

ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ
ИМЕНИ А. А. ХАРКЕВИЧА РАН

На правах рукописи

Янович Юрий Александрович

АСИМПТОТИЧЕСКИЕ СВОЙСТВА ПРОЦЕДУР СТАТИСТИЧЕСКОГО
ОЦЕНИВАНИЯ НА МНОГООБРАЗИЯХ

01.01.05 — Теория вероятностей и математическая статистика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
доктор физико-математических наук,
проф. Бернштейн Александр Владимирович

Москва
2017

Содержание

Введение	4
1 Локальное поведение случайных выборок на многообразии	11
1.1 Описание задачи	11
1.2 Модель данных	14
1.3 Основные результаты главы	18
1.4 Некоторые определения и леммы	20
1.4.1 Локальная линеаризация	20
1.4.2 Леммы Муавра-Лапласа для медленно убывающей вероятности	22
1.4.3 Вероятность больших отклонений для ограниченных случайных величин	23
1.4.4 Замена области интегрирования	24
1.4.5 Конечные сети	25
1.5 Доказательство основных теорем для попадания в окрестность	25
2 Непараметрическое оценивание на многообразии	31
2.1 Рассматриваемые статистики	31
2.2 Основные результаты главы	34
2.3 Доказательство основных теорем главы	36
3 Выборочное оценивание непрерывных интегральных операторов	46
3.1 Постановка задачи	46
3.2 Иллюстративный пример	47
3.3 Рассматриваемые статистики и основные результаты главы	49
3.4 Доказательство основных теорем главы	52
4 Свойства процедур моделирования многообразий	52
4.1 Оценивание многообразий как задача вложения многообразий	53
4.2 Типичная схема алгоритмов вложения многообразий	54
4.2.1 Шаг первый: построение окрестности	54
4.2.2 Шаг второй: описание окрестностей	54
4.2.3 Шаг третий: глобальное описание	55

4.2.4	Шаг четвертый: расширение задачи для точек вне выборки	55
4.3	Основной результат главы	55
Заключение		58
Список литературы		60
А Определения и леммы из дифференциальной геометрии		66
A.1	Ковариантное дифференцирование	66
A.2	Квадратичные формы многообразия и кривизна Риччи	68
A.3	Локально-римановы координаты и экспоненциальное отображение	70
В Доказательство лемм		72
B.1	Леммы Муавра-Лапласа для медленно убывающей вероятности .	73
B.2	Вероятность больших отклонений для ограниченных случайных величин	78
B.3	Замена области интегрирования	79
B.4	Конечные сети	80
B.5	Моменты случайных величин	82
B.6	Случайные величины у границы	87

Ведение

Актуальность темы

Развитие вычислительной техники и информационных технологий привело к возможности хранить, передавать по каналам связи и быстро обрабатывать большие массивы данных, осуществлять быстрый удаленный доступ к ним. Появление в результате таких возможностей “шквала данных”, называемого обычно парадигмой *больших данных* (Big Data), и новые возможности для работы с ними, позволили ставить и эффективно решать новые научные и прикладные задачи в области промышленности и сельского хозяйства, биологии и медицины, обработки сигналов, речи, текстов и изображений, и др. Для решения таких задач были развиты методы работы с большими данными, научную основу которым составляет новая мультидисциплинарная область знаний, выделившаяся в XXI веке в отдельную академическую и университетскую дисциплину *наука о данных* (Data Science), в которой сконцентрированы методы математики и статистики, распознавания образов, визуализации и машинного обучения, информационных технологий и *интеллектуального анализа данных*.

Понятие *большие данные* включает в себя не только большой объем данных, но и их высокую размерность, так как реальные данные обычно имеют очень высокую размерность (например, размерность цифровой черно-белой фотографии равна числу ее пикселей и может достигать сотен тысяч; изображения головного мозга, получаемые ежесекундно с помощью функциональной магнито-резонансной томографии, имеют размерность порядка полутора миллионов). Однако многие традиционные методы и алгоритмы становятся неэффективными или просто неработоспособными для данных высокой размерности (не только в вычислительном, но и в статистическом смысле), и этот феномен назван *проклятием размерности* [1]. Известный статистик Д. Донохо сказал в 2000 году на конференции, посвященной математическим вызовам 21-го века: “мы можем с полной уверенностью сказать, что в наступающем веке анализ многомерных данных станет очень важным занятием, и совершенно новые методы многомерного анализа данных будут разработаны, просто мы еще не знаем, каковы они будут” [2].

Однако совокупность конкретных “реальных” данных, полученных из реальных источников, в силу наличия различных зависимостей между компонен-

тами данных и ограничений на их возможные значения, занимает, как правило, малую часть высокоразмерного пространства наблюдений, имеющую невысокую внутреннюю размерность (например, множество всех черно-белых портретных изображений человеческих лиц с исходной размерностью порядка сотен тысяч, имеет внутреннюю размерность не выше ста). Следствием невысокой внутренней размерности является возможность построения низкоразмерной параметризации таких данных. Поэтому многие алгоритмы для работы с высокоразмерными данными начинаются с решения задачи снижения размерности, результатом которого являются низкоразмерные описания таких данных.

Традиционные методы математической статистики в задаче снижения размерности рассматривают, в основном, лишь случай, когда высокоразмерные данные сосредоточены вблизи неизвестного низкоразмерного аффинного подпространства, которое можно оценить, например, с помощью метода главных компонент К. Пирсона. Однако многие реальные данные имеют нелинейную, “искривленную” структуру, для нахождения которой в прошлом веке были предложены различные эвристические методы нелинейного снижения размерности (например, репликативные нейронные сети [3] или нелинейные методы многомерного шкалирования [4]), свойства которых невозможно было исследовать строгими методами в силу отсутствия математической модели для обрабатываемых данных. Имелись лишь отдельные работы (например, работы В.М. Бухштабера [5, 6]) в которых строгие дифференциально-геометрические и статистические методы использовались при решении нелинейных статистических задач.

Лишь в 2000 году появилась первая математическая модель многомерных нелинейных данных, названная *моделью многообразия* [7]), в соответствии с которой, высокоразмерные данные расположены на (или вблизи) неизвестного низкоразмерного нелинейного многообразия (*многообразия данных*), вложенного в высокоразмерное пространство наблюдений, а наблюдаемая выборка случайно извлекается из многообразия данных в соответствии с неизвестным вероятностным распределением на нем. Размерность *многообразия данных* также может быть неизвестной и оцениваться по выборке.

Область анализа данных, в которой статистические задачи рассматриваются в рамках этой модели, получила название *моделирование многообразий* [8] и в настоящее время является бурно развивающимся разделом *науки*

о данных. Изначально под моделированием многообразий понимались только задачи нелинейного снижения размерности [7, 9], впоследствии в эту область были включены задачи регрессии на многообразии [10], оценки плотности на многообразии [11], и др.

Методы и алгоритмы моделирования многообразий существенно опираются на геометрическую структуру данных и позволяют эффективно решить большое количество прикладных задач анализа данных. Эти методы “эксплуатируют” различные дифференциально-геометрические свойства многообразий данных (например, возможность аппроксимации окрестности выбранной точки многообразия касательной плоскостью невысокой размерности, описание кратчайших расстояний на многообразии геодезическими линиями, и другие). По существу, появился новый раздел многомерного статистического анализа — анализ данных, лежащих на нелинейном многообразии меньшей размерности (manifold valued data), в котором существенную роль играют такие дифференциально-геометрические характеристики нелинейного многообразия данных как кривизна, риманова структура и др.

Большинство работ по моделированию многообразий содержит описание различных процедур, свойства которых исследуются лишь с помощью вычислительных экспериментов. Лишь в небольшом числе работ [12, 13] математически строго исследовались статистические свойства конкретных функций от данных (статистик). При этом известные фундаментальные методы анализа асимптотического поведения статистик, развитые во второй половине прошлого века, значительный вклад в которые внесли, в том числе, такие видные советские и российские исследователи как Л.Н. Большев, И.А. Ибрагимов, Р.З. Хасьминский, А.А. Боровков, Д.М. Чибисов и др., разработаны для данных на линейных подпространствах и непосредственно неприменимы для данных на нелинейных многообразиях.

В процедурах моделирования многообразий используются, как правило, три типа статистик:

1. локальные статистики, построенные по подвыборке, состоящей из точек попавших в малую окрестность выбранных точек, и оценивающие локальные и геометрические характеристики многообразия (касательное пространство [14], Риманов метрический тензор [15], и др.); распределение таких статистик зависит, в том числе, от локально-геометрических свойств

многообразия;

2. глобальные (интегральные) статистики, являющиеся усреднением локальных статистик по точкам выборки и которые могут также зависеть от параметров (например, искомым неизвестных низкоразмерных представлений);
3. статистики, являющиеся результатом оптимизации глобальных статистик по этим параметрам.

Оптимизация выборочных интегральных статистики во многих случаях сводится к оптимизации выборочных квадратичных форм от этих параметров (алгоритмы IsoMap [16], Locally-linear Embedding [17], Local Tangent Space Alignment [18], Laplacian Eigenmaps [19], Hessian Eigenmaps [20] и Grassmann&Stiefel Eigenmaps [21], и др.), решение которых основаны на спектральных методах [22] (обобщенные задачи на собственные векторы). Решение таких выборочных задач часто сводится к решению спектральных непрерывных задач на многообразии. Например, оптимизируемая выборочная квадратичная форма в алгоритме Laplacian Eigenmaps [14, 19] является выборочным аналогом оператора Лапласа-Бельтрами на определенном классе функций, заданных на многообразии данных, а наилучшие низкоразмерные представления сходятся к конкретным собственным функциям этого оператора [13].

Однако до сих пор отсутствуют результаты об асимптотическом поведении статистик на многообразиях для достаточно общих классов, включающих в себя как ранее неисследованные статистики из различных существующих алгоритмов, так и новые статистики для создаваемых алгоритмов с желаемыми свойствами. Именно отсутствие таких общих результатов для классов статистик и явилось мотивацией исследования. Тем самым, целью диссертационной работы является изучение асимптотических свойств выбранных классов статистик всех трех типов.

