроля качества с ограничениями на среднюю долю ошибочно принятых решений. Отмечается, что игнорирование априорного распределения в построении гарантийных процедур приводит не только к увеличению необходимого числа наблюдения, но и не позволяет контролировать именно те характеристики процедур, которые в наибольшей степени

соответствуют существованию решаемой статистической проблемы. Для наиболее употребительных в практике выборочного контроля качества и тестирования надежности вероятностных моделей строятся d -гарантийные критерии и приводятся таблицы для НОВ, гарантирующих ограничения на d -риски.

Второй параграф посвящен проблеме выделения генов с повышенной экспрессией. Предлагается вероятностная модель наблюдений, согласующаяся с данными. Для этой модели осуществляется построение критерия, удовлетворяющего заданным ограничениям либо на долю ложно отвергнутых нулевых гипотез, либо на долю ложно принятых нулевых гипотез, и при этом минимизируется доля ложных альтернативных решений. Показывается, что методы d -апостериорного подхода позволяют ставить и решать классические задачи на оптимум.

Предложен вариант вероятностной модели в задаче множественного тестирования, для которой найден вид оптимальной d -гарантийной процедуры (глава 2, §2, теорема 2.1). Дана общая схема построения процедур множественного сравнения, позволяющая контролировать не только долю ложно отвергнутых нулевых гипотез, а также долю ложно принятых нулевых гипотез (глава 2, §2). Кроме того, намечен путь решения проблемы множественного тестирования для задачи различения более двух гипотез (глава 2, §2).

Все полученные в диссертации результаты новые и их достоверность подтверждается приведенными доказательствами. Диссертационная работа завершается списком насущных нерешенных проблем, который позволяет надеяться на дальнейшее продвижение тематики диссертации.

Сделаем несколько замечаний.

1. Автор предлагает находит более точные асимптотические формулы НОВ только для двух вероятностных моделей. Причина, по которой это не сделано для модели «бета-Бернулли», осталась непонятной.

2. Метод доказательства бесконечности среднего значения момента остановки универсальной процедуры в рамках нормально-нормальной модели, по-видимому, может быть применен и для двух других моделей.

3. Хотелось бы видеть, как влияет точность оценок параметров априорного распределения на величину необходимого объема выборки. Например, можно было численно проанализировать изменение НОВ при изменении параметров в границах доверительных интервалов.

4. Во введении заявляется, что разрабатываемые в диссертации методы применительно к проблеме множественного тестирования позволяют контролировать не только «долю ложных открытий», но и долю ошибок среди принятых нулевых гипотез. Однако сравнение процедур в основном проводится снова по показателю «Частота ложного обнаружения» (FDR).

5. В Теоремах 1.5 и 1.6 исследуется случай, когда истинный параметр θ «отделен» от границы θ_0 различаемых гипотез или, наоборот, совпадает с этой границей. Интересно было бы исследовать случай «контигуальных» альтернатив, когда истинный параметр θ стремится «с правильной скоростью» к θ_0 с ростом объема выборки.

6. В диссертации имеются небольшое число опечаток. Например, на стр. 4 неправильно указано математическое ожидание гамма-распределения, на стр. 91 во втором абзаце пропущен предлог «на».

Однако указанные замечания не умаляют значимости диссертации, не снижают ее общей положительной оценки и во многом могут рассматриваться как предложения для будущих исследований.

По теме диссертации опубликовано восемь работ автора, три из которых изданы в журналах, входящих в базы данных Web of Science и Scopus. Результаты диссертации докладывались автором на трех международных конференциях. Диссертация соответствует паспорту специальности 01.01.05 – «Теория вероятностей и математическая статистика», автореферат правильно отражает содержание диссертационной работы.

В целом диссертация Д.С. Симушкина является законченным научно-квалификационным исследованием, выполненным под руководством д.ф.-м.н. И.Н. Володина, содержащим новые решения актуальной научной задачи построения статистических процедур в рамках так называемого d-апостериорного подхода, имеющим несомненный математический интерес для специалистов в области теории вероятностей и математической статистики и существенное значение для многих реальных прикладных задач. Диссертация по критериям актуальности, научной новизны, обоснованности и достоверности выводов удовлетворяет всем требованиям, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 01.01.05 – «Теория вероятностей и математическая статистика», а ее автор Симушкин Дмитрий Сергеевич заслуживает присуждения ему этой ученой степени.

Официальный оппонент:

профессор Сколковского института науки и технологий,
д.ф.-м.н. по специальности 01.01.05, профессор

Александр Владимирович Бернштейн

29 сентября 2020 года

Контактные данные:

телефон: +7 985 4162762;

e-mail: a.bernstein@skoltech.ru

Подпись Бернштейн А.В. подтверждено

МЕНЕДЖЕР
ПО ПЕРСОНАЛУ
ПОЧЕПЦОВ

