

## Заключение по содержанию диссертации

Райгородский Андрей Михайлович

(Ф.И.О. члена диссертационного совета)

Волков Никита Алексеевич

(Ф.И.О. соискателя ученой степени)

О некоторых свойствах вероятностных распределений

и их применении в задачах машинного обучения

(Название диссертации, ученая степень, на которую представлена диссертация, специальность)

представлена на соискание степени кандидата физико-математических наук по специальности 05.13.17 (теоретические основы информатики)

Дата защиты 15.12.20

Оценка соответствия диссертации требованиям Положения о присуждении ученых степеней кандидата наук, доктора наук в МФТИ (далее - Положение):

1. Актуальность тематики диссертации:

При разложении  $e^n$  в ряд Тэйлора частичная сумма ряда начинает превосходить половину всей суммы при суммировании более  $n$  членов. Это было замечено Рамануджаном, и им было инициировано изучение отклонения суммы первых  $n$  членов ряда от  $0.5$ . Этот вопрос имеет прямое отношение к изучению поведения вероятности  $P(X < b)$  для пуассоновской случайной величины с параметром  $b$ . Аналогичный вопрос о биномиальной случайной величине был изучен Джогдео и Самуэльсом, после чего на некоторые из поставленных ими вопросов были получены ответы Дмитриевым и Жуковским. Тем не менее общий случай (поведения упомянутой вероятности для биномиальной случайной величины с параметрами  $n$  и  $b/(n+c)$  для произвольного  $c$ ) оставался открытым. Именно этому случаю и посвящена первая глава диссертации. Вторая часть диссертации посвящена оцениванию параметров смеси многомерных распределений Стьюдента. Задача мотивирована некоторыми вопросами анализа данных, которые связаны наличием выбросов в данных. Предложенный метод позволяет учесть недостатки других известных методов и, как показал Н.А. Волков в третьей главе диссертации, может быть применен в задаче оценки качества проб пластовых флюидов.

2. Научная новизна выносимых на защиту результатов:

Все результаты диссертации являются новыми. Все теоретические результаты подтверждены строгими математическими доказательствами. Н.А. Волковым получено обобщение ранее известного свойства функции распределения  $P(X < b)$  биномиальной случайной величины  $X$  с параметрами  $n$  и  $b/n$ . Этот результат позволил доказать гипотезу, сформулированную Дмитриевым и Жуковским. Разработан новый метод оценивания параметров смеси многомерных распределений Стьюдента по выборкам с пропусками. Наконец, предложен и реализован новый метод машинного обучения, основанный на упомянутых оценках.

3. Теоретическая и практическая значимость диссертационной работы:

Первая и вторая глава диссертационной работа носят теоретический характер, результаты этих глав могут быть использованы для решения смежных задач теории вероятностей и математической статистики. Третья глава носит практический характер - ее результаты могут быть использованы для решения различных задач анализа данных (в том числе, как отмечено диссертации, для оценки качества проб пластовых флюидов).

4. Полнота опубликования основных результатов диссертации в рецензируемых научных изданиях в соответствии с требованиями Положения:

Результаты диссертации полностью опубликованы: 5 статей в рецензируемых журналах (все входят в перечень ВАК), из них три индексируются Scopus. Работа апробирована на нескольких семинарах и российских конференциях.

5. Вопросы и замечания (в соответствии с п. 4.13 Положения соискатель отвечает на сформулированные здесь вопросы и замечания на заседании по защите диссертации):

Возможно ли доказать гипотезу Самуэльса методами, предложенными в первой главе диссертации? Какие именно следствия в духе этой гипотезы можно получить из полученных результатов?

6. Общая характеристика диссертации (не включает резолютивную часть):

Во-первых, в диссертации обобщены ранее известные результаты о поведении функции  $F$  биномиального распределения с параметрами  $n$  и  $b/(n+c)$ . А именно доказано, что есть некоторое пороговое значение, при котором меняется монотонность функции  $F(b)$  и найдена асимптотика этого порогового значения при всех  $c$  в отрезке  $[0,1]$ . Во-вторых, в диссертации обобщены известные ранее методы оценки параметров смеси многомерного распределения Стюдента, что позволяет избежать их прежнего недостатка – невозможности использовать для данных, в которых есть пропуски и выбросы. Достоверность всех теоретических результатов, сформулированных в диссертации, подтверждается строгими математическими доказательствами. В-третьих, на основании упомянутых оценок, Н.А. Волков разработал метод машинного обучения для анализа данных с пропущенными значениями и успешно применил его для оценивания качества проб пластовых флюидов.

Дата \_\_\_\_\_

Подпись \_\_\_\_\_

/ Райгородский Андрей Михайлович /