Основные задачи исследования

Для достижения сформулированной цели были поставлены и решены следующие взаимосвязанные асимптотические статистические задачи:

1. исследование статистических свойств случайных подвыборок, попавших в асимптотически малую окрестность выбранной точки многообразия;
2. исследование асимптотического поведения выбранного класса локальных статистик в окрестности рассматриваемой точки, зависящего, в том числе, от дифференциально-геометрических свойств многообразия в этой точке;
3. исследование асимптотического поведения глобальных (интегральных) статистик получаемых путем усреднения локальных статистик по точкам выборки;
4. исследование сходимости статистик, являющихся решениями выборочных спектральных оптимизационных задач, к решениям предельных непрерывных оптимизационных задач на многообразии.

Основные выносимые на защиту результаты и их новизна

Все приведенные ниже результаты диссертации являются новыми и получены лично соискателем:

1. найдены асимптотическое распределение числа точек в медленно убывающей окрестности выбранной точки многообразия и условное распределение этих точек. Получена равномерная по многообразию верхняя оценка для вероятности больших отклонений числа точек в окрестности от своего среднего значения;
2. найдено асимптотическое распределение локальных статистик рассматриваемого класса, ранее известное лишь для конкретных статистик [14, 23] (оценки элементов локальной выборочной ковариационной матрицы и оценки оператора Лапласа-Бельтрами специального вида). Получена равномерная по многообразию верхняя оценка на вероятности больших отклонений локальных статистик от их средних значений;
3. найдены предельные значения глобальных статистик на многообразиях, принадлежащих рассматриваемому классу; для выбранного класса статистик получена верхняя оценка на вероятности больших отклонений глобальной статистики от своего предельного значения;

4. получена оценка для отклонения собственных чисел и собственных функций, являющихся решениями рассматриваемых выборочных спектральных оптимизационных задач для выбранных глобальных статистик от собственных чисел и собственных функций предельных непрерывных спектральных операторов.

Основные методы исследования

В работе используются методы теории вероятностей и математической статистики для асимптотического поведения случайных величин; методы математического анализа и дифференциальной геометрии для изучения локальной структуры многообразия — носителя данных; методы функционального анализа для получения равномерных по многообразию оценок и доказательства сходимости результата оптимизации выборочных функционалов к их предельным непрерывным аналогам.

Теоретическая ценность и практическая значимость

Работа носит теоретический характер. Её результаты могут быть использованы для изучения свойств и совершенствования существующих алгоритмов анализа лежащих на многообразии данных, а также, для создания новых алгоритмов с желаемыми свойствами. В частности, результаты диссертации были использованы в работах по развитию новых алгоритмов моделирования многообразий и их использованию в прикладных задачах анализа данных [4-11], написанных в соавторстве с соискателем. В этих работах лично соискателю принадлежат результаты и выводы, основанные на полученных в диссертации результатах применительно к конкретным рассматриваемым в этих работах статистикам.

Результаты апробации

Результаты диссертации были доложены на следующих мероприятиях:

1. семинар отдела теории вероятностей и математической статистики МИАН (2017, Москва, Россия);

2. семинар “Теория риска и смежные вопросы” кафедры математической статистики ВМК МГУ (2016, Москва, Россия);
3. математические и статистические семинары ИППИ РАН (Москва, Россия): Добрушинской математической лаборатории (2016), “Статистический кластерный анализ” (2016), сектора интеллектуального анализа данных (2014);
4. междисциплинарные школы-конференции “Информационные технологии и системы”: 40-я (2016, Санкт-Петербург, Россия), 39-я (2015, Сочи, Россия), 37-я (2013, Калининград, Россия);
5. International Conference on Algebra, Analysis and Geometry (2016, Kazan, Russia);
6. The 8th International Conference on Machine Vision, ICMV (2015, Barcelona, Spain);
7. The 14th International Conference on Machine Learning and Applications (2015, Miami, USA; ICMLA);
8. The third international symposium on statistical learning and data sciences, SLDS (2015, Royal Holloway, University of London, United Kingdom);
9. International IEEE Conference on Data Science and Advanced Analytics, DSAA (2015, Paris, France);
10. Yandex School of Data Analysis Conference “Machine Learning: Prospects and Applications” (2015, Berlin, Germany);
11. Premolab-WIAS Workshop “Advances in predictive modeling and optimization” (2014, WIAS, Berlin, Germany);
12. International Workshop on Statistical Learning (2013, Moscow, Russia);
13. The 29th European Meeting of Statisticians, EMS (2013, Budapest, Hungary);
14. Neural Information Processing Systems Conference (NIPS 2013, Lake Tahoe, USA).

Публикации

Список работ автора по теме диссертации: [24–37]. Основные результаты диссертации содержатся в работах автора [24–26]. Работы [24–30] опубликованы в изданиях, индексируемых в системах Scopus или Web of Science.

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения, двух глав приложения и списка литературы, включающего 58 наименований. Объем диссертации составляет 87 страниц, включая 1 рисунок.

Благодарности

Автор благодарен своему научному руководителю Александру Владимировичу Бернштейну за всегда интересные и плодотворные обсуждения, постоянную поддержку и участие.

1 Локальное поведение случайных выборок на многообразии

1.1 Описание задачи

Модель данных включает в себя описание носителя данных и механизма извлечения выборки. Предполагается, что носителем данных является “достаточно хорошее” (предположения строго сформулированы как $M1 - M10$ в следующем разделе) неизвестное многообразие данных \mathbb{M} с известной размерностью q , вложенным в \mathbb{R}^p , $p > q$. Предполагается, что на многообразии имеется “достаточно хорошая” (предположения строго сформулированы как $S1 - S3$ в диссертации) неизвестная мера μ , абсолютно непрерывная относительно меры Римана на многообразии и обладающая плотностью p_μ . Рассматривается выборка данных $\mathbb{X}_N = \{X_1, \dots, X_N\}$ состоит из независимых одинаково распределенных величин распределенных в соответствии с мерой μ .

Многообразию \mathbb{M} в окрестности каждой точки X может быть приближено своим q -мерным касательным пространством $T_X(\mathbb{M})$: в малой окрестности

$U_X \subset \mathbb{M}$ точки $X \in \mathbb{M}$ существует взаимнооднозначное соответствие между точками этой окрестности и точками окрестности TU_X нулевого вектора в касательном пространстве, представленных в виде вектора $t\theta$, где $t \geq 0$ и θ принадлежит единичной сфере S_{q-1} в \mathbb{R}^q . Существует так называемое экспоненциальное отображение \exp_X точек TU_X в касательном пространстве в точки $X' = \exp_X(t\theta)$ окрестности U_X (Рис. 1). Обратное отображение из U_X подмножества касательного пространства $T_X(\mathbb{M})$ задает локальные римановы q -мерные координаты $t\theta$ точек $X' = \exp_X(t\theta) \in U_X(\mathbb{M})$ многообразия.

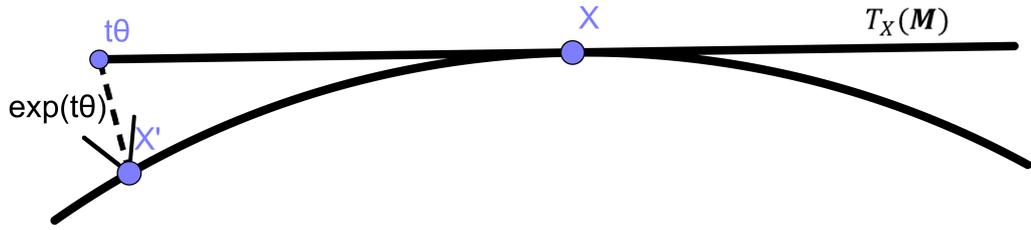


Рис. 1: Экспоненциальное отображение \exp_X в окрестности точки X многообразия \mathbb{M} из касательного пространства $T_X(\mathbb{M})$ к многообразию \mathbb{M} в точке X , $t \in [0, \infty)$, $\theta \in S_{q-1} \subset T_X(\mathbb{M})$ — вектор единичной длины.

Различные методы построения и исследования локальных статистик на многообразии, точки которого принадлежат высокоразмерному пространству \mathbb{R}^p , в той или иной форме используют локальное приближение многообразия в окрестности рассматриваемой точки его q -мерным касательным пространством $T_X(\mathbb{M})$ и порождаемые им q -мерные координаты точек окрестности. Рассматривая в качестве окрестности U_X пересечение $U_{X,\varepsilon}$ ε -шара в \mathbb{R}^p с центром в $X \in \mathbb{M}$ с многообразием \mathbb{M} . Очевидно, что подобное приближение будет тем лучше, чем меньше радиус окрестности ε . Однако, для обеспечения состоятельности рассматриваемых локальных оценок, радиус ε должен быть достаточно большим, чтобы в окрестность $U_{X,\varepsilon}$ попало достаточно много точек при стремлении размера выборки N к бесконечности. Точные требования $P1 - P3$ к поведению радиуса $\varepsilon = \varepsilon_N$ как функции от размера выборки N , в частности, требование $N \cdot \varepsilon^q \rightarrow \infty$, сформулированы в диссертации, этим требованиям удовлетворяет, например, функция

$$E(N) = C \cdot N^{-\frac{1}{q+2}}, \quad (1)$$

где $C > 0$ — константа.

Пусть X — фиксированная точка многообразия \mathbb{M} , и $X' \in \mathbb{M}$ — случайная точка, имеющая распределение μ . Обозначим $I_\varepsilon(X'|X)$ бернуллиевскую случайную величину являющуюся индикатором события $|X' - X| < \varepsilon$, с вероятностью успеха

$$P_\varepsilon = \mu(|X' - X| < \varepsilon) = \mathbb{E}_\mu I_\varepsilon(X'|X), \quad (2)$$

где $|\cdot|$ — обычная евклидова норма в \mathbb{R}^p . При малых ε главный член вероятности успеха P_ε ведет себя [38] (с точностью до членов большего порядка малости) как

$$P_\varepsilon \approx \varepsilon^q \cdot p_\mu(X) \cdot V_q, \quad (3)$$

где V_q — объем q -мерного единичного шара. Обозначим

$$N_\varepsilon(X) = \sum_{n=1}^N I_\varepsilon(X_n|X) \quad (4)$$

биномиальную случайную величину равную числу точек выборки \mathbb{X}_N , попавших в окрестность $U_{X,\varepsilon}$ точки X . Для $N_\varepsilon(X)$ справедливо

$$\mathbb{E}N_\varepsilon(X) = N \cdot P_\varepsilon \approx N \cdot \varepsilon_N^q \cdot p_\mu(X) \cdot V_q. \quad (5)$$

Исследуется поведение случайной величины $N_\varepsilon(X)$ (16) в схеме серий ($N \rightarrow \infty, \varepsilon_N \rightarrow 0, N \cdot \varepsilon_N^q \rightarrow \infty$) с учетом приближений (3) и (5). Известные результаты [39] для евклидова случая (закон больших чисел, центральная предельная теорема, условная асимптотическая равномерность распределения точек, попавших в убывающую окрестность рассматриваемой точки), в которых вероятность успеха P_ε (15) медленно стремится к нулю, с использованием стандартной техники [14, 19] обобщены в диссертации на случай случайных величин, лежащих на многообразии, в полученном обобщении дополнительно исследовалось “смещение” математического ожидания, вызванное кривизной многообразия в точке X . Эти обобщенные результаты сформулированы как Утверждения 1-3 в виде, удобном для дальнейшего использования в диссертации (например, определяют главный член стохастического разложения $N_\varepsilon(X)$ по ε).

