«УТВЕРЖДАЮ»
Директор Федерального государственного
бюджетного учреждения науки
Центрального экономико-математического
института РАН
академик В.Л. Макаров

____

«28» ноября 2016 г.

Отзыв ведущей организации

на диссертационную работу **Савелова Максима Павловича**
«Экстремальные характеристики критериев выбора статистических гипотез»,
представленную на соискание ученой степени
кандидата физико-математических наук по специальности
01.01.05 – теория вероятностей и математическая статистика

Задача о различении нескольких простых гипотез относится к классическим задачам математической статистики. В случае двух гипотез ее решение дается известной леммой Неймана–Пирсона, которая имеет огромное число разнообразных применений. Для трех и более гипотез проблема построения оптимального (в том или ином смысле) критерия и нахождения связи между вероятностями ошибок и характеристиками распределений (определяющих простые гипотезы) остается актуальной. Базовым результатом здесь является лемма Фано, дающая оценки вероятностей ошибок через расстояния Кульбака–Лейблера между распределениями. Более общие результаты в этом направлении были получены А.А. Гущиным в терминах *f*-дивергентных расстояний.

Диссертация М.П. Савелова посвящена изучению характеристик критериев качества в задаче различения трех и более гипотез для ситуации, когда известны только расстояния по вариации между гипотезами. Хотя такое расстояние является частным случаем общих *f*-дивергентных расстояний, в данной диссертации, в отличие от предыдущих работ, рассматриваются общие нерандомизированные критерии и приводится ряд результатов, позволяющих свести исходную (вообще говоря, бесконечномерную) задачу к более простой экстремальной задаче в конечномерном пространстве, а для некоторых линейных или кусочно-линейных критериев получить точные формулы или оценки.

Диссертация состоит из Введения, где приводится общая постановка задачи, обзор существующей литературы по данной тематике и кратко излагаются основные результаты работы, трех глав, краткого заключения и списка литературы из 24 наименований.

В первой главе изучаются характеристики качества нерандомизированных критериев выбора в задаче различения n простых гипотез в случае, если заданы только расстояния по вариации между гипотезами. Рассматривается набор мер $\vec{\mu}=(\mu_1,...,\mu_n)$, представляющих простые гипотезы, и (непересекающиеся) разбиения $\vec{C}=(C_1,...,C_n)$ пространства элементарных исходов на области принятия соответствующих гипотез. Качество критерия оценивается с помощью произвольной функции $f(\vec{\mu}(\vec{C}))$, заданной на всевозможных наборах $\vec{\mu}(\vec{C})=\left(\mu_j(C_k),\ j,k=1,...,n\right)$, характеризующих как вероятности принятия «правильных» гипотез, так и вероятности ошибок. Основная проблема, изучаемая в диссертации, состоит в нахождении границ изменения качества критерия f в условиях, когда известны только попарные расстояния по вариации между мерами из семейства гипотез $\vec{\mu}$. С этой целью автор вводит «экстремальные» величины $\overline{R}=\sup\limits_{\vec{\mu}}\sup\limits_{\vec{C}} f(\vec{\mu}(\vec{C}))$ и $\underline{R}=\inf\limits_{\vec{\mu}}\sup\limits_{\vec{C}} f(\vec{\mu}(\vec{C}))$, где внутренний супремум берется по всем непересекающимся разбиениям пространства элементарных исходов, а внешний экстремум — по всевозможным наборам мер с заданными попарными расстояниями по вариации.

Основное внимание в этой главе уделяется случаю трех гипотез. Доказывается, что нахождение величин $\overline{R}$ и $\underline{R}$ сводится к решению экстремальных задач на выпуклых замкнутых многогранниках, лежащих в пространствах достаточно большой размерности (54 и 18 соответственно) — теорема 4. Показано также, что размерность пространств можно снизить до 6, однако соответствующие множества, по которым берутся экстремумы, будут иметь более сложную структуру (теорема 5). Наряду с общим случаем, автор рассматривает три конкретные кусочно-линейные функции качества критерия f_+, f_Δ, $f_{\min}$, связанные с так называемой «проблемой деления торта», и выводит точные формулы для соответствующих верхних и нижних значений $\overline{R}$ и $\underline{R}$ (или только верхних значений $\overline{R}$) — теоремы 1–3.

