

## ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию ПАЛИОННОЙ Софьи Игоревны

«Асимптотические свойства оценок риска в задачах множественной проверки гипотез», представленную на соискание ученой степени кандидата физико-математических наук по специальности 1.1.4 – теория вероятностей и математическая статистика

### Актуальность темы исследования

Современный уровень развития техники и информационных технологий позволяет агрегировать удивительно большие объемы данных. Практика ставит перед теорией вопросы разработки инновационных методов поиска существенных влияющих переменных, форматов экономного представления больших объемов данных, удаления шума на фоне большого числа информативных факторов.

Со стороны математической статистики для решения указанных задач привлекает внимание исследователей понятие множественной проверки гипотез. В диссертации рассматриваемая задача множественной проверки гипотез сводится к задаче *пороговой обработки*. За последние десятилетия были предложены различные методы контроля за числом ошибок первого рода в случае нескольких нулевых гипотез. По-видимому, наиболее перспективным в настоящее время является метод учета *доли ложных срабатываний* (обозначается как FDR, False Discovery Rate). Применение «критических значений», найденных по методу FDR, в качестве порогового значения («FDR-порога») регулярно встречается, например, в современных работах по компьютерному зрению. В первую очередь это обусловлено хорошей статистической мощностью критериев на основе FDR для случая большого объема данных. Выбор именно этой характеристики диссертантом позволяет использовать результаты ряда ученых (Д. Донохо, И. Джонстон, Ф. Абрамович и др.), посвященные свойствам FDR-порога.

Рассматриваемые постановки задач бесспорно являются актуальными для широкого круга прикладных областей, в которых возникает задача сведения сложной модели с большим количеством факторов к упрощенному виду и без потери полезного сигнала. В списке литературы в

диссертации 30 % работ (20 из 65, исключая работы диссертанта) по пороговой обработке, появились за последние 20 лет. Таким образом, актуальность темы диссертации не вызывает сомнений.

### **Характеристика работы**

Во *Введении* автором изложена история рассматриваемого вопроса, а также обосновывается выбор направления исследования.

В *первой главе* кратко напоминаются различные методы решения проблемы множественных сравнений: метод Бонферрони, FWER, FDR, pFDR, FCR, NMP (все сокращения расшифрованы в тексте диссертации). Так как задача множественной проверки гипотез сводится к задаче пороговой обработки, то для сравнения рассматриваемых методов используется среднеквадратичная погрешность, или риск. В тексте диссертации автор подчеркивает невозможность вычисления на практике значения риска при использовании того или иного порога. В связи с этим используется оценка риска, которая в силу независимости от исходного сигнала может быть вычислена по наблюдаемым данным. Таким образом оценку риска можно использовать для сравнения различных порогов на практике. При рассмотрении задачи пороговой обработки автором приведены аргументы в пользу выбора FDR-меры в качестве порогового значения. Сделан обзор результатов, демонстрирующих положительные свойства FDR-порога в ситуации разреженного вектора исходных данных. Основные результаты первой главы содержатся в теоремах 1.5–1.7, доказывающих сильную состоятельность и асимптотическую нормальность оценки риска при использовании FDR-порога, а также в теоремах 1.8 и 1.9, в которых получены оценки скорости сходимости распределений оценки риска к нормальному распределению.

Во *второй главе* рассматривается постановка важной для приложений задачи, в которой исходный сигнал данных преобразовывается под действием линейного однородного оператора. Возникающая при этом задача обращения оператора по дискретным наблюдениям решается путем оценивания коэффициентов вейвлет-разложения искомой функции с использованием метода пороговой обработки. Именно здесь обращение к классам разреженности обретает смысл как выявление действительно значимых базовых функций в вейвлет-разложении. В теоремах 2.1

и 2.2 доказываются сильная состоятельность и асимптотическая нормальность оценки риска при использовании FDR-порога. В теореме 2.3 находится оценка скорости сходимости распределения оценки риска к нормальному закону.

В заключении диссертационной работы в соответствии с требованиями кратко перечислены полученные результаты, отмечены возможные направления для дальнейшей работы по данной теме.

Обе главы диссертации соединены общим методическим приемом исследования — исследования риска для пороговой обработки с FDR-порогом, что придает всей работе приятную целостность.

### **Характеристика результатов (новизна, достоверность и обоснованность)**

Изложенные в диссертационной работе результаты являются новыми, получены в ходе нетривиальных доказательств, демонстрирующих знания в области математического анализа, теории вероятностей и математической статистики.