Однако, для дальнейшего использования в диссертационной работе, требуются результаты о поведении случайной величины $N_\varepsilon(X)$ (16) для всех точек многообразия “сразу”, оценивая, в частности, вероятность того, что в окрестности всех отделенных от границы точек попадет необходимое бесконечное число точек выборки. Поэтому в диссертации случайные величины $N_\varepsilon(X)$, $X \in \mathbb{M}$, рассматривается как “параметризованное” случайное поле на многообразии данных \mathbb{M} . В рассматриваемом нелинейном случае, ошибка аппроксимации в (3) вызвана не только тем, что плотность $p_\mu(X')$ меняется в окрестности $U_{X,\varepsilon}$, но и отличием $U_{X,\varepsilon}$ от q -мерного шара в касательной плоскости, вызванного кривизной многообразия в точке X . Если поведение членов более высокого порядка малости, обусловленное изменчивостью плотности $p_\mu(X')$ внутри окрестности $U_{X,\varepsilon}$, широко исследовалось и известно в литературе, то влияние кривизны носителя в точках X мало изучено.

1.2 Модель данных

Предполагается, что

- М1. Множество $\mathbb{M} \subset \mathbb{R}^p$ является q -мерным многообразием, покрытым одной картой. То есть, для некоторого открытого $\mathbb{B} \subset \mathbb{R}^q$ и гомеоморфизма $f: \mathbb{B} \rightarrow \mathbb{R}^p$ выполнено: $\mathbb{M} = f(\mathbb{B})$;
- М2. Назмерность q известна;
- М3. Множество \mathbb{B} ограничено;
- М4. Собственные значения матрицы порядка q , являющейся произведением матриц Якоби $J_f^T(b) \cdot J_f(b)$ отображения f , равномерно отделены от 0 и ∞ ;
- М5. Гессиан отображения f существует и ограничен на \mathbb{M} ;
- М6. Третьи производные отображения f существуют и ограничены на \mathbb{M} ;
- М7. Число обусловленности $c(\mathbb{M})$ ограничено, где $c(\mathbb{M})$ — наименьшее действительное число такое, что для каждой точки $Z \in \mathbb{R}^p$, удаленной от \mathbb{M} не более чем на $1/c(\mathbb{M})$, существует единственная проекция на \mathbb{M} [40];
- М8. Многообразие \mathbb{M} геодезически выпукло, т.е. для любой пары точек на \mathbb{M} существует геодезическая линия, соединяющая их, и являющаяся кратчайшей;

- М9. Существует такое положительное число ε_B и положительная константа C_B , что для любого $\varepsilon < \varepsilon_B$ Риманова мера точек $dV(\mathbb{M} \setminus \mathbb{M}_\varepsilon)$, удаленных от границы многообразия не более чем на ε может быть ограничена через Риманову меру многообразия $dV(\mathbb{M})$: $dV(\mathbb{M} \setminus \mathbb{M}_\varepsilon) \leq \varepsilon \cdot C_B \cdot \varepsilon \cdot dV(\mathbb{M})$;
- М10. Существует такое положительное число ε_C и положительная константа C_c , то для любого $\varepsilon < \varepsilon_c$ и любой точки $X \in \mathbb{M}$ Риманова мера пересечения p -мерного шара радиуса ε с центром в точке X ограничено снизу величиной $C_c \cdot \varepsilon^q$.

Замечание 1. *Предположение М1 эквивалентно существованию глобальных q -мерных координат на многообразии.*

Предположение М3 используется для получения равномерных результатов.

Предположения М4-М6 задают условия гладкости, которые используются для связывания Евклидовых расстояний и объемов с расстояниями и объемами на многообразии.

Свойство М7 гарантирует, что многообразие не содержит “коротких замыканий”: близость точек многообразия в p -мерном пространстве равносильна близости точек на многообразии. Иными словами, пересечение малого Евклидова p -мерного шара точки с многообразием является связным множеством.

Предположение М8 является техническим упрощением, полезным для разложения Тейлора.

Предположение М9 гарантирует, что малая окрестность границы \mathbb{M} является множеством малой меры.

Предположение М10 гарантирует, что \mathbb{M} не содержит “иголок”, фактически, каждая точка входит в многообразие с некоторым непустым геометрическим конусом.

Множество \mathbb{V} и производные f ограничены, поэтому многообразию \mathbb{M} также ограничено. Обозначим a — длину ребра описанного вокруг многообразия \mathbb{M} p -мерного куба:

$$a = \inf \{a' : a', a_1, \dots, a_p | \mathbb{M} \subset \otimes_{i=1}^p [a_i, a_i + a']\}. \quad (6)$$

Обозначим c_J и C_J минимальное и максимальное собственные значения матрицы (метрического тензора) $J_f^T(b) \cdot J_f(b)$, $b \in \mathbb{B}$ соответственно:

$$c_J = \inf_{b \in \mathbb{B}} \min_{\lambda \in \sigma(J_f^T(b)J_f(b))} \lambda; \quad (7)$$

$$C_J = \sup_{b \in \mathbb{B}} \max_{\lambda \in \sigma(J_f^T(b)J_f(b))} \lambda. \quad (8)$$

Обозначим C_H максимальный элемент гессиана отображения (\mathbb{B}, f) :

$$C_H = \sup_{b \in \mathbb{B}, i, j=1, \dots, q} \left\| \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \right\| = \sup_{b \in \mathbb{B}, i, j=1, \dots, q} \sqrt{\sum_{k=1}^p \left(\frac{\partial^2 f_k(b)}{\partial b_i \partial b_j} \right)^2}, \quad (9)$$

где $X = (f_1(b), \dots, f_p(b))^T$. Обозначим

$$C_T = \sup_{b \in \mathbb{B}, i, j, m=1, \dots, q} \left\| \frac{\partial^3 f(b)}{\partial b_i \partial b_j \partial b_m} \right\| = \sup_{b \in \mathbb{B}, i, j=1, \dots, q} \sqrt{\sum_{k=1}^p \left(\frac{\partial^3 f_k(b)}{\partial b_i \partial b_j \partial b_m} \right)^2}. \quad (10)$$

Многообразии \mathbb{M} неизвестно и представлено конечной выборкой

$$\mathbb{X}_N = \{X_1, \dots, X_N\} \subset \mathbb{M} \subset \mathbb{R}^p,$$

состоящей из N точек. Более того, относительно выборки \mathbb{X}_N предполагается:

- S1. Точки из \mathbb{X}_N являются независимыми одинаково распределенными (i.i.d.) с некоторой вероятностной мерой μ на \mathbb{M} такой, что $\mathbb{M} = \text{supp} \mu$;
- S2. Мера μ является непрерывной относительно Римановой меры на многообразии, и её плотность p_μ отделена от 0 и ∞ ;
- S3. Плотность p_μ дважды дифференцируема на \mathbb{M} и её вторые производные ограничены.

Замечание 2. На многообразии \mathbb{M} его вложением в \mathbb{R}^p определена Риманова мера $dV(X)$, которая в главном члене (Лемма 1) совпадает с q -мерным объемом. Пусть $(\Omega, \mathfrak{B}, \mathbb{P})$ — вероятностное пространство, тогда борелевская функция $\mathfrak{B} \rightarrow \mathbb{M}: X = X(\omega)$ называется случайной величиной на многообразии. Будем называть индуцированную на \mathbb{M} меру μ . Если для каждого борелевского

левского множества $\mathfrak{X} \in \mathfrak{B}$

$$\mu(X \in \mathfrak{X}) = \int_{\mathfrak{X}} p_{\mu}(X) dV(X),$$

то функция $p_{\mu}(X)$, $X \in \mathbb{M}$ называется функцией плотности вероятности [41].

Обозначим p_{\min} и p_{\max} максимальное и минимальное значения p_{μ} (существуют по S2):

$$p_{\min} = \inf_{X \in \mathbb{M}} p_{\mu}(X); \quad (11)$$

$$p_{\max} = \sup_{X \in \mathbb{M}} p_{\mu}(X). \quad (12)$$

Введем обозначения для максимальных значений первой и второй производных (существуют по S3):

$$C_{p,1} = \sup_{X \in \mathbb{M}, \theta \in T_X(\mathbb{M}): \|\theta\|=1} \|\nabla_{\theta} p_{\mu}(X)\|; \quad (13)$$

$$C_{p,2} = \sup_{X \in \mathbb{M}, \theta \in T_X(\mathbb{M}): \|\theta\|=1} \|\nabla_{\theta} \nabla_{\theta} p_{\mu}(X)\|, \quad (14)$$

где ∇_{θ} — ковариантная производная (см. приложение А.1), которая является обобщением производной по направлению для многообразий.

Замечание 3. Предположения S1-S2 являются стандартными, гарантирующими соответствие между выборкой \mathbb{X}_N и изучаемым многообразием \mathbb{M} . Предположение S3 полезно для получения равномерных результатов.

О радиусе окрестности $\varepsilon = \varepsilon(N)$ предполагается при $N \rightarrow \infty$:

- P1. При $\varepsilon \rightarrow 0$;
- P2. При $N \cdot \varepsilon^q \rightarrow \infty$;
- P3. $N \cdot \varepsilon^{q+4} \rightarrow 0$.

Замечание 4. Предположение P1 означает, что размер окрестности стремится к нулю, и, следовательно, главным членом разложения функций будет

член с минимальной степенью длины.

