Московский государственный университет
имени М. В. Ломоносова

На правах рукописи

ЗУБАЙРАЕВ Тимур Асламбекович

АСИМПТОТИЧЕСКИЙ АНАЛИЗ ВЫРОЖДЕННЫХ U-СТАТИСТИК ВТОРОГО ПОРЯДКА: ОЦЕНКИ ТОЧНОСТИ АППРОКСИМАЦИИ И ФУНКЦИЙ КОНЦЕНТРАЦИИ

Специальность 01.01.05 — теория вероятностей
и математическая статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата физико–математических наук

Москва - 2012

Работа выполнена на кафедре математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова.

Научный руководитель:	доктор физико-математических наук, профессор В.В. Ульянов
Официальные оппоненты:	доктор физико-математических наук, профессор А.Н. Тихомиров
	кандидат физико-математических наук, доцент А.Е. Кондратенко
Ведущая организация:	Институт математики им. С.Л. Соболева Сибирского отделения РАН

Защита диссертации состоится ________ 2012 г. в _____ часов на заседании диссертационного совета Д 501.001.44 в Московском государственном университете имени М.В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ, 2-й учебный корпус, факультет ВМиК, аудитория _____.

С диссертацией можно ознакомиться в библиотеке факультета ВМиК МГУ. С текстом автореферата можно ознакомиться на официальном сайте ВМиК МГУ http://cs.msu.ru в разделе "Наука"-"Работа диссертационных советов" - "Д 501.001.44"

Автореферат разослан " ___ " _____ 2012 г.

Ученый секретарь
диссертационного совета
профессор _______________ Н.П. Трифонов

Общая характеристика работы

Актуальность:

Одним из центральных объектов в теории вероятностей являются суммы независимых случайных величин. Интерес к задачам, связанным с суммированием независимых случайных величин, появился в математике еще в XVIII веке. Невозможность прямых вычислений распределений сумм независимых случайных величин приводит к необходимости получения и изучения асимптотических формул для них, то есть таких формул, которые позволяют находить с нужной точностью требующиеся нам вероятности, связанные с суммами случайных величин. Эти формулы даются предельными теоремами теории вероятностей. Таким образом, аппроксимация многократных сверток распределений потребовала развития содержательной математической теории, которая называется теорией предельных теорем для сумм независимых случайных величин.

Впервые исследования по U-статистикам были проведены в работах Халмоша [1], фон Мизеса [2] и Хефдинга [3] в конце 40-х годов XX века. U-статистики являются алгебраическим обобщением суммы независимых случайных величин и относится к классу симметрических функций. Интерес к таким случайным объектам возник сначала в математической статистике в задачах оценки функционалов от распределений. U-статистики, как объект вероятностной теории суммирования, с одной стороны "алгебраически" более сложны, чем суммы независимых случайных величин и векторов, с другой - содержат в себе существенные элементы зависимости, проявляющиеся в мартингальных свойствах (см. например монографию Королюк и Боровских [4]). Кроме того, U-статистики занимают одно

[1] *Halmos P. R.* The theory of unbiased estimation. Ann. Math. Statist.- 1946.- №. 17.- Pp. 34-43.

[2] *Mises R. von.* On asymptotic distribution of differentiable statistical functions.// Ann. Math. Statist.- 1947.- V. 18, № 2.- Pp. 309-348.

[3] *Hoeffding W.* A class of statistics with asymptotically normal distribution.// Ann. Math. Statist.- 1948.- №. 19.- Pp. 293-325.

[4] *Королюк В. С., Боровских В. Ю.* Теория U-статистик.- Киев: Наукова думка, 1989.

из центральных мест в статистических задачах.

С самого начала теория U-статистик развивалась под влиянием классической теории суммирования независимых случайных величин. Вместе с тем, асимптотическая теория U-статистик имеет специфические черты и качественно отличается от теории сумм независимых случайных величин. Многие асимптотические свойства U-статистик зависят от ранга r ядра, связанного со статистикой, и условий на моменты ядра. Для невырожденного ядра и ранга $r = 1$ асимптотическое поведение U-статистик сводится к асимптотическому поведению суммы случайных величин. Если $r \geq 2$ (ядро вырождено), то предельные распределения U-статистик представляют собой более широкий класс, включающий в себя бесконечно делимые распределения, распределения с функционалами, определенными ядром, конечно и бесконечномерными нормальными векторами. Такие особенности U-статистик создают новые идеи в контексте предельных теорем и требуют развития специальных методов исследования.

Историю развития оценок точности аппроксимации для U-статистик в случае бесконечномерного пространства можно разделить на два этапа. Первый этап связан с получением оценок, оптимальных по N - объему выборки, а второй по зависимости от характеристик оператора, ассоциированного со статистикой.

Настоящая работа посвящена асимптотическому анализу вырожденных U-статистик второго порядка.

Для вырожденных U-статистик второго порядка скорость сходимости к предельному распределению исследовалась в 1973 году Грэмсом и Серфлингом [5], которые получили оценку типа Берри-Эссеена $O(N^{-1/2+\varepsilon}), \varepsilon > 0$. Оценка улучшена в 1974 году Бикелем [6], а затем в 1977 году Чаном и Виерманом [7], Каллаертом, Янссе-

[5] *Grams W. F., Serfling R. J.* Convergence rates for U-statistics and related statistics. // Ann. Statist.- 1973.- no. 1.- Pp. 153-160.

[6] *Bickel P.* Edgeworth expansions in nonparametric statistics. // Ann. Statist.- 1974.- no. 2.- Pp. 1-20.

[7] *Chan Y.-K., Wierman J.* On the Berry-Esseen theorem for U-statistics // The Annals of Probability.- 1977.- Vol. 5, no. 1.- Pp. 136-139.