Для произвольного числа гипотез $n\geq 3$ приводится результат о понижении размерности (теорема 9), аналогичный тому, что был для трех гипотез, а для конкретных функций качества вместо точных значений выписываются верхние и нижние оценки или неравенства, связывающие между собой соответствующие экстремальные значения $\overline{R}$ и $\underline{R}$ для разных функций (теоремы 7 и 8). В конце главы (раздел 1.6) автор останавливается на рандомизированных критериях и отмечает, что в случае, когда критерий качества представляет собой сумму вероятностей принятия «правильных» гипотез, рандомизация не увеличивает оптимальное значение критерия.

Во второй главе автор рассматривает обобщения экстремальных задач, изучавшихся в первой главе, на класс нерандомизированных «интервальных критериев». Под «интервальными» автор понимает такие критерии, когда наблюдение (результат эксперимента) может не противоречить не одной, а сразу нескольким гипотезам. Каждому такому критерию в случае трех гипотез соответствует разбиение пространства элементарных исходов на 7 множеств. Как и в первой главе, здесь вводятся экстремальные величины $\overline{Ri}$, $\underline{Ri}$, аналогичные $\overline{R}$, $\underline{R}$. Нахождение величин $\underline{Ri}$, $\overline{Ri}$ сведено к решению экстремальных задач на выпуклых замкнутых многогранниках (теоремы 11 и 12 соответственно). Кроме того, можно еще больше понизить размерность в задаче нахождения $\underline{Ri}$, но при этом множество, по которому берется экстремум, уже не будет многогранником (теорема 13). В конце главы рассматривается также случай «интервальных критериев» для произвольного числа простых гипотез и приводится результат о понижении размерности экстремальных задач нахождения величин $\overline{Ri}$, $\underline{Ri}$ (теорема 14).

Тематика третьей главы несколько отличается от первых двух. В ней рассматривается полиномиальная схема независимых испытаний и классическая статистика Пирсона, являющаяся основой известного критерия χ^2. Основным объектом изучения здесь является последовательность статистик Пирсона с увеличивающимся объемом выборки, а основным результатом — сходимость конечномерных распределений такой последовательности к конечномерным распределениям квадрата процесса Бесселя, соответствующим образом нормированного (теорема 16). Из этой сходимости и формулы для плотности многомерного χ^2–распределения (являющегося предельным для «кратной» статистики Пирсона) из работы Захарова, Сарманова, Севастьянова (1969) автор выводит явную формулу для конечномерных совместных плотностей процесса Бесселя (теорема 17). Стоит отметить, что, хотя процессам Бесселя посвящено огромное число работ, формула, касающаяся k-мерных распределений этих процессов (с $k \geq 2$), получена, по-видимому, впервые. Во всяком случае, в «обозримой» литературе ее аналогов найти не удалось.

В Заключении кратко сформулированы основные результаты, представленные в диссертации.

Написанная работа не лишена отдельных недостатков, носящих, в основном, редакционный характер. Среди них можно отметить следующие.

После сведения задачи нахождения величин $\overline{R}$, $\underline{R}$ к экстремальным задачам по выпуклым замкнутым многогранникам, для случая непрерывной кусочно-линейной целевой функции говорится о сведении к задаче линейного программирования, но корректнее было бы говорить о задаче кусочно-линейного программирования. Хотя такие задачи, как правило, сводятся к последовательности задач линейного программирования, но, с одной стороны, при соответствующем переходе иногда могут возникнуть особенности, а с другой стороны, сведение к линейным задачам не всегда является наиболее эффективным способом решения задач кусочно-линейного программирования.

Во Введении на стр. 6 вместо «кусочно-непрерывных функций» должно стоять «непрерывных кусочно-линейных функций», т.к. именно для таких функций рассматриваемые экстремальные задачи сводятся к кусочно-линейному программированию.

Восприятие теоремы 4 стало бы лучше, если бы было упомянуто, что величины $x_{\sigma}^{(j),k}$, стоящие в аргументе целевой функции, являются компонентами вектора $\vec{x} \in N_{18}$ (а не какими-то другими конструкциями), тем более что в другой формуле компоненты вектора $\vec{x} \in N_6$ обозначаются уже по-другому.