Предположение P2 гарантирует попадание бесконечного числа точек выборки в окрестность, несмотря на стремление радиуса окрестности к нулю.

Предположение P3 сильнее P1 и гарантирует, что вклад смещения порядка ε^2 , усредненный по выборке из окрестности, будет бесконечно мал.

1.3 Основные результаты главы

Обозначим $I_\varepsilon(X'|X)$ бернуллиевскую случайную величину являющуюся индикатором события $U_{X,\varepsilon} = \{|X' - X| < \varepsilon\}$, с вероятностью успеха

$$P_\varepsilon = \mu(|X' - X| < \varepsilon) = \mathbb{E}_\mu I_\varepsilon(X'|X), \quad (15)$$

где $|\star|$ — обычная евклидова норма в \mathbb{R}^p .

Обозначим $N_\varepsilon(X)$ — биномиальную случайную величину — число точек выборки \mathbb{X}_N , попавших в p -мерную шаровую окрестность точки $X \in \mathbb{R}^p$ радиуса ε :

$$N_\varepsilon(X) = \sum_{n=1}^N I_\varepsilon(X_n|X). \quad (16)$$

Обозначим V_q — объем q -мерного шара радиуса 1:

$$V_q = \frac{\pi^{q/2}}{\Gamma(q/2 + 1)}, \quad (17)$$

где $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$, $z > 0$ — гамма функция, т.е. $V_{2m} = \frac{\pi^m}{m!}$ и $V_{2m+1} = \frac{2^m \pi^m}{(2m+1)!!}$ для натуральных m .

Ниже приведены основные результаты главы, сформулированные в предположениях M1-M8, S1-S3 и P1-P3.

Утверждение 1 (слабый закон больших чисел). Для любой точки $X \in \mathbb{M}$ выполнено при $N \rightarrow \infty$:

$$\frac{N_\varepsilon(X)}{N \cdot \varepsilon^q} \xrightarrow{P} p_\mu(X) \cdot V_q,$$

где $N_\varepsilon(X)$ — число точек в окрестности из (16) и V_q — объем q -мерного единичного шара (17).

Утверждение 2 (условно равномерное распределение). Для любой точки $X \in \mathbb{M}$ при $N \rightarrow \infty$ и для $\varepsilon = \varepsilon(N)$:

$$\sup_{A \in \mathfrak{A}} \left| \frac{\mu(A)}{\mu(U_{X,\varepsilon})} - \frac{dV(A)}{dV(U_{X,\varepsilon})} \right| \rightarrow 0,$$

где \mathfrak{A} — Борелевская сигма-алгебра на $U_{X,\varepsilon}$, dV — мера Лебега на многообразии.

Утверждение 3 (центральная предельная для $N_\varepsilon(X)$). Для любой точки $X \in \mathbb{M}$ выполнено при $N \rightarrow \infty$:

$$\frac{N_\varepsilon(X) - N\varepsilon^q p_\mu(X) V_q}{\sqrt{N\varepsilon^q p_\mu(X) V_q}} \rightarrow^d N(0, 1),$$

а $N(0, 1)$ — стандартное нормальное распределение.

Теорема 1 (большие отклонения для $N_\varepsilon(X)$). Для любой точки $X \in \mathbb{M}$, удаленной от границы многообразия не менее чем на ε , и $0 < z < 1/16$ выполнено:

$$P \left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \geq z + \varepsilon^2 \cdot \frac{C_E}{V_q p_\mu(X)} \right) \leq \exp(-4z^2 \cdot N\varepsilon^q V_q p_\mu(X)),$$

$$P \left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \geq -z - \varepsilon^2 \cdot \frac{C_E}{V_q p_\mu(X)} \right) \leq \exp(-4z^2 \cdot N\varepsilon^q V_q p_\mu(X)),$$

и, если не требовать гладкость функции плотности, то

$$P \left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q} \leq p_{\min} \cdot (1 - z) - \varepsilon^2 \cdot \frac{C_E \cdot p_{\max}}{V_q p_{\min}} \right) \leq \exp(-4z^2 \cdot N\varepsilon^q V_q p_{\min}^2 / p_{\max}),$$

где C_E — константа, зависящая от многообразия \mathbb{M} , p_{\min}, p_{\max} — минимальное и максимальное значение p_μ на многообразии.

Форма записи вероятности больших отклонений обусловлена тем, что для анализа алгоритмов оценивания многообразий полезны утверждения вида “с большой вероятностью отклонение рассматриваемых статистик от среднего мало”.

Теорема 2 (равномерные большие отклонения). Для достаточно малого $\varepsilon \leq \varepsilon_0$, $0 < z < 1/4$ и всех отделенных от границы точек $X \in \mathbb{M}_\varepsilon$:

$$P \left(\inf_{X \in \mathbb{M}_\varepsilon} \frac{N_\varepsilon(X)}{N\varepsilon^q V_q} \leq p_{\min} \cdot (1 - z) - \varepsilon^2 \cdot \frac{C_E \cdot p_{\max}}{V_q p_{\min}} \right) \leq \left(\frac{6a\sqrt{p}}{\varepsilon} \right)^p \cdot \exp \left(-4z^2 \cdot N\varepsilon^q V_q p_{\min}^2 / (9p_{\max}) \right),$$

где C_E, ε_0 — константы, p_{\min}, p_{\max}, a — константы из (11), (12) и (6).

Замечание 5. Если $\varepsilon(N) \geq C_\alpha \cdot N^{-\alpha}$, $\alpha < 1/q$, $C_\alpha > 0$, то правая часть неравенств Теорем 1.1 и 1.2 стремится к нулю быстрее любой степени $N^{-\beta}$, $\beta > 0$ как произведение полиномиальных и экспоненциальных множителей.

1.4 Некоторые определения и леммы

Необходимые определения и леммы сформулированы и доказаны в данном разделе.

1.4.1 Локальная линеаризация

Окрестность каждой точки $X \in \mathbb{M}$ близка к q -мерному линейному касательному пространству $T_X(\mathbb{M})$ в точке X . Дифференцирование на многообразии немного отличается от дифференцирования в Евклидовых координатах в \mathbb{R}^k . Отличия носят преимущественно технический характер, и основной особенностью является то, что производные определены лишь в направлениях из касательного пространства $T_X(\mathbb{M})$ (см. приложение А.3).

В достаточно малой окрестности $X \in \mathbb{M}$ существует взаимнооднозначное соответствие между точками многообразия \mathbb{M} и элементами касательного пространства $T_X(\mathbb{M})$. В такой окрестности векторы $V \in T_X(\mathbb{M})$ задают координаты. Будем записать V в полярных координатах: $V = t\theta$, $t \in (0, \infty)$ и $\theta \in S_{q-1} = S_{p-1} \cap T_X(\mathbb{M})$, где S_{q-1} — единичная сфера в q -мерном пространстве. Такие координаты называются Римановыми (см. Приложение А.3).

Элемент объема многообразия и q -мерный объем в касательном пространстве связаны следующей леммой:

Лемма 1. (см. [42]). Риманова мера в полярных координатах в окрестности точки $X \in \mathbb{M}$ имеет вид

$$dV(\exp_X(t\theta)) = J(t, \theta) dt d\theta,$$

где для некоторых $\tilde{t} \in [0, t]$, $\tilde{X} \in \{\exp_X \tilde{t}\theta\}$ и $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{M})$:

$$\begin{aligned} J(t, \theta) &= t^{q-1} + t^{q+1} Ric_X(\theta, \theta) + O(t^{q+2}), \\ J(t, \theta) &= t^{q-1} + t^{q+1} Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta}), \end{aligned}$$

где $Ric_X(\theta, \theta)$ — кривизна Риччи (66) в точке X в направлении θ , $\tilde{X} = \exp_X(\tilde{t}\theta)$, $\tilde{t} \in [0, t]$, $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{M})$.

Расстояние между близкими точками многообразия в p -мерном пространстве и расстояние между ними в q -мерных Римановых координатах [14] связано Леммой:

Лемма 2. (см. [43]). Для $X, \tilde{X} \in \mathbb{M}$, таких, что $\tilde{X} = \exp_X(t\theta)$, для малых t (и малых $h = \|X - \tilde{X}\|$):

$$\begin{aligned} t &= h + \frac{1}{24} \|II_X(\theta, \theta)\| \cdot h^3 + O(h^3); \\ t &= h + \frac{1}{24} \|II_{\tilde{X}}(\tilde{\theta}, \tilde{\theta})\| \cdot h^3, \end{aligned}$$

где $\tilde{X} \in \{\exp_X \tilde{t}\theta | \tilde{t} \in [0, t]\}$, $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{M})$, $II_X(V, V)$ — вторая фундаментальная форма многообразия (61).

Значения правых частей Леммы 1 и 2 могут быть ограничены в Предположениях M1-M8:

Лемма 3. Для $X \in \mathbb{M}$ и $\theta \in T_X(\mathbb{M}) \cap S_{p-1}$:

$$\|II_X(\theta, \theta)\| \leq C_{II} \equiv \frac{C_H}{c_J} \cdot q,$$

где C_H — максимальный элемент Гессiana отображения f (9), c_J — минимальное собственное значение метрического тензора \mathbb{M} (7).

Лемма 4. Для $X \in \mathbb{M}$ и $\theta \in T_X(\mathbb{M}) \cap S_{p-1}(1)$:

$$\begin{aligned} \|Ric_X(\theta, \theta)\| &\leq C_{Ric} \equiv 2q^3 \cdot \frac{C_H^2 + C_T \sqrt{C_J}}{c_J^{3/2}} + \\ &+ q^5 \cdot \left(18C_H^2 \cdot \frac{C_J}{c_J} + 4C_H \frac{\sqrt{C_J}}{c_J} \right) / c_J, \end{aligned}$$

где C_H — максимальный элемент Гессиана отображения f (9), c_J и C_J — минимальное и максимальное собственные значения метрического тензора \mathbb{M} (7) и (8), C_T — максимальная норма третьей производной (10).

Доказательства Лемм 3 и 4 приведены в Приложении А.3.