ном в 1978 году [8] и Хелмерсом и ван Цветом [9] в 1982 году, которые получили порядок сходимости $O(N^{-1/2})$. В 1979 году оценка $O(N^{-1+\varepsilon}), \varepsilon > 0$ доказана Гетце [10] с использованием неравенства симметризации Вейля при определенных условиях на моменты и предположении о том, что ненулевых собственных значений некоторого оператора, ассоциированного с U-статистикой, бесконечно много. Данный результат не является оптимальным ни по зависимости от N, ни по зависимости от характеристик оператора. Позже асимптотическими свойствами и получением разложения Эджворта для вырожденных U-статистик занимались Гетце [11], Королюк и Боровских. В монографии Королюк, Боровских [4] оценка улучшена до порядка $o(N^{-1/2})$. В работе Гетце и Зитикиса [12] получена аппроксимация со слагаемыми $O(N^{-1})$ при определенных предположениях о гладкости ядра и распределения элементов выборки. В 1999 году Гетце и Бенткус [13] доказали оценку точности аппроксимации для U-статистик, которая является оптимальной по N и получена при оптимальных моментных условиях. В той же работе построены оценки для функций концентрации U-статистик. Оба результата содержат константу, которая имеет экспоненциальный порядок зависимости от собственных значений некоторого оператора, ассоциированного с U-статистикой. Такая зависимость от собственных значений оператора, ассоциированного с U-статистикой может быть заметно улучшена, что и сделано в настоящей работе.

Исследования U-статистик ведутся и по другим направлениям: для хвостов распределений вырожденных U-статистик произволь-

---

[8] *Callaert H., Janssen P.* The Berry-Esseen theorem for U-statistics // The Annals of Statistics.- 1978.- Vol. 6, no. 2.- Pp. 417-421.

[9] *Helmers R., van Zwet W.* The Berry-Esseen bound for U-statistics //Statistical Decision Theory an Related Topics, III(S.S. Gupta and J.O. Berger, eds.).- 1982.- Vol. 1.- Pp. 497-512.

[10] *Götze F.* Asymptotic expansions for bivariate von Mises functionals //Z. Wahrsch. Verw. Gebiete.- 1979.- no. 50.- Pp. 333-355.

[11] *Götze F.* Expansions for von mises functions // Z. Wahrsch. Verw. Gebiete.0 1984.- no. 65.- Pp. 599-625.

[12] *Götze F., Zitikis R.* Edgeworth expansions and bootstrap for degenerate von Mises statistics. // Probab. Math. Statist.- 1995.- no. 15.- Pp. 327-351.

[13] *Bentkus V., Götze F.* Optimal bounds in non-Gaussian limit theorems for U-statistics // The Annals of Probability.- 1999.- no. 1.- Pp. 454-521.

ного порядка, построенных по выборкам из последовательности стационарно связанных наблюдений Борисовым И.С. и Володько Н.В. [14] получены экспоненциальные неравенства. Помимо этого, Борисов И.С. и Жечев В.А. доказали предельные теоремы для U-статистик от зависимых наблюдений [15], Борисов И.С. и Володько Н.В. доказали предельную теорему для U-процессов от стационарно связанных наблюдений [16]. Для статистик Мизеса - частного случая U-статистик, Борисов И.С. и Саханенко Л.А. [17] доказали центральную предельную теорему. Тихомировым А.Н., Гетце и Юрченко [18] получены асимптотические разложения в центральной предельной теореме для квадратичных форм.

Цель работы:

Целью данной диссертации является уточнение оценок точности аппроксимации и уточнение оценок функций концентрации для вырожденных U-статистик второго порядка.

Методика исследования:

Для оценки характеристической функции обобщенной U-статистики использована лемма симметризации, а также техника перехода к дискретным случайным величинам, которая ранее была предложена Юринским [19]. Характеристическая функция иссле-

---

[14] *Борисов И.С., Володько Н.В.* Экспоненциальные неравенства для распределений U- и V-статистик от зависимых наблюдений // Матем. тр.- 2008.- Т. 2, № 11.- С. 3-19.

[15] *Борисов И.С., Жечев В.А.* Функциональная предельная теорема для канонических U-процессов от зависимых наблюдений // Сиб. матем. журн.- 2011.- Т. 52, № 4.- С. 754-764.

[16] *Борисов И.С., Володько Н.В.* Ортогональные ряды и предельные теоремы для канонических U- и V-статистик от стационарно связанных наблюдений // Матем. тр.- 2008.- Т. 11, № 1.- С. 25-48.

[17] *Борисов И.С., Саханенко Л.А.* Центральная предельная теорема для обобщенных статистик Мизеса с вырожденными ядрами // Матем. тр.- 2001.- Т. 4, № 1.- С. 3-17.

[18] *Götze F., Tikhomirov A., Yurchenko V.* Asymptotic expansion in the central limit theorem for quadratic forms // Зап. научн. сем. ПОМИ.- 2007.- Vol. 341.- Pp. 81-114.

[19] *Yurinskii V.* On the accuracy of normal approximation of the probability of hitting a ball // Theory Probability Applications.- 1982.- no. 27.- Pp. 280-289.

дуемой статистики оценивается сверху характерестической функцией дискретной случайной величины. Данная техника также использовалась в работе Бенткуса и Гетце [13]. Для оценки точности аппроксимации используется усовершенствованный метод характеристических функций. Оценки для разностей характеристических функций U-статистики и предельного распределения получены с использованием неравенств типа Бергстрёма, которые применялись в работах Бенткуса и Гетце.

Научная новизна:

Все основные результаты диссертации являются новыми и состоят в следующем:

1. Для распределения вырожденной U-статистики второго порядка при оптимальных моментных условиях получена оценка точности аппроксимации коротким асимптотическим разложением, которая имеет степенной порядок зависимости от собственных значений оператора, ассоциированного с ядром U-статистики.

2. Для функций концентрации вырожденной U-статистики второго порядка получена оценка, которая имеет степенной порядок зависимости от собственных значений оператора, ассоциированного с ядром U-статистики.

Практическая значимость:

Результаты диссертации имеют теоретический характер и одновременно допускают применение к решению различных практических задач, связанных с использованием оценок точности аппроксимации распределений статистик.

Апробация работы:

Результаты работы неоднократно докладывались и обсуждались на научно-исследовательском семинаре кафедры математической статистики факультета ВМК МГУ (2009, 2010, 2011 гг.), X

Всероссийском симпозиуме по прикладной и промышленной математике (осенняя открытая сессия) (Сочи - Дагомыс, 1 - 8 октября 2009), XVIII Международной научной конференции студентов, аспирантов и молодых ученых "Ломоносов-2011"(11-15 апреля 2011), заседании кафедры математической статистики факультета ВМК МГУ имени М.В.Ломоносова, г. Москва, 14 сентября 2011 г.