Основные результаты диссертации были опубликованы в трех статьях Палионной С.И. без соавторов и двух статьях в соавторстве с научным руководителем Шестаковым О.В., которому принадлежат постановки задач. Все пять статей по теме диссертации опубликованы в журналах, индексируемых в базах Web of Sciences, Scopus, РИНЦ и входящих в список ВАК. Кроме того, основные результаты диссертации докладывались на четырех конференциях, одна из которых является международной, а тезисы к данным выступлениям приведены в сборниках трудов конференций.

Следует отметить высокий научный уровень диссертационной работы. Доказательства проведены тщательно, никаких серьезных ошибок в них не выявлено, что подтверждает достоверность полученных в диссертации результатов. Автореферат содержит все основные моменты диссертационной работы и в полной мере отражает ее суть.

### **Замечания по работе**

Тем не менее имеются общие замечания к диссертационной работе, которые хотелось бы отметить.

1) Теоремы 1.5, 1.6 и 1.7 в своих формулировках не уточняют, для

каких видов пороговой обработки они справедливы. И если в ходе *доказательства* теоремы 1.6 оговаривается, что будут приведены аргументы для мягкого порога, то в доказательстве теоремы 1.5 таких указаний нет. Возникает вопрос, насколько общим является утверждение этой теоремы, может ли оно быть перенесено (без изменений) и на другие, новые виды пороговой обработки.

2) Имеются некоторые неудачные определения. Например, определение символа  $\|\mu\|_0$  должно иметь вид  $|\{i: \mu_i \neq 0\}|$  вместо  $\#\{\mu_i: \mu_i \neq 0\}$ , так как, во-первых, если некоторые ненулевые  $\mu_i$  совпадут, то их нужно считать с учетом кратности, а не только их общее значение (по определению, элементы множества должны быть различными). Тут читателю приходится придерживаться словесного разъяснения обозначения скорее чем определения через формулу. Во-вторых, ранее на стр. 11 мощность конечного множества обозначалась вертикальными чертами, а не символом « $\#$ ». На стр. 16 и на стр. 54 при определении порога  $t_k$  через квантиль Гауссова распределения оказывается, что порог отрицательный, поскольку  $\alpha$  обычно выбирают близко к нулю. Но в таком случае, все положительные величины  $|X|_{(k)}$  окажутся больше отрицательных порогов  $t_k$ , так что  $k_F$  всегда будет равен  $n$ . По-видимому, речь должна идти о верхнем квантиле, соответствующем правому хвосту распределения.

3) В разделе 2.1 при записи основной модели в виде  $X_i = (Kf)_i + \varepsilon_i$  используется не введенное ранее обозначение  $(Kf)_i$ , то есть формального индекса  $i$  — «номера наблюдения» — вместо аргумента функции  $((Kf)(x) — это функция из  $L^2(\mathbb{R})$ ), что существенно усложняет понимание связи этой формулы с возможными приложениями, например, при анализе результатов компьютерной томографии, физики плазмы, для телетрафика.$

### **Заключение**

Указанные выше замечания не умаляют значимости диссертационного исследования. Диссертация отвечает требованиям, установленным Московским государственным университетом имени М.В. Ломоносова к работам подобного рода. Содержание диссертации соответствует специальности 1.1.4 — теория вероятностей и математическая статистика (по физико-математическим наукам), а также критериям, определенным пп.

2.1–2.5 Положения о присуждении ученых степеней в Московском государственном университете имени М.В. Ломоносова, а также оформлена согласно требованиям Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук Московского государственного университета имени М.В. Ломоносова. Таким образом, соискатель Палионная Софья Игоревна заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 1.1.4 – теория вероятностей и математическая статистика.

Официальный оппонент

заведующий кафедрой теории вероятностей и анализа данных федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского», доктор физико-математических наук (специальность 01.01.05 – теория вероятностей и математическая статистика), доцент по кафедре прикладной теории вероятностей

Зорин Андрей Владимирович

 /А. В. Зорин/

13.06.2023

Контактные данные:

тел.: +7 (920) 045-16-02, e-mail: andrei.zorine@itmm.unn.ru



Адрес места работы: 603022, г. Нижний Новгород, пр. Гагарина, 23, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского»,

Телефон: (831) 462-30-85, E-mail: unn@unn.ru