1.4.2 Леммы Муавра-Лапласа для медленно убывающей вероятности

В работе используются локальные и интегральные леммы Муавра-Лапласа для медленно стремящейся к нулю вероятности успеха в схеме серий. В классической формулировке, вероятность предполагается фиксированной. Однако, доказательство почти не изменится, если предположить, что параметр p_N убывает с увеличением размера выборки N , но $p_N \cdot N \rightarrow \infty$ при $N \rightarrow \infty$. Единственное отличие состоит в разложении функций по малому параметру $\frac{1}{p_N \cdot N}$, вместо $\frac{1}{N}$.

Лемма 5. (Локальная Муавра-Лапласа для медленно убывающей вероятности успеха). Пусть вероятность успеха в схеме Бернулли $p_N > 0$ зависит от размера выборки N и $q_N = 1 - p_N$. Также пусть $p_N \cdot N \rightarrow \infty$ при $N \rightarrow \infty$. Тогда при $N \rightarrow \infty$ для числа успехов k , такого что $\frac{|k - Np_N|}{(Np_Nq_N)^{2/3}} \rightarrow 0$:

$$P_N(k) \sim \frac{1}{\sqrt{2\pi Np_Nq_N}} \exp\left(\frac{-(k - Np_N)^2}{2Np_Nq_N}\right),$$

то есть

$$\sup_{\{k: |k - Np_N| \leq \phi(Np_Nq_N)\}} \left| \frac{P_N(k)}{\frac{1}{\sqrt{2\pi Np_Nq_N}} \exp\left(\frac{-(k - Np_N)^2}{2Np_Nq_N}\right)} - 1 \right| \rightarrow 0,$$

где $\phi(Np_Nq_N)$ — любая неотрицательная функция такая, что $\phi(Np_Nq_N) =$

$o((Np_Nq_N)^{2/3})$.

Следствие 1. Утверждение Леммы 5 можно переформулировать следующим образом: для всех $z \in \mathbb{R}$ таких, что $z = o((Np_Nq_N)^{1/6})$, а $Np + z\sqrt{Np_Nq_N}$ — целые числа из множества $\{1, \dots, N\}$,

$$p(Np + z\sqrt{Np_Nq_N}) \sim \frac{1}{\sqrt{2\pi Np_Nq_N}} e^{-z^2/2},$$

то есть при $N \rightarrow \infty$

$$\sup_{\{z: |z| \leq \psi(N)\}} \left| \frac{p_N(Np_N + z\sqrt{Np_Nq_N})}{\frac{1}{\sqrt{2\pi Np_Nq_N}} e^{-z^2/2}} - 1 \right| \rightarrow 0,$$

где $\psi(N) = o((Np_Nq_N)^{1/6})$.

Лемма 6. (интегральная Муавра-Лапласа для убывающей вероятности успеха). Пусть вероятность успеха в схему Бернулли $p_N > 0$ зависит от числа испытаний N , и $q_N = 1 - p_N$. При этом $p_N \cdot N \rightarrow \infty$ при $N \rightarrow \infty$. Обозначим $P_N(k) = C_N^k p_N^k q_N^{N-k}$, $P_N(a, b) = \sum_{a < z \leq b} P_N(Np_N + z\sqrt{Np_Nq_N})$. Тогда

$$\sup_{-\infty \leq a < b \leq \infty} \left| P_N(a, b) - \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-z^2/2) dz \right| \rightarrow 0, \quad n \rightarrow \infty.$$

Леммы 5 и 6 доказаны в Приложении В.1.

1.4.3 Вероятность больших отклонений для ограниченных случайных величин

Для оценивания вероятности больших отклонений будем использовать следующую лемму, доказанную в Приложении В.2:

Лемма 7. Пусть χ_1, \dots, χ_N — независимые одинаково распределенные случайные величины, $\mathbb{E}\chi_1 < \infty$, $|\chi_1 - \mathbb{E}\chi_1| \leq C < \infty$ и константы m_1, \dots, m_N такие, что $\max_{k=1, \dots, N} |m_k - \mathbb{E}\chi_k| \leq m$. Обозначим $\bar{\chi} = \frac{1}{N} \sum_{k=1}^N \chi_k - m_k$ и $\sigma^2 = \mathbb{E}(\chi_1 - \mathbb{E}\chi_1)^2$. Тогда для $H \geq 2C$ при $0 \leq x \leq \frac{1}{H}$:

$$\begin{aligned} P(\bar{\chi} \geq x \cdot \sigma^2 + m) &\leq \exp(-x^2 \sigma^2 \cdot N/4), \\ P(\bar{\chi} \leq -x \cdot \sigma^2 - m) &\leq \exp(-x^2 \sigma^2 \cdot N/4) \end{aligned}$$

и при $x \geq \frac{1}{H}$:

$$\begin{aligned} P(\bar{\chi} \geq x \cdot \sigma^2 + m) &\leq \exp(-x\sigma^2 N/4H), \\ P(\bar{\chi} \leq -x \cdot \sigma^2 - m) &\leq \exp(-x\sigma^2 N/4H). \end{aligned}$$

1.4.4 Замена области интегрирования

Обозначим $B_\varepsilon(X)$ — пересечение евклидовой ε -окрестности точки $X \in \mathbb{M}$ как точки p -мерного пространства с многообразием \mathbb{M} :

$$B_\varepsilon(X) = \{\tilde{X} | \tilde{X} \in \mathbb{M} \cap |\tilde{X} - X| < \varepsilon\}. \quad (18)$$

Обозначим $\tilde{B}_\varepsilon(X)$ — ε окрестность в локально-римановых координатах с центром в $X \in \mathbb{M}$:

$$\tilde{B}_\varepsilon(X) = \{\tilde{X} | \exists V \in T_X(\mathbb{M}) : \tilde{X} = \exp_X(V) \cap |V| < \varepsilon\}. \quad (19)$$

Обозначим \mathbb{M}_ε — множество внутренних точек, отстоящих от границы не менее чем на ε :

$$\mathbb{M}_\varepsilon = \{X \in \mathbb{M} | \forall V \in T_X(\mathbb{M}) \cap |V| < \varepsilon : \exp_X(V) \in \mathbb{M}\} \quad (20)$$

Лемма 8. Для произвольной ограниченной функции $g(X, \tilde{X}) = \tilde{g}(X, t, \theta)$, $\varepsilon \leq C_{int}$ и $X \in \mathbb{M}_\varepsilon$, $\tilde{X} \in \mathbb{M}$, $\theta \in T_X(\mathbb{M}) \cap S_{p-1}$ выполнено

$$\begin{aligned} \left| \int_{B_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) \right| &\leq \\ &\leq 8 \cdot V_q \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \cdot \varepsilon^{q+2}, \end{aligned}$$

где $c(\mathbb{M})$ — число обусловленности (M7), V_q — объем q -мерного шара (17), C_{II} и C_{Ric} — константы из Лемм 3 и 4, ограничивающие вторую квадратичную форму и кривизну Риччи,

$$C_{int} = \min \left\{ \frac{1}{c(\mathbb{M})}, \sqrt{\frac{24 \cdot (\sqrt[3]{2} - 1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}} \right\} \quad (21)$$

Лемма доказана в разделе В.3.

1.4.5 Конечные сети

Для доказательства равномерности оценок будем использовать дополнительное построение. δ -сеть метрического пространства \mathbb{Z} есть множество $\mathbb{Z}_{net}(\delta) \subset \mathbb{Z}$ такое, что для любой точки $Z \in \mathbb{Z}$ найдётся точка $Z_{net} \in \mathbb{Z}_{net}(\delta)$, удалённая от Z не более чем δ .

Так как множество \mathbb{M}_ε (19) точек, удаленных от границы не менее чем на ε , является подмножеством \mathbb{M} и многообразием для которого также выполнены M2 и M4, то применяя к нему лемму 11, получаем

Следствие 2. (Следствие леммы 11). Для каждого $\delta > 0$ и $\varepsilon > 0$ найдется конечная δ -сеть \mathbb{M}_ε , содержащая не более $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ элементов, где $a > 0$ — ребро описанного вокруг \mathbb{M} p -мерного гиперкуба.

1.5 Доказательство основных теорем для попадания в окрестность

Обозначим χ индикатор попадания случайной точки \tilde{X} , распределенной плотностью p_μ с многообразия в p -мерную ε -окрестность точки $X \in \mathbb{X}$:

$$\chi = \chi(\tilde{X}|X, \varepsilon) = I(|X - \tilde{X}| \leq \varepsilon), \quad (22)$$

где $I(A)$ — индикатор события A (то есть 1, если событие наступило; и 0 иначе). Заметим, что $N_\varepsilon(X)$ (16) — биномиальная случайная величина:

$$N_\varepsilon(X) = \sum_{n=1}^N \chi(X_n|X, \varepsilon). \quad (23)$$

Оценим первые два начальных момента χ .

Лемма 9. Пусть $\varepsilon \leq \min \left\{ \frac{1}{c(\mathbb{X})}, \sqrt{\frac{24 \cdot (\sqrt[4]{2}-1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}} \right\}$, $X \in \mathbb{X}_\varepsilon$.

Тогда для центральных моментов случайной величины χ (22) выполнено

$$\mathbb{E}\chi^2 = \mathbb{E}\chi = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2 \delta_{\mathbb{E}}),$$

где

$$|\delta_{\mathbb{E}}| \leq 8 \cdot p_{\max} + C_{Ric} + \varepsilon^2 \cdot \left((C_{Ric} \cdot C_{p,1} + C_{p,2}) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \right), \quad (24)$$

а константы V_q , p_{\min} , p_{\max} , C_{II} , C_{Ric} , $C_{p,1}$ и $C_{p,2}$ определены в (17), (11), (12), Лемме 3, Лемме 4, (13) и (14) соответственно.