Публикации:

Материалы диссертации опубликованы в 4 печатных работах, из них 2 статьи опубликованы в журналах, включенных в перечень ВАК.

Структура и объем диссертации:

Диссертация состоит из введения, трех глав, разбитых на разделы, и списка литературы из 44 наименований. Общий объем работы составляет 83 страницы.

Краткое содержание диссертации

Введение

Введение содержит обоснование актуальности темы диссертации и исторический обзор, связанный с темой работы. Кроме этого, в нем также формулируются и обсуждаются основные результаты работы. Доказательство этих результатов содержится в главах 1-3, поэтому здесь мы ограничимся основными определениями и обозначениями.

Пусть $X, \bar{X}, X_1, ..., X_N$ - независимые одинаково распределенные случайные величины, принимающие значения в произвольном измеримом пространстве $(\mathfrak{X}, \mathfrak{B})$. Пусть φ_1: $\mathfrak{X} \to \mathbb{R}$ и φ: $\mathfrak{X}^2 \to \mathbb{R}$ - измеримые функции. Предположим, что φ симметрична, т.е. $\varphi(x, y) = \varphi(y, x)$, для любых x, $y \in \mathfrak{X}$. Рассмотрим U-статистику:

$$T = \frac{1}{N} \sum_{1 \le i < j \le N} \varphi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \varphi_1(X_i), \qquad (1.1)$$

предполагая, что

$$\mathbb{E}\varphi_1(X) = 0, \mathbb{E}\varphi(x, X) = 0, \text{для всех } x \in \mathfrak{X},$$

$$\mathbb{E}\varphi^2(x,X) < \infty, \mathbb{E}\varphi_1^2(X) < \infty.$$

Одной из задач математической статистики является аппроксимация функций распределения статистик, имеющих сложный вид, некоторой более простой функцией распределения. В диссертационной работе эта задача решается для функции распределения U-статистики T, которую обозначим через $F(x)$. Для $F(x)$ мы будем использовать приближение, так называемым, коротким разложением. Это разложение включает в себя как предельную функцию $F_0(x)$, так и первый член асимптотического разложения $F_1(x)$ порядка $O(1/\sqrt{N})$. В работе получена оценка сверху для величины

$$\Delta_N = \sup_x |F(x) - F_0(x) - F_1(x)|,$$

где $F_1(x)$ - поправка Эджворта, определенная в главе 3. Заметим, что $F_1 = 0$ если $\varphi_1 = 0$ или для всех $x \in \mathfrak{X}$ справедливы равенства:

$$\mathbb{E}\varphi_1^3(X) = \mathbb{E}\varphi_1^2(X)\varphi(X,x) = \mathbb{E}\varphi_1(X)\varphi^2(X,x) = \mathbb{E}\varphi^3(X,x) = 0. \tag{1.2}$$

Рассмотрим измеримое пространство $(\mathfrak{X}, \mathfrak{B}, \mu)$ с мерой $\mu = \mathcal{L}(X)$ - распределением X. Пусть $L^2 = L^2(\mathfrak{X}, \mathfrak{B}, \mu)$ - Гильбертово пространство интегрируемых в квадрате вещественных функций. Оператор Гильберта-Шмидта $\mathbb{Q}$: $L^2 \to L^2$ определяется следующим образом

$$\mathbb{Q}f(x) = \int_{\mathfrak{X}} \varphi(x,y) f(y)\mu(dy) = \mathbb{E}\varphi(x,X)f(X). \tag{1.3}$$

Пусть $\{e_j : j \geq 1\}$ - полная ортонормированная система, составленная из собственных функций оператора $\mathbb{Q}$, соотвествующих собственным значениям $q_1, q_2 \ldots$ таким, что $|q_1| \geq |q_2| \geq \ldots$. Тогда

$$\sigma^2 = \mathbb{E}\varphi^2(\bar{X}, X) = \sum_{j\geq 1} q_j^2 \ , \ \varphi(x,y) = \sum_{j \geq 1} q_j e_j(x) e_j(y). \tag{1.4}$$

Так как $\mathbb{Q}$ оператор Гильберта-Шмидта и ядро φ вырождено ряд (1.4) сходится в $L^2(\mathfrak{X}, \mathfrak{B}, \mu \times \mu)$. Рассмотрим подпространство $L^2(\varphi, \varphi_1) \subset L^2$, образованное φ_1 и собственными функциями e_j, соответствующим ненулевым собственным значениям q_j. Добавляя собственную функцию e_0, $\mathbb{Q}e_0 = 0$,мы можем предположить, что функции $e_0, e_1, \ldots$ образуют ортонормированный базис в $L^2\,(\varphi, \varphi_1)$. Таким образом, справедливо следующее разложение

$$\varphi_1(X) = \sum_{j\geq 0} a_j e_j(x) \text{ в } L^2, \beta = \mathbb{E}\varphi_1^2(X) = \sum_{j\geq 0} a_j^2, \tag{1.5}$$

где $a_j = \mathbb{E}\varphi_1(X)e_j(X)$. Легко видеть, что $\mathbb{E}e_j(X) = 0$, для любого j. Следовательно, система $(e_j(X))_{j\geq 0}$ является ортонормированной системой случайных элементов с нулевым средним.