В теоремах о понижении размерности можно было бы явно указать, с какой размерности и по какую это снижение происходит. Если для случая трех мер все понятно, то для произвольного числа мер (раздел 1.5) это уже менее очевидно. В тексте говорится лишь об обобщениях «с естественными изменениями» (стр. 60), но точных формулировок не приводится.

По ходу доказательства теоремы 6 несколько раз говорится о задачах на «экстремум» (стр. 37, 38), хотя корректнее было бы говорить о задачах конкретно на минимум или на максимум.

Название раздела 1.4 «Экстремальные значения некоторых конкретных функций качества для произвольного числа мер» не совсем точно отражает его содержание, поскольку речь в нем идет не о точных формулах для экстремальных значений функций f_+, f_Δ, $f_{\min}$ (как это было в разделе 1.2 для трех мер), а лишь о неравенствах для этих экстремальных значений.

Термин «интервальные критерии», когда не отвергаться может не одна, а несколько гипотез, не кажется удачным. «Интервальность» обычно ассоциируется с «интервальными оценками» параметров и, вообще, с чем-то непрерывным. А в случае конечного числа простых гипотез более подходящим для этой ситуации представляется название типа «неоднозначные (множественные) критерии».

На стр. 78 для полиномиальной схемы число $\sum_{t=1}^{n} I_j(t)$ осуществлений исхода j в первых n испытаниях почему-то называется «частотой» исхода j, хотя под частотой обычно понимают отношение этой величины к n.

Ковариация случайных величин обозначается то «cov» (на стр. 79, 84), то «Cov» (на стр. 80). А на стр. 86 «Cov» обозначает уже ковариационную матрицу случайного вектора.

Читатель, который после формулировки теоремы 16 (стр. 80) захотел посмотреть ее доказательство (в разделе 3.4), сразу сталкивается с не определенными величинами $I_j^*(u)$, которые вводятся в другом разделе при доказательстве теоремы 15. Наверное, стоило бы повторить их определение (или хотя бы дать отсылку) и в разделе 3.4.

Немного странно читать, что носителем дискретного случайного вектора $\vec{\chi}(u)$ *является* гиперплоскость (стр. 85). Наверное, здесь речь должна идти о том, что носитель распределения *лежит* в гиперплоскости.

Работа написана в целом хорошим языком, хотя несколько раз встречаются повторы обозначений и определений, а также стилистические неточности.

Несмотря на высказанные замечания, общая оценка работы Савелова М.П. остается положительной. Автор продемонстрировал хорошее владение математической техникой, его результаты являются новыми и достоверными.

Работа носит теоретический характер, ее результаты могут быть полезны при исследовании задач различения гипотез и асимптотических методов математической статистики. Они могут быть использованы специалистами по математической статистике, работающими, в частности, в Московском государственном университете им. М.В. Ломоносова, Математическом институте им. В.А. Стеклова РАН, Центральном экономико-математическом институте РАН, Национальном исследовательском университете «Высшая школа экономики».

По результатам диссертации опубликованы 4 работы, в том числе 3 статьи в журналах, входящих в перечень ВАК. Основные результаты прошли апробацию на нескольких российских и международных конференциях. Автореферат полно и правильно отражает содержание диссертации.

Работа «Экстремальные характеристики критериев выбора статистических гипотез» является законченным научным исследованием и полностью удовлетворяет всем требованиям «Положения о порядке присуждения ученых степеней» Высшей аттестационной комиссии Министерства образования и науки Российской Федерации, а ее автор, Савелов Максим Павлович, заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 01.01.05 (теория вероятностей и математическая статистика).

Отзыв обсужден и одобрен на заседании научного семинара ЦЭМИ РАН «Вероятностные проблемы управления и стохастические модели в экономике, финансах и страховании» от 15.11.2016 (протокол № 10).

Ведущий научный сотрудник Федерального
государственного бюджетного учреждения науки
Центрального экономико-математического института РАН,
кандидат физико-математических наук
Сластников Александр Дмитриевич

117418 Москва, Нахимовский пр-т, 47,
тел.: 8(499)724-24-56,
электронная почта: slast@cemi.rssi.ru

Подпись А.Д. Сластникова заверяю
Ученый Секретарь ЦЭМИ РАН к.э.н. — Ставчиков А.И.

28.11.2016г.