Также (без использования S3):

$$\mathbb{E}\chi^2 = \mathbb{E}\chi \geq \varepsilon^q \cdot V_q \cdot p_{\min} - V_q \cdot \varepsilon^{q+2} \cdot (C_{Ric} + 8p_{\max}).$$

Доказательство. Заметим, что $\chi^2 = \chi$. Поэтому $\mathbb{E}\chi^2 = \mathbb{E}\chi$. Пусть точка $\tilde{X} \in \mathbb{X}$ выбирается из многообразия случайно с плотностью $p_{\mu}(\tilde{X})$. Для математического ожидания χ выполнено $\mathbb{E}\chi = \int_{B_{\varepsilon}(X)} p_{\mu}(\tilde{X}) dV(\tilde{X})$. По Лемме 8 для функции $p_{\mu}(\tilde{X})$, учитывая (12):

$$\left| \int_{B_{\varepsilon}(X)} p_{\mu}(\tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_{\varepsilon}(X)} p_{\mu}(\tilde{X}) dV(\tilde{X}) \right| \leq \leq 8 \cdot V_q \cdot p_{\max} \cdot \varepsilon^{q+2}.$$

Далее, учитывая разложение

$$p_{\mu}(\tilde{X}) = p(X) + t \cdot \nabla_{\theta} p(X) + \frac{t^2}{2} \nabla_{\tilde{\theta}} \nabla_{\tilde{\theta}} p(\tilde{X}), \tilde{X} \in \tilde{B}_{\varepsilon}(X)$$

и симметричность области $\tilde{B}_\varepsilon(X)$

$$\begin{aligned}
& \left| \int_{\tilde{B}_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} p_\mu(X) dV(\tilde{X}) \right| = \\
& = \left| \int_{\tilde{B}_\varepsilon(X)} \left(p_\mu(\tilde{X}) - p_\mu(X) \right) dV(\tilde{X}) \right| = \\
& = \left| \int_{S^{q-1}} \int_0^\varepsilon \left(t \cdot \nabla_\theta p(X) + \frac{t^2}{2} \nabla_{\tilde{\theta}} \nabla_{\tilde{\theta}} p(\tilde{X}) \right) \cdot \left(t^{q-1} + t^{q+1} Ric_{\tilde{X}_2}(\tilde{\theta}, \tilde{\theta}) \right) dt d\theta \right| \leq \\
& \leq \left| \int_{S^{q-1}} \int_0^\varepsilon t^q \cdot \nabla_\theta p(X) dt d\theta \right| + \\
& \left| \varepsilon^4 \cdot (C_{Ric} \cdot C_{p,1} + C_{p,2}) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q-1} dt d\theta \right| + \\
& \left| \varepsilon^6 \cdot (C_{Ric} \cdot C_{p,2}) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q-1} dt d\theta \right| = \\
& = 0 + V_q \cdot \varepsilon^{q+4} \cdot \left((C_{Ric} \cdot C_{p,1} + C_{p,2}) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \right).
\end{aligned}$$

Далее

$$\begin{aligned}
& \left| \int_{\tilde{B}_\varepsilon(X)} p_\mu(X) dV(\tilde{X}) - V_q \cdot \varepsilon^q p_\mu(X) \right| = \\
& = \left| p_\mu(X) \cdot \int_{S^{q-1}} \int_0^\varepsilon t^{q+1} Ric_{\tilde{X}_2}(\tilde{\theta}, \tilde{\theta}) dt d\theta \right| \leq \\
& \leq p_\mu(X) \cdot \varepsilon^{q+2} \cdot V_q \cdot C_{Ric}.
\end{aligned}$$

Таким образом, если плотность достаточно гладкая (выполнено S3):

$$\mathbb{E}\chi = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2 \delta_{\mathbb{E}}),$$

где

$$|\delta_{\mathbb{E}}| \leq 8 \cdot p_{\max} + C_{Ric} + \varepsilon^2 \cdot \left((C_{Ric} \cdot C_{p,1} + C_{p,2}) + \varepsilon^2 \cdot C_{Ric} \cdot C_{p,2} \right).$$

Если не использовать S3, то $\mathbb{E}\chi$ можно оценить снизу:

$$\begin{aligned}
\int_{B_\varepsilon(X)} p_\mu(\tilde{X}) dV(\tilde{X}) & \geq \int_{\tilde{B}_\varepsilon(X)} p_{\min} dV(\tilde{X}) - 8 \cdot V_q \cdot p_{\max} \cdot \varepsilon^{q+2} \geq \\
& \geq \varepsilon^q \cdot V_q \cdot p_{\min} - V_q \cdot \varepsilon^{q+2} \cdot (C_{Ric} + 8p_{\max}).
\end{aligned}$$

Лемма доказана. □

Следствие 3. Если

$$\varepsilon \leq \min \left\{ \frac{1}{c(\mathbb{X})}, \sqrt{\frac{24(\sqrt[q]{2} - 1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}}, \frac{1}{\sqrt{C_{Ric}C_{p,2}}}, \frac{1}{\sqrt{C_{Ric}C_{p,1} + C_{p,2} + 1}} \right\}$$

и $X \in \mathbb{X}_\varepsilon$, то в лемме 9 для $\delta_{\mathbb{E}}$ (24) выполнено

$$|\delta_{\mathbb{E}}| \leq C_{\mathbb{E}} \equiv 8 \cdot p_{\max} + C_{Ric} + 1.$$

Следствие 4. Если

$$\varepsilon \leq \min \left\{ \frac{1}{c(\mathbb{X})}, \sqrt{\frac{24(\sqrt[q]{2} - 1)}{\max\{1, C_{II}\}}}, \frac{1}{2\sqrt{C_{Ric}}}, \frac{1}{\sqrt{C_{Ric}C_{p,2}}}, \frac{1}{\sqrt{C_{Ric}C_{p,1} + C_{p,2} + 1}} \right\},$$

и $X \in \mathbb{X}_\varepsilon$,

$$\begin{aligned} \text{Var}\chi &\leq \varepsilon^q V_q \cdot (p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4C_{\mathbb{E}}^2); \\ \text{Var}\chi &\geq \varepsilon^q V_q \cdot (p_\mu(X) - C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q - \varepsilon^{q+2}C_{\mathbb{E}} - \varepsilon^4C_{\mathbb{E}}^2). \end{aligned}$$

Доказательство. Преобразуем, учитывая Лемму 9 и Следствие 3:

$$\begin{aligned} \text{Var}\chi &= \mathbb{E}\chi^2 - (\mathbb{E}\chi)^2 = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}}) \cdot (1 - \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}})) \leq \\ &\leq \varepsilon^q V_q \cdot (p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4C_{\mathbb{E}}^2); \\ \text{Var}\chi &= \mathbb{E}\chi^2 - (\mathbb{E}\chi)^2 = \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}}) \cdot (1 - \varepsilon^q \cdot V_q \cdot (p_\mu(X) + \varepsilon^2\delta_{\mathbb{E}})) \geq \\ &\geq \varepsilon^q V_q \cdot (p_\mu(X) - C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q - \varepsilon^{q+2}C_{\mathbb{E}} - \varepsilon^4C_{\mathbb{E}}^2). \end{aligned}$$

Следствие доказано. □

Доказательство Утверждения 1. Учитывая представление (23):

$$\mathbb{E}N_\varepsilon(X) = N \cdot \mathbb{E}\chi.$$

Пользуясь открытостью \mathbb{B} (предположение M1), из следствия 3 для достаточно

малых ε :

$$\left| \frac{N_\varepsilon(X)}{N} - \varepsilon^q V_q p_\mu(X) \right| \leq \varepsilon^{q+2} V_q C_{\mathbb{E}}.$$

Воспользуемся неравенством Чебышева для случайной величины $\frac{N_\varepsilon(X)}{N}$ и порога $\varepsilon^q \cdot \psi(N)$, где $\psi(N) = (N\varepsilon^q)^{-1/4}$ (можно взять произвольное $\psi(N) : N\varepsilon^q \cdot \psi(N)^2 \rightarrow \infty$ и $\psi(N) \rightarrow 0$ при $N \rightarrow \infty$):

$$P \left(\left| \frac{N_\varepsilon(X)}{N} - \mathbb{E}\chi \right| \geq \varepsilon^q \psi(N) \right) \leq \frac{\text{Var}\chi}{N\varepsilon^{2q}\psi(N)^2}.$$

Следовательно

$$P \left(\left| \frac{N_\varepsilon(X)}{N} - \varepsilon^q V_q p_\mu(X) \right| \geq \varepsilon^{q+2} V_q C_{\mathbb{E}} + \varepsilon^q \right) \leq \frac{\text{Var}\chi}{N\varepsilon^{2q}\psi(N)^2}.$$

Заметим, что при $N \rightarrow \infty$

$$\frac{\text{Var}\chi}{N\varepsilon^{2q}\psi(N)^2} \leq \frac{V_q \cdot (p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4C_{\mathbb{E}}^2)}{N\varepsilon^q\psi(N)^2} \rightarrow 0;$$

$$\varepsilon^{q+2}V_qC_{\mathbb{E}} + \psi(N) \rightarrow 0.$$

Таким образом,

$$\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \xrightarrow{p} 0.$$

Утверждение доказано. □

Доказательство Утверждения 2. $\forall A \in \mathfrak{X} \cap U_{X,\varepsilon}$:

$$\begin{aligned} |\mu(A) - p_\mu(X) \cdot dV(A)| &= \left| \int_A p_\mu(X') dV(X') - \int_A p_\mu(X) dV(X') \right| = \\ &= \left| \int_A (p_\mu(X') - p_\mu(X)) dV(X') dV(X') \right| \leq \varepsilon \cdot C_{p,1} \cdot dV(U_{X,\varepsilon}). \end{aligned}$$

Аналогично,

$$\begin{aligned} |\mu(A) - p_\mu(X) \cdot dV(A)| &= \left| \int_A p_\mu(X') dV(X') - \int_A p_\mu(X) dV(X') \right| = \\ &= \left| \int_A (p_\mu(X') - p_\mu(X)) dV(X') dV(X') \right| \leq \varepsilon \cdot C_{p,1} \cdot \mu(U_{X,\varepsilon}). \end{aligned}$$

При $N \rightarrow \infty$:

$$\begin{aligned} & \left| \frac{\mu(A)}{\mu(U_{X,\varepsilon})} - \frac{dV(A)}{dV(U_{X,\varepsilon})} \right| = \\ &= \left| \frac{(\mu(U_{X,\varepsilon}) - p_\mu(X) dV(U_{X,\varepsilon})) dV(A) + (\mu(A) - p_\mu(X) dV(A)) dV(U_{X,\varepsilon})}{\mu(U_{X,\varepsilon}) dV(U_{X,\varepsilon})} \right| \leq \\ & \leq \varepsilon \cdot C_{p,1} \cdot \left(\frac{dV(A)}{dV(U_{X,\varepsilon})} + \frac{\mu(A)}{\mu(U_{X,\varepsilon})} \right) \leq 2\varepsilon \cdot C_{p,1} \rightarrow 0. \end{aligned}$$

Утверждение доказано. □

Доказательство Утверждения 3. Так как χ (22) — бернуллевская случайная величина, то, учитывая лемму 9 и ее следствия, получаем, что для χ выполнены условия леммы 6: $\mathbb{E}\chi \rightarrow 0$ и $N \cdot \mathbb{E}\chi \rightarrow \infty$ при $N \rightarrow \infty$.