Гильбертово пространство $\ell_2 \subset \mathbb{R}^\infty$ состоит из элементов $x = (x_1, x_2 \ldots) \in \mathbb{R}^\infty$ таких, что

$$\mid x \mid^2 =_{def} \langle x, x\rangle, \mid x \mid < \infty, \langle x, y\rangle = \sum_{j\geq 0} x_j y_j.$$

Рассмотрим случайный вектор

$$\mathbf{X} =_{def} (e_0(X), e_1(X), e_2(X), \ldots),$$

принимающий значения в $\mathbb{R}^\infty$. Поскольку $(e_j(X))_{j\geq 0}$ система некоррелированных случайных величин с единичными дисперсиями, случайный вектор $\mathbf{X}$ имеет единичную матрицу ковариаций и нулевое среднее. В силу (1.4) и (1.5) справедливы равенства

$$\varphi(X, \bar{X}) = \langle \mathbb{Q}\mathbf{X}, \bar{\mathbf{X}}\rangle,\ \varphi_1(X) = \langle a, \mathbf{X}\rangle, \tag{1.6}$$

где $\mathbb{Q}x = (0, q_1x_1, q_2x_2, \ldots)$, для $x \in \mathbb{R}^\infty$ и $a = (a_j)_{j\geq 0} \in \mathbb{R}^\infty$. Равенства в (1.6) позволяют взять нам в качестве измеримого пространства $\mathfrak{X}$ простанство $\mathbb{R}^\infty$. Случайная переменная X - случайный вектор, принимающий значения в $\mathbb{R}^\infty$ с нулевым средним и единичной матрицей ковариаций и такой, что

$$\varphi(X, \bar{X}) = \langle \mathbb{Q}X, \bar{X}\rangle,\ \varphi_1(X) = \langle a, X\rangle.$$

Не ограничивая общности рассуждений, будем считать, что $\varphi(x, y)$ и $\varphi_1(x)$ - линейные функции по каждому из своих аргументов. Такое предположение значительно упрощает выкладки.

Введем ряд обозначений, которые мы будем использовать далее. Через $e(x)$ мы обозначим $exp\{ix\}$, т.е. $e(x) = \exp\{ix\}$. Будем писать $A \ll B$ или $A \ll_s B$, если $A \leq cB$ или $A \leq c_s B$ соответственно. $A \asymp B$ или $A \asymp_s B$ означает, что $A \ll B \ll A$ или $A \ll_s B \ll_s A$ соответственно.

Для матрицы $\mathbb{A} = (a_{ij}), \quad 1 \leq i, j \leq s$ (или соответствующего линейного оператора) через $|\mathbb{A}| = \sup_{|x|=1} |\mathbb{A}x|$ мы будем обозначать норму оператора. Кроме этого мы будем использовать следующие нормы:

$$|\mathbb{A}|_2^2 = \sum_{1 \leq i,j \leq s} a_{ij}^2, \quad |\mathbb{A}|_\infty = \max_{1 \leq i,j \leq s} |a_{ij}|.$$

Матрица $\mathbb{I}$ - единичная матрица (или единичный оператор). Введем функцию $\mathcal{M}(t; N) = 1/\sqrt{|t|N} + \sqrt{|t|}$ для $|t| > 0$. Если случайный вектор X принимает значения в линейном простанстве, тогда $\bar{X}$ - его независимая копия, а $\tilde{X} = X - \bar{X}$ - симметризация X. Для условных математических ожиданий мы будем использовать следующие обозначения:

$$\mathbb{E}\{f(X,Y)|Y\} = \mathbb{E}_X f(X,Y) = \mathbb{E}^Y f(X,Y).$$

Введем ряд обозначений для моментов:

$$\beta_s = \mathbb{E}|\varphi_1(X)|^s, \gamma_s = \mathbb{E}|\varphi(X,\bar{X})|^s,$$
$$\sigma^2 = \gamma_2, \gamma_{s,r} = \mathbb{E}(\mathbb{E}\{|\varphi(X,\bar{X})|^s|X\})^r,$$

также преположим, что

$$\beta_2 < \infty, 0 < \sigma^2 < \infty.$$

Тогда дисперсию T можно записать в виде:

$$\mathbb{E}T^2 = \beta_2 + \frac{N-1}{2N}\sigma^2.$$

Статистика T является вырожденной из-за условия $\sigma^2 > 0$, в силу которого квадратичная часть статистики не является асимптотически пренебрежимой и, следовательно, распределение статистики

T не является асимптотически нормальным. Более точно, асимптотическое распределение T задается распределением случайной величины

$$T_0 = \frac{1}{2} \sum_{j \geq 1} q_j (\eta_j^2 - 1) + \sum_{j \geq 0} a_j \eta_j,$$

где η_j последовательность независимых одинаково распределенных стандартных нормальных случайных величин, $a_0, a_1, \ldots$ последовательность суммируемых в квадрате весов и $|q_1| \geq |q_2| \geq \ldots$ собственные значения оператора Гильберта-Шмидта $\mathbb{Q}$, ассоциированного с ядром φ. Оператор $\mathbb{Q}$ определен в (1.3). Далее, будем считать $\sigma^2 = 1$.

Рассмотрим статистику T_*:

$$T_* = \sum_{1 \leq i < k \leq N} \varphi(X_j, X_k) + f_1(X_1, \ldots, X_M) + f_2(X_{M+1}, \ldots, X_N), \tag{1.7}$$

где $1 \leq M \leq N/2$. $f_1 = f_1(X_1, \ldots, X_M)$ - произвольная статистика, зависящая только от $X_1, \ldots, X_M$, $f_2 = f_2(X_{M+1}, \ldots, X_N)$ также произвольная статистика не зависящая от $X_1, \ldots, X_M$. Заметим, что класс статистик T_* является более общим, чем класс статистик T.

Для статистики T_* определим функцию концентрации:

$$Q(T_*; \lambda) = \sup_x \mathbb{P}\{x \leq T_* \leq x + \lambda\}, \lambda \geq 0.$$

Глава 1

В главе 1 получены вспомогательные результаты, в число которых входит оценка характеристических функций статистики T_* и мультипликативное неравенство для U-статистик. Аналогичные результаты имеются в работе Бенткуса и Гетце [13], однако, правые части полученных неравенств не содержат зависимости от собственных значений оператора Гильберта-Шмидта, которая понадобится нам для улучшения оценок. Получить такую зависимость удалось благодаря предложенным в настоящей работе условиям невырожденности и замене леммы 6.6 из работы Бенткуса и Гетце на модифицированную лемму 2.5 из работы Ульянова и Гетце [20].

Лемма 1.1. *Пусть* $\mathbb{A}$ *невырожденная матрица размера* $s \times s$, $X \in \mathbb{R}^s$*- случайный вектор с матрицей ковариаций* C. *Предположим, что существует константа* c_s *такая, что*

$$\mathbb{P}\{|X| \le c_s\} = 1,\ |\mathbb{A}| \le c_s,\ |C^{-1}| \le c_s.$$

Пусть U *и* V *независимые случайные векторы, являющиеся суммами* n *независимых копий* X. *Тогда*

$$|\mathbb{E}e\{t\langle \mathbb{A}U, V\rangle\}| \le c(s)|det\mathbb{A}|^{-1}\mathcal{M}^{2s}(t;N) \text{ для } |t| > 0,$$

Оценка в последней лемме зависит от детерминанта матрицы $\mathbb{A}$. Сформулируем условия невырожденности, использованные в работе Бенткуса и Гетце [13]. Условия невырожденности считаются выполненными для распределения случайного вектора Z, ядра φ, параметров $0 < p < 1, \delta \ge 0$ и $s \in \mathbb{N}$, если выполнено неравенство

$$\mathbb{P}\left\{|\mathbb{A}(Z) - \mathbb{I}|_\infty \le \delta\right\} = \mathbb{P}\left\{\max_{1 \le i,j \le s} |\varphi(Z_i, \bar{Z}_j) - \delta_{ij}| \le \delta\right\} \ge p,$$

где $\mathbb{A}$ - случайная матрица, составленная следующим образом:

$$\mathbb{A} = \mathbb{A}(Z) = (a_{ij})_{1 \le i,j \le s},$$

где $a_{ij} = \varphi(Z_i, \bar{Z}_j), Z_i, \bar{Z}_j -$ независимые копии Z.

Условия невырожденности, на которые опирается наша работа, наложены не на вероятность отклонения значений оператора от единичного оператора, а на детерминант самой матрицы $\mathbb{A}$. Будем говорить, что вектор Z удовлетворяет условиям невырожденности $\mathcal{N}(\delta, p)$, если:

$$\mathbb{P}\{W(\bar{Z}) > \delta\} \ge p,$$

[20] *Ulyanov V., Götze F.* Uniform approximations in the CLT for balls in euclidian spaces // University of Bielefeld.- 2000.- no. SFB 343.

$$\mathbb{P}\{|\varphi(Z_i, \bar{Z}_j)| \le c\} \ge c_1,\ 1 \le i, j \le s,$$

$$\text{где } W(\bar{Z}) = (det\mathbb{A})^2, \mathbb{A} = \{a_{ij}\}_{i,j=1}^s, a_{ij} = \varphi(Z_i, \bar{Z}_j),$$

$$Z_i, \bar{Z}_j - \text{ независимые копии вектора } Z.$$

При этом параметр p мал, а параметр c_1 близок к единице.

Благодаря условию $W(\bar{Z}) > \delta = q_1^2 \dots q_9^2$ нам удалось в явном виде выписать зависимость от собственных значений в оценке характеристической функции T_*. Сформулируем основной результат главы 1, в котором получена оценка характеристической функции статистики T_*.

Теорема 1.1. *Пусть $m \in \mathbb{N}$ и $Y = (2m)^{-1/2}(\tilde{X_1} + \ldots + \tilde{X_m})$ удовлетворяет условиям невырожденности $\mathcal{N}(q_1^2 \dots q_9^2, p)$. Тогда для любой статистики T_*, определенной в (1.7), справедливо неравенство:*

$$|\mathbb{E}e\{tT_*\}| \ll_s \frac{1}{|q_9|^9} \mathcal{M}^{2s}(tm; pM/m). \tag{1.8}$$

Неравенство (1.8), полученное в данной теореме, содержит зависимость от собственных значений $q_j, j = \overline{1,9}$ оператора $\mathbb{Q}$. Аналогичное неравенство получено в работе Бенткуса и Гетце [13], однако, в нем не указана явная зависимость от собственных значений оператора $\mathbb{Q}$. Помимо этого, теорема 1.1 используется в доказательстве основных результатов глав 2 и 3. Для доказательства теоремы 1.1 использовалась техника перехода к дискретным случайным векторам, предложенная в работе Юринского [19]. Суть данного метода состоит в оценке сверху характеристической функции некоторой случайной величины характерестической функцией дискретной случайной величины.

Важную роль в построении оценки для функций концентрации и оценки точности аппроксимации играет мультипликативное неравенство, полученное в лемме 1.2. Определим функцию $\psi(t) = |\mathbb{E}\exp(itT^\vartheta)|$, где T^ϑ - некоторая вспомогательная статистика, используемая в доказательстве ряда результатов.

Лемма 1.2. *Пусть $d \geq 0$ и $s \in \mathbb{N}$. Предположим, что сумма $Y = (2m)^{-1/2}(\tilde{X}_1 + \ldots + \tilde{X}_m)$ удовлетворяет условиям невырожденности $\mathcal{N}(q_1^2 \ldots q_9^2, p)$. Тогда существуют константы $c_1(s,d)$ и $c_2(s,d)$ такие, что событие*

$$D = \{\psi(t-\gamma)\psi(t+\gamma) \leq c_1(s,d)\frac{1}{|q_9|^9}\mathcal{M}^s(\gamma m; pM/m)\},$$

удовлетворяет условию

$$\mathbb{P}\{D\} \geq 1 - c_2(s,d)(pM/m)^{-d}.$$

Доказательство мультипликативного неравенства в диссертационной работе аналогично доказательству леммы 7.1 в работе Бенткуса и Гетце [13], но результат, опять же, отличается правой частью неравенства, которая зависит от собственных значений оператора $\mathbb{Q}$.

Глава 2

Основным результатом главы 2 является теорема 1.2, в которой получена оценка функций концентрации U-статистики T_*.

Теорема 1.2. *Пусть $\lambda \geq 0$, тогда для функции концентрации статистики T_* при $q_9 \neq 0$ справедливо неравенство*

$$Q(T_*;\lambda) \ll |q_9|^{-18}\frac{\max\{\lambda; m_0\}}{M},$$

где $m_0 \asymp |q_1 \ldots q_9|^{-3}p^{-1}(|q_1 \ldots q_9|^{-3}p^{-1}\gamma_{2,3/2} + \gamma_3), p \asymp c$.