Остается заметить, что при $N \rightarrow \infty$

$$\begin{aligned} \frac{\mathbb{E}\chi}{\varepsilon^q V_q p_\mu(X)} &\rightarrow 1; \\ \frac{\text{Var}\chi}{\varepsilon^q V_q p_\mu(X)} &\rightarrow 1. \end{aligned}$$

Откуда, по лемме 9:

$$\frac{N_\varepsilon(X) - N\varepsilon^q p_\mu(X) V_q}{\sqrt{N\varepsilon^q p_\mu(X) V_q}} \xrightarrow{d} N(0, 1).$$

Утверждение доказано. □

Доказательство Теоремы 1. Проверим выполнение условий Леммы 7 для χ (22), $m_k = \varepsilon^q \cdot V_q \cdot p_\mu(X)$, $m = \varepsilon^{q+2} C_{\mathbb{E}}$ (3) и $C = 1$:

$$\begin{aligned} |\chi - \mathbb{E}\chi| &\leq 1; \\ |m_k - \mathbb{E}\chi_k| &\leq \varepsilon^{q+2} C_{\mathbb{E}}. \end{aligned}$$

Для $\varepsilon < \min \left\{ \frac{1}{\sqrt{4C_{\mathbb{E}}p_{\min}}}, \sqrt[4]{\frac{1}{4p_{\min}}} \right\}$ по следствию 4:

$$\begin{aligned} \text{Var}\chi &\leq \varepsilon^q V_q \cdot (p_\mu(X) + C_{\mathbb{E}}\varepsilon^2 - p_\mu(X)^2\varepsilon^q + \varepsilon^{q+2}C_{\mathbb{E}} + \varepsilon^4 C_{\mathbb{E}}^2) \leq \\ &\varepsilon^q V_q \cdot p_\mu(X) (1 + 1/4 + 1/4 + 1/4) < 2\varepsilon^q V_q \cdot p_\mu(X); \\ \text{Var}\chi &\geq 1/4 \cdot \varepsilon^q V_q \cdot p_\mu(X). \end{aligned}$$

Таким образом, условия леммы 7 выполнены, и для $z \leq 1/16$:

$$\begin{aligned} P \left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \geq z + \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu} \right) &\leq \exp(-z^2 \cdot N\varepsilon^q V_q p_\mu(X)/4); \\ P \left(\frac{N_\varepsilon(X)}{N\varepsilon^q V_q p_\mu(X)} - 1 \leq -z - \varepsilon^2 \cdot \frac{C_{\mathbb{E}}}{V_q p_\mu} \right) &\leq \exp(-z^2 \cdot N\varepsilon^q V_q p_\mu(X)/4) \end{aligned}$$

Аналогично, если не выполнено S3, то заменяя оценки с разложением плотности по формуле Тейлора на их грубые аналоги: *Теорема доказана.* \square

Доказательство Теоремы 2. Ввиду разрывности индикаторной функции, $N_\varepsilon(X)$ является разрывной функцией. Однако, ввиду вложенности шаров $\tilde{B}_{\varepsilon/3}(X) \subset \tilde{B}_\varepsilon(\tilde{X})$ для $X \in \tilde{B}_{\varepsilon/3}(\tilde{X})$ для малых ε :

$$N_\varepsilon(\tilde{X}) \geq N_{\varepsilon/3}(X).$$

Откуда, по леммам 11, 12 и Теореме 1, получаем условие теоремы. \square

2 Непараметрическое оценивание на многообразии

2.1 Рассматриваемые статистики

Рассматриваемые локальные статистики для фиксированной точки X имеют вид

$$F_N(X) = \frac{\sum_{n=1}^N K_\varepsilon(X, X_n) \cdot F(X, X_n)}{\varepsilon^d \cdot \sum_{n=1}^N K_\varepsilon(X, X_n)}, \quad (25)$$

где функция $F(X, X_n)$ определяется решаемой статистической задачей, а ядерная функция (ядро) $K_\varepsilon(X, X_n)$ играет роль веса для слагаемого $F(X, X_n)$ та-

кого, $K_\varepsilon(X, X_n) > 0$ только для точек X_n из ε -окрестности $U_{X,\varepsilon}$ точки X , здесь $d \in \mathbb{N} \cup \{0\}$ — числовой параметр, определяемый поведением функции $F(X, X_n)$ при $X_n \rightarrow X$.

В качестве глобальных статистик рассматриваются усреднения локальных статистик (25)

$$F_N = \frac{1}{N} \sum_{n=1}^N F_N(X_n). \quad (26)$$

по всем точкам выборки.

Примеры ядерных функций $K_\varepsilon(X, X')$, используемые в различных статистиках при *моделировании многообразий*:

- индикатор $I(X' \in U_{X,\varepsilon})$ в алгоритмах Local Linear Embedding [17] и Local Tangent Space Alignment [18];
- heat-ядро в алгоритме Laplacian Eigenmaps [19] $\exp(-|X' - X|^2/T) \cdot I(X' \in U_{X,\varepsilon})$, где $T > 0$ — параметр “температуры”;
- ядро Епанечникова в алгоритме Vector Diffusion Maps [14] $(1 - |X' - X|^2/\varepsilon^2) \cdot I(X' \in U_{X,\varepsilon})$;
- ядро $K_{GSE}(X, X')$ в алгоритме Grassmann&Stiefel Eigenmaps [21], которое зависит не только от расстояния $|X - X'|$, но и от расстояния Бине-Коши между касательными пространствами $T_X(\mathbb{M})$ и $T_{X'}(\mathbb{M})$ к \mathbb{M} в точках X и X' .

При фиксированной точке X , представим точку $X' \in U_{X,\varepsilon}$ через её локально римановы координаты: $X' = \exp_X(t\theta)$, где $t \in [0, \infty)$ и $\theta \in S_{q-1} \subset T_X(\mathbb{M})$, где S_{q-1} — единичная сфера в q -мерном пространстве. В работе введено параметрическое семейство гладких финитных ядер $\{K(X, \theta, t), t \geq 0, \theta \in S_{q-1}\}$, удовлетворяющих введенным в диссертации предположениям $K1 - K4$, и рассматриваются ядерные функции $K_\varepsilon(X, X')$ представимые в виде

$$K_\varepsilon(X, X') = K_\varepsilon(X, \exp_X(t\theta)) = K(X, \theta, t/\varepsilon). \quad (27)$$

В диссертации, как и в других теоретических исследованиях статистик на многообразиях, предполагается, что ядра являются гладкими функциями своих аргументов. Однако в приведенных выше примерах ядра хотя и допускают

представление (27) но являются разрывными. Однако эти ядра можно заменить их “сглаженными” аналогами, с сохранением всех других процедур в использующих их алгоритмах, что обычно мало влияет на основные характеристики алгоритмов [44]. Примером такого удовлетворяющего предположениям диссертации гладкого финитного ядра является ядро $K(X, \theta, t) = \exp\left(\frac{1}{z^2-1}\right) \cdot I(t \in (-1, 1))$.

Таким образом, относительно ядерной функции предполагается $K_\varepsilon(X, X')$ для $X' = \exp_X t\theta \in U_{X,\varepsilon}$ предполагается:

- К1. $K_\varepsilon(X, X') = K(X, \theta, \frac{|X-X'|}{\varepsilon})$, где $K(X, \theta, z)$ — фиксированная неотрицательная симметричная по θ и z ограниченная функция с ограниченной производной;
- К2. Функция $K(X, \theta, z)$ — финитная по z : $K(X, \theta, z) = 0$ для $z \geq 1$;
- К3. Существует положительные $\rho_{E,0}(X, \theta)$ (35) и $\rho_{V,0}(X, \theta)$ (39);
- К4. Функция $K(X, \theta, z)$ Липшицева по X с константой \tilde{C}_K ;
- К5. $\inf_{X \in \mathbb{M}} \int_{S_q} \int_{1/2}^1 K(X, \theta, z) z^{q-1} d\theta dt \geq C_{K,1/2}$.

Обозначим

$$C_K = \sup_{X, \theta, z} K(X, \theta, z); \quad (28)$$

$$C_{K,1} = \sup_{X, \theta, z} \left| \frac{\partial K(X, \theta, z)}{\partial z} \right|. \quad (29)$$

Примеры функций $F(X, X')$ используемых в локальных статистиках вида (25):

1. $F(X, X') = \psi(X')$, тогда статистика $F_N(X)$ (25) является непараметрической оценкой [45] значения $\psi(X)$ в точке X неизвестной заданной на многообразии функции ψ по её известным значениям $\psi(X')$ в точках выборки $X' \in \mathbb{X}_N$;
2. $F(X, X') = (X' - X) \cdot (X' - X)^T$, $X, X' \in \mathbb{R}^p$, тогда $p \times p$ матрица $F_N(X)$ (25) является оценкой ковариационной матрицы случайной точки $X' \in \mathbb{M}$ попавшей в окрестность $U_{X,\varepsilon}$ точки $X \in \mathbb{M}$. Эта оценка используется в локальном методе главных компонент для построения оценки касательного пространства $T_X(\mathbb{M})$ к многообразию \mathbb{M} в точке X^{14} ;

3. $F(X, X') = \left| \psi(X) - \psi(\tilde{X}) \right|^2$, где ψ — скалярная или векторная функция, тогда локальная статистика $F_N(X)$ (25) используется в процедурах снижения размерности^{19,21}.

Параметр d определяется порядком малости функции $F(X, X') = \Phi(X, \theta, t) = O(t^d)$ при $t \rightarrow 0$:

$$\Phi(X, \theta, t) = t^d \cdot \phi(X, \theta, t),$$

где $\phi \not\equiv 0$ и ограничена, d не зависит от точки X и θ .