Построенная в теореме 1.2 оценка содержит зависимость от собственных значений оператора $\mathbb{Q}$ вида $O(|q_9|^{-c}), c > 0$. Ранее, оценка функций концентрации порядка $O(\exp\{-c|q_9|\})$ была доказана Бенткусом и Гетце [13]. Таким образом, в диссертационной работе оценка существенно улучшена по зависимости от собственных значений оператора $\mathbb{Q}$.

Данная теорема доказывается в три этапа. На первом этапе мы докажем, что, если для случайного вектора $Y = (2m)^{-1/2}(X_1 + \ldots + X_m)$ выполнены условия невырожденности $\mathcal{N}(q_1^2 \ldots q_9^2, 2p)$, тогда

$$Q(T_*;\lambda) \ll |q_9|^{-18}\frac{\max\{\lambda; m\}}{pM}. \tag{1.9}$$

Для доказательства (1.9) используется лемма 3 из книги Петрова [21], согласно которой

$$Q(T_*;\lambda) \le 2\max\left\{\lambda;\frac{1}{A}\right\}\int_{-A}^{A}\left|\Psi\hat{(}t)\right|dt,$$

$$\text{где } \Psi\hat{(}t) = \int_{\mathbb{R}} e\{tx\}d\Psi(x), \Psi(x) = \mathbb{P}\left\{T_* \le x\right\}.$$

Полученный интеграл оценивается с помощью леммы 1.3, приведенной ниже. Для доказательства леммы 1.3 мы используем теорему 1.1 об оценке характеристических функций и мультипликативное неравенство для U-статистик из леммы 1.2.

Для $A \ge t_0, t_1 \ge 0$ определим интегралы

$$I_0 = \int_{-t_1}^{t_1}|\hat{\Psi}(t)|dt, I_1 = \int_{t_0\le|t|\le A}|\hat{\Psi}(t)|\frac{dt}{|t|}.$$

Лемма 1.3. *Пусть* $m \in \mathbb{N}$. *Предположим, что случайный вектор* $Y = (2m)^{-1/2}(\tilde{X_1} + \ldots + \tilde{X_m})$ *удовлетворяет условиям невырожденности* $\mathcal{N}(q_1^2\ldots q_9^2, p)$ *и* $s \ge 9$. *Введем ряд обозначений и условий*

$$k = \frac{pM}{m}, t_0 = \frac{c_0(s)}{m}k^{-1+2/s}, t_1 = \frac{c_1(s)}{m}k^{-1/2},$$

$$\frac{c_2(s)}{m} \le A \le \frac{c_3(s)}{m},$$

где $c_j(s), 0 \le j \le 3$ *- некоторые положительные константы. Тогда*

$$I_0 \ll_s |q_9|^{-9}(pM)^{-1}, I_1 \ll_s |q_9|^{-18}m(pM)^{-1}.$$

Сформулируем лемму, необходимую для оценки интеграла I_1 .

[21] *Петров В. В.* Суммы независимых случайных величин. М.: Наука, 1972.

Лемма 1.4. *Пусть $\varphi(t), t \geq 0$ - непрерывная функция такая, что $\varphi(0) = 1, 0 \leq \varphi \leq 1$. Предположим, что для $s > 8$ выполнено неравенство:*

$$\varphi(t)\varphi(t+\tau) \leq \Theta \mathcal{M}^s(\tau; N),$$

для всех $t \geq 0$ и $\tau \geq 0$ и некоторого $\Theta \geq 1$, не зависящего от t и τ.

Тогда для любых $A \geq 1, 0 < B \leq 1$ и $N \geq 1$ справедливо неравенство

$$\int_{B/\sqrt{N}}^{A} \varphi(t)\frac{dt}{t} \ll_s \frac{\Theta^2(1+logA)}{N} + \Theta^2 B^{-s/2} N^{-s/4}.$$

На втором этапе доказательства теоремы нам необходимо показать, что, если гауссовское распределение G удовлетворяет условиям невырожденности $\mathcal{N}(4q_1^2 \ldots q_9^2, p)$, тогда

$$Q(T_*; \lambda) \ll |q_9|^{-18} \frac{\max\{\lambda; m_0\}}{pM}, \tag{1.10}$$

где $m_0 \asymp c|q_1 \ldots q_9|^{-3} p^{-1}(|q_1 \ldots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3)$.

Мы используем (1.9) и лемму 1.5, в которой показано, что, начиная с некоторого номера m, выполнение наших условий невырожденности $\mathcal{N}(4q_1^2 \ldots q_9^2, p)$ для гауссовского вектора влечет за собой выполнение условий невырожденности $\mathcal{N}(q_1^2 \ldots q_9^2, 2p)$ и для суммы случайных векторов $S_m = m^{-1/2}(X_1 + \ldots + X_m)$.

Лемма 1.5. *Пусть гауссовский вектор G удовлетворяет условиям невырожденности $\mathcal{N}(4q_1^2 \ldots q_9^2, 1-p)$.*

Тогда, при

$$m \geq c_s |q_1 \ldots q_9|^{-3} p^{-1}(|q_1 \ldots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3),$$

сумма $S_m = m^{-1/2}(X_1 + \ldots + X_m)$ удовлетворяет условиям невырожденности $\mathcal{N}(q_1^2 \ldots q_9^2, 1-2p)$.

Для доказательства результата теоремы 1.2 нам потребуется (1.10), а также оценка вероятности p, которая будет получена с помощью леммы 1.6 и неравенства

$$\mathbb{P}\{Z > 0,5\} \geq 0,25A^{-2},$$

где Z - неотрицательная случайная величина, такая, что $\mathbb{E}Z = 1, \mathbb{E}Z^2 \leq A$. Ниже выпишем лемму, в которой получено логарифмическое неравенство для моментов второго порядка, а также найден момент первого порядка детерминанта случайной матрицы $\mathbb{A}$.

Лемма 1.6. *Пусть $G_1, \dots, G_s, G'_1, \dots, G'_s$ независимые, одинаково распределенные случайные элементы в $\mathbb{R}^\infty$ такие, что $G_i = (G_{i1}, G_{i2}, \dots), G'_i = (G'_{i1}, G'_{i2}, \dots)$, где G_{ij}, G'_{ij} - независимые, стандартные нормальные случайные элементы. Пусть $|q_1| \geq |q_2| \geq \dots$ - собственные значения оператора Гильберта-Шмидта $\mathbb{Q}$. $W = (det\mathbb{A})^2$, где $\mathbb{A}(G) = \{a_{ij}(G)\}_{i,j=1}^s$, $a_{ij}(G) = \varphi(G_i, G'_j) = \langle \mathbb{Q}G_i, G_j \rangle$. Тогда*

$$\mathbb{E}W = (s!)^2 \sum_{1 \leq i_1 < \dots < i_s < \infty} (q_{i_1} \dots q_{i_s})^2,$$

$$(\mathbb{E}W^2)^{1/2} \leq c(s)\mathbb{E}W.$$

Глава 3

В главе 3 получена оценки точности аппроксимации функции распределения U-статистики и их приложения к статистикам фон Мизеса. Основной результат данной главы сформулирован в виде теоремы 1.3, в которой получена оценка точности аппроксимации функции распределения U-статистики в оптимальных моментных условиях.

Теорема 1.3. *(i) Пусть*
$m_0 \asymp |q_1 \dots q_9|^{-3} p^{-1} (|q_1 \dots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3)$, $p \asymp c$.
Тогда

$$\Delta_N \ll \frac{m_0}{N|q_9|^{18}} + \frac{(\beta_3^2 + \gamma_{2,2})}{N|q_{13}|^{13}} + \frac{(\beta_4 + \gamma_3)}{N|q_9|^9}.$$

(ii) Пусть выполнено условие (1.2).
Тогда

$$\Delta_N \ll \frac{m_0}{N|q_9|^{18}} + \frac{1}{N|q_9|^9} \cdot (\beta_4 + \beta_3^2 + \gamma_3 + \gamma_{2,2}).$$

В работе Бенткуса и Гетце [13] аналогичная оценка содержит константу, которая имеет порядок $O(\exp{(-c|q_j|)}), j = 9, 13$, по зависимости от собственных значений оператора $\mathbb{Q}$. В диссертационной работе оценка существенно улучшена по зависимости от $|q_i|$ и имеет степенной порядок $O(|q_j|^{-c}), j = 9, 13$

Доказательство данной теоремы опирается на лемму 1.7

Лемма 1.7. *Предположим, что Y удовлетворяет условиям невырожденности $\mathcal{N}(q_1^2 \dots q_9^2, p)$.*

(i) Пусть $s \geq 13$ и $m_0 \asymp |q_1 \dots q_9|^{-3} p^{-1} (|q_1 \dots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3)$. Тогда

$$\Delta_N \ll_s \frac{m_0}{pN|q_9|^{18}} + \frac{(\beta_3^2 + \gamma_{2,2})}{N|q_s|^s} + \\ + \frac{1}{p^6 N|q_9|^9} \cdot (\beta_4 + \beta_3^2 + \gamma_3 + \gamma_{2,2}).$$

(ii) Пусть выполнено условие (1.2) и $s \geq 9$. Тогда

$$\Delta_N \ll_s \frac{m_0}{pN|q_9|^{18}} + \frac{(\beta_3^2 + \gamma_{2,2})}{N|q_s|^s} + \\ + \frac{1}{p^4 N|q_9|^9} \cdot (\beta_4 + \gamma_3 + \gamma_{2,2}).$$

Константа p оценивается также, как и в теореме 1.2, а величина m_0 получена в лемме 1.5. Доказательство леммы 1.7 опирается на леммы 1.8 и 1.9.

Определим статистику $T^{(r)}$

$$T^{(r)} = \frac{1}{N} \sum_{1 \leq i < j \leq N} \varphi(Z_i, Z_j) + \sum_{1 \leq i \leq N} f_i(Z_i),$$

где

$$Z_j = X_j \text{ для } 1 \le j \le r, r \in \mathbb{Z}^+, Z_j = G_j \text{ для } r < j \le N.$$

В следующей лемме получено разложение функции распределения для статистики $T^{(r)}$. В доказательстве данной леммы мы использовали результаты с оценкой характеристической функции в теореме 1.1 и мультипликативное неравенство из леммы 1.2.

Лемма 1.8. *Пусть $m \in \mathbb{N}, s \ge 9$ и $t_0 = m^{-1}(pN/m)^{-1+2/s}$. Предположим, что случайный вектор Y удовлетворяет условиям невырожденности $\mathcal{N}(q_1^2 \dots q_9^2, p)$. Тогда, для $pN > m$ и $m^{-1} \ge t_* \ge t_0$ функция распределения $F^{(r)}$ статистики $T^{(r)}$ удовлетворяет равенству*

$$F^{(r)}(x) = \frac{1}{2} + \frac{i}{2\pi} V.P. \int_{-Nt_*}^{Nt_*} e\{-xt\} \hat{F}^{(r)}(t) \frac{dt}{t} + R,$$

$$\text{где } |R| \ll_s m/(pN|q_9|^{18}).$$

В лемме 1.9 доказана оценка поправки Эджворта.

Лемма 1.9. *Преобразование Фурье-Стилтьеса поправки Эджворта F_1 удовлетворяет неравенству:*

$$|\hat{F}_1(t)| \ll N^{-1/2} |t|^3 (\beta_3^2 + \gamma_{2,2})^{1/2} \prod_{j \ge 1} (1 + 2t^2 q_j^2/25)^{-1/4}.$$

Далее, при $s \ge 7$ справедливы неравенства:

$$\int_{|t| \ge \lambda} |\hat{F}_1(t)| \frac{dt}{|t|} \ll N^{-1/2} (\beta_3^2 + \gamma_{2,2})^{1/2} |q_s|^{-s/2} \lambda^{3-s/2}, \text{ для } \lambda > 0,$$

$$\int_{\mathbb{R}} |\hat{F}_1(t)| \frac{dt}{|t|} \ll N^{-1/2} (\beta_3^2 + \gamma_{2,2})^{1/2} |q_s|^{-3}.$$

Благодаря лемме 1.8, мы можем оценить величину Δ_N сверху следующим образом

$$\Delta_N \le J + R,$$

$$J =_{def} \int_{-Nt*}^{Nt_*} |\hat{F}(t) - \hat{F}_0(t) + \hat{F}_1(t)| \frac{dt}{|t|}.$$

R - остаточный член, который оценивается с помощью сформулированных ранее лемм 1.9 и 1.8.

Определим $\kappa = \kappa(t)$

$$\kappa(t) = \kappa(t, N, \varphi, \mathcal{L}(X)) = \kappa_1(t) + \kappa_2(t), l = [(N-2)/20]$$

где

$$\kappa_1(t) = \sup_L |\mathbb{E}e\{tN^{-1} \sum_{1 \leq j < k \leq l} \varphi(X_j, X_k) + L(X_1, \ldots, X_l)\}|,$$

$$\kappa_2(t) = \sup_L |\mathbb{E}e\{tN^{-1} \sum_{1 \leq j < k \leq l} \varphi(G_j, G_k) + L(G_1, \ldots, G_l)\}|,$$

супремум берется по всем линейным статистикам L, которые могут быть представлены как $L(x_1, \ldots, x_l) = \sum_{j=1}^{l} f_j(x_j)$ с некоторыми функциями $f_1, \ldots, f_l$. Мы можем получить оценку для $\kappa(t)$, применив теорему 1.1 с заменой N на l.

Оценим J. Чтобы оценить подынтегральное выражение выпишем лемму 1.10, в которой доказана оценка для $\hat{\Delta}_N = |\hat{F}(t) - \hat{F}_0(t) + \hat{F}_1(t)|$.

Лемма 1.10. *Справедливо неравенство:*

$$\hat{\Delta}_N \ll \kappa N^{-1}(t^4\beta_4 + t^6\beta_3^2 + t^2\gamma_2 + |t|^3\gamma_3 + |t|^5\gamma_2\gamma_3 + t^2\gamma_{2,2} + t^6\gamma_2\gamma_{2,2}).$$

Если выполнено условие (1.2), тогда $F_1\hat{(t)} = 0$ *и выполнено неравенство*

$$\hat{\Delta}_N \ll \kappa N^{-1}(t^4\beta_4 + t^2\gamma_2 + |t|^3\gamma_3 + t^4\gamma_2\gamma_{2,2}).$$

В разделе 3.3 главы 3 построена оценка для статистик фон Мизеса вида

$$M = \frac{1}{2N} \sum_{1 < i,j < N} \varphi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \leqslant i \leqslant N} \varphi_1(X_i).$$

Ядра φ и φ_1 вырождены. Результат доказан аналогично оценке точности аппроксимации U-статистик.

Рассмотрим функцию $\psi(x) = (\varphi(x,x) - \nu)/2$, где $\nu = \mathbf{E}\varphi(X,X)$. Перепишем предыдущее равенство для M

$$M - \frac{\nu}{2} = T + \frac{1}{N} \sum_{1 \leqslant i \leqslant N} \psi(X_i),$$

где T определена в (1.1). Мы применим оценку для данных статистик, предполагая, что $\mathbf{E}\psi(X) = 0$ и $\varrho = \mathbf{E}\psi^2(X) < \infty$. Пусть F_* - функция ограниченной вариации (при условии $q_3 \neq 0$) с преобразованием Фурье-Стилтьеса

$$\hat{F}_*(t) = \frac{it}{\sqrt{N}} \mathbf{E}\psi(G) e\{tT_0\} = \frac{it}{\sqrt{N}} \mathbf{E}\psi_0(G) e\{tT_0\},$$

и таким, что $F_*(-\infty) = 0$. Запишем $H_1 = F_1 + F_*$, пусть H - функция распределения $M - \nu/2$. Определим

$$\delta_N = \sup_x |\delta_N(x)|, \qquad \delta_N(x) = H(x) - F_0(x) - H_1(x).$$

Основным результатом диссертационной работы для статистик фон Мизеса является сформулированная ниже теорема 1.4.

Теорема 1.4. *(i) Предположим, что* $q_{13} \neq 0$, $m_0 \asymp |q_1 \dots q_9|^{-3} p^{-1} (|q_1 \dots q_9|^{-3} p^{-1} \gamma_{2,3/2} + \gamma_3)$, $p \asymp c$. *Тогда мы имеем*

$$\delta_N \leqslant \frac{m_0}{N|q_9|^{18}} + \frac{(\beta_3^2 + \gamma_{2,2})}{N|q_{13}|^{13}} + \frac{(\beta_4 + \gamma_3 + \varrho)}{N|q_9|^9}.$$

(ii) Предположим, что (1.2) выполнено и $q_9 \neq 0$. *Тогда мы имеем*

$$\delta_N \leqslant \frac{m_0}{N|q_9|^{18}} + \frac{1}{N|q_9|^9} \cdot (\beta_4 + \beta_3^2 + \gamma_3 + \gamma_{2,2} + \varrho).$$

Зависимость от собственных значений q_j оператора $\mathbb{Q}$ имеет степенной порядок. Результата теоремы 1.4 существенно улучшает аналогичную оценку, которая построена в работе Бенткуса и Гетце [13] и имеет степенной порядок.

Работа выполнена под руководством доктора физико-математических наук, профессора Владимира Васильевича Ульянова, которому автор выражает искреннюю благодарность за постановку задачи, терпение и постоянное внимание к работе.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. *Zubayraev T.* Asymptotic analysis for U-statistics and its application to von Mises statistics // Open Journal of Statistics.- 2011.- Vol. 1, no. 3.- Pp. 139-144.

2. *Зубайраев Т.А.* Об асимптотическом анализе U-статистик: оценка функций концентрации // Вестник Тверского государственного университета. Серия: Прикладная математика.- 2011.- № 22.- С.73-84

3. *Зубайраев Т.А.* Об асимптотическом анализе U-статистик: оценка точности аппроксимации функции распределения // Сборник статей молодых ученых факультета ВМК МГУ.- 2010.- Т. 1, № 7.- С. 99-108.

4. *Зубайраев Т.А.* Оценка функций концентрации U-статистик // Сборник тезисов XVIII Международной научной конференции студентов, аспирантов и молодых ученых "ЛОМОНОСОВ - 2011"; секция "Вычислительная математика и кибернетика"; Москва, МГУ им. М.В. Ломоносова.- 2011.- С. 15-16.