Таким образом, относительно функции $F(X, X')$ для $X' = \exp_X t\theta \in U_{X,\varepsilon}$ предполагается:

- F1. Существует равномерный по X порядок $d \in \mathbb{N} \cup \{0\}$ главного члена разложения $\tilde{F}(X, \theta, t)$ по t при $t \rightarrow 0$;
- F2. Справедливо представление

$$\Phi(X, \theta, t) = t^d \cdot \varphi_0(X, \theta, 0) + t^{d+1} \cdot \varphi_1(X, \theta, 0) + t^{d+2} \cdot \varphi_2(X, \theta, \tilde{t})$$

при $t \rightarrow 0$ и $\tilde{t} \in [0, t]$, где $\varphi_0(X, \theta, t)$ и $\varphi_1(X, \theta, t)$ — гладкие и ограниченные, а $\varphi_2(X, \theta, t)$ — ограничена и Липшицева.

Обозначим

$$C_\varphi = \sup_{X \in \mathbb{M}, t, \theta} |\varphi(X, \theta, t)|; \quad (30)$$

$$C_{\varphi,1} = \sup_{X \in \mathbb{M}, t, \theta} |\varphi_1(X, \theta, t)|; \quad (31)$$

$$C_{\varphi,2} = \sup_{X \in \mathbb{M}, t, \theta} |\varphi_2(X, \theta, t)|. \quad (32)$$

2.2 Основные результаты главы

Обозначим

$$\bar{F}(X) = \frac{\int_{S^{q-1}} \rho_{E,d}(X, \theta) \tilde{F}(X, \theta, 0) d\theta}{\int_{S^{q-1}} \rho_{E,0}(X, \theta) d\theta}, \quad (33)$$

и

$$\bar{F} = \int_{\mathbb{M}} \bar{F}(X) d\mu(X), \quad (34)$$

где

$$\rho_{V,m}(X, \theta) = \int_0^1 K(X, \theta, t) t^{m+q-1} dt. \quad (35)$$

Ниже приведены основные результаты главы, сформулированные в предположениях M1-M10, S1-S3, P1-P3, F1-F2 и K1-K5.

Утверждение 4 (состоятельность $F_N(X)$). Для любой точки $X \in \mathbb{M}$ выполнено при $N \rightarrow \infty$:

$$F_N(X) \rightarrow^p \bar{F}(X).$$

.

Теорема 3 (асимптотическое разложение $F_N(X)$). Для любой точки $X \in \mathbb{M}$ выполнено при $N \rightarrow \infty$:

$$\sqrt{N\varepsilon^q} \cdot (F_N(X) - \bar{F}(X)) \rightarrow^d N(m(X), d(X)),$$

где $m(X)$ и $d(X)$ — функции (42), (41), зависящая от многообразия, меры, функции и ядра, причем $m(X) \equiv 0$ при $d:2$.

Теорема 4 (вероятность больших уклонений $F_N(X)$). Для любой удаленной от его края точки многообразия $X \in \mathbb{M}_\varepsilon$, для любого $N > N_0$ и $\varepsilon < \varepsilon_0$ выполнено:

$$\begin{aligned} P \left(|F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2} \right) &\leq \\ &\leq 4 \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{\sigma^2(X)} \right), \end{aligned}$$

где $N_0, \varepsilon_0, C_{LD1}, C_{LD2}$ — положительные константы (43), (45), (46), (47),

$$\sigma^2(X) = p_\mu(X) \cdot \int_{S_{q-1}} \rho_{D,2d}(X, \theta) (1 + \varphi(X, \theta, 0))^2 d\theta, \quad (36)$$

и $\sigma^2(X) \leq C_\sigma$, где C_σ — положительная константа (44).

Теорема 5 (равномерная вероятность больших уклонений $F_N(X)$). Для любого $z \in [0, 1]$, для $N > N_0$, $\varepsilon < \tilde{\varepsilon}_0$ для каждой точки многообразия $X \in \mathbb{M}_\varepsilon$, удаленной от его границы, выполнено

$$\begin{aligned} P \left(\sup_{X \in \mathbb{M}_\varepsilon} |F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot \tilde{C}_{LD2} \right) &\leq \\ &\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3} \right)^p \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma} \right) + \\ &+ \left(\frac{6a\sqrt{p}}{\varepsilon} \right)^p \cdot \exp \left(-1/16 \cdot N\varepsilon^q V_{qp_{\min}}^2 / (9p_{\max}) \right). \end{aligned}$$

$N_0, \varepsilon_0, C_{LD1}$ — положительные константы (43), (45), (46), \tilde{C}_{LD2} — положительная константа, определенная в доказательстве.

Утверждение 5 (состоятельность F_N). При $N \rightarrow \infty$:

$$F_N \rightarrow^p \bar{F}.$$

Теорема 6 (вероятность больших уклонений F_N). Найдутся положительные $N_{U,0}, C_{U,1}, C_{U,2}, C_{U,p}$ такие, что для любого $z \in [0, 1]$ и $N > N_{U,0}$:

$$P \left(|F_N - \bar{F}| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{U,1} + \varepsilon^2 \cdot C_{U,2} \right) \leq \exp \left(-z^2 \cdot N\varepsilon^q \cdot C_{U,p} \right).$$

2.3 Доказательство основных теорем главы

В доказательстве будем использовать следующую лемму

Лемма 10 (моменты случайных величин). Пусть $X \in \mathbb{M}_\varepsilon$, $\varepsilon \leq C_{int}$ (21) и точка $X' \in \mathbb{M}$ выбрасывается случайно в s с плотностью $p_\mu(X')$. Тогда для центральных моментов случайной величины $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$ выполнено

$$\begin{aligned} \mathbb{E}\xi &= p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta + \\ &+ \varepsilon \cdot E_1(X) \cdot I(d \neq 2) + \varepsilon^2 \cdot \alpha_E(X) \cdot C_{E,2}; \end{aligned} \quad (37)$$

$$\begin{aligned} \text{Var}\xi = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta + \\ + \varepsilon^{\min\{-q+2,0\}} \cdot \alpha_V(X) \cdot C_{V,2}, \end{aligned} \quad (38)$$

где

$$\begin{aligned} |\alpha_E(X)| &\leq 1; \\ |\alpha_V(X)| &\leq 1; \\ E_1(X) &\leq C_{E,1}, \end{aligned}$$

$E_1(X)$ определена в (71), $C_{E,1}, C_{E,2}, C_{V,2}$ — константы (72), (73), (74) соответственно, $I(d \not\equiv 2)$ — индикатор нечетности d , и

$$\rho_{V,m}(X, \theta) = \int_0^1 K(X, \theta, t)^2 t^{m+q-1} dt. \quad (39)$$

Доказательство. Лемма доказана в разделе В.5. □

Следствие 5 (Из Леммы 10). Пусть $X \in \mathbb{M}_\varepsilon$,

$\varepsilon \leq C_{int}$ (21) и точка $X' \in \mathbb{M}$ выбрасывается случайно в S плотностью $p_\mu(X')$. Тогда для центральных моментов случайных величин

$$\begin{aligned} \xi_{K1,F0} &= \frac{1}{\varepsilon^q} \cdot K_\varepsilon(X, X'); \\ \xi_{K1,F2} &= \frac{1}{\varepsilon^{q+2d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')^2; \\ \xi_{K2,F1} &= \frac{1}{\varepsilon^{2q+d}} \cdot K_\varepsilon(X, X')^2 \cdot F(X, X'); \\ \xi_{K2,F2} &= \frac{1}{\varepsilon^{2q+2d}} \cdot K_\varepsilon(X, X')^2 \cdot F(X, X')^2 \end{aligned}$$

выполнено

$$\begin{aligned}
& |\mathbb{E}\xi_{K1,F0} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta| \leq \varepsilon^2 \cdot \tilde{C}_{E,2}; \\
& |\mathbb{E}\xi_{K1,F2} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta| \leq \varepsilon^2 \cdot \tilde{C}_{E,2}; \\
& |\mathbb{E}\xi_{K2,F1} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{V,d}(X, \theta) \varphi(X, \theta, 0) d\theta| \leq \\
& \quad \leq \varepsilon \cdot I(d \neq 2) \cdot \tilde{C}_{E,1} + \varepsilon^2 \cdot \tilde{C}_{E,2}; \\
& |\mathbb{E}\xi_{K2,F2} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{V,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta| \leq \varepsilon^2 \cdot \tilde{C}_{E,2},
\end{aligned}$$

где

$$\begin{aligned}
\tilde{C}_{E,1} &= V_q \cdot (1 + C_\rho)^2 \cdot (p_{max} \cdot (1 + C_{\varphi,1})^2 + C_{p,1} \cdot (1 + C_\varphi)^2); \\
\tilde{C}_{E,2} &= V_q \cdot (1 + C_K + C_{K,1})^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2 \cdot \\
& \quad \cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + \\
& \quad + 8 \cdot V_q \cdot p_{max} \cdot (1 + C_K)^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2.
\end{aligned}$$

Доказательство Утверждения 4. Применяя Лемму 10 для $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$ и $\tilde{\xi} = \frac{1}{\varepsilon^q} \cdot K_\varepsilon(X, X') \cdot 1$, учитывая $\varepsilon^{-q} \cdot N^{-1} \rightarrow 0$ из P2, по закону больших чисел получаем:

$$\begin{aligned}
& \frac{1}{N\varepsilon^{q+d}} \cdot \sum_{n=1}^N K_\varepsilon(X, X_n) \cdot F(X, X_n) \xrightarrow{P} \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta; \\
& \frac{1}{N\varepsilon^q} \cdot \sum_{n=1}^N K_\varepsilon(X, X_n) \cdot F(X, X_n) \xrightarrow{P} \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta > 0.
\end{aligned}$$

Откуда следует сходимость частного к частному пределов. □

Доказательство Теоремы 3. Преобразуем

$$\begin{aligned}
& F_N(X) - \frac{\mathbb{E} \frac{1}{\varepsilon^{q+d}} K_\varepsilon(X, X_1) \cdot F(X, X_1)}{\mathbb{E} \frac{1}{\varepsilon^q} K_\varepsilon(X, X_1)} = \\
& = \left(\frac{1}{N} \sum_{n=1}^N \frac{1}{\varepsilon^{2q+d}} \left(K_\varepsilon(X, X_i) F(X, X_i) \mathbb{E} K_\varepsilon(X, X_n) - \right. \right. \\
& \quad \left. \left. - K_\varepsilon(X, X_n) (\mathbb{E} K_\varepsilon(X, X_n) F(X, X_n)) \right) \right) / \\
& \quad / \left(\mathbb{E} \frac{1}{\varepsilon^q} K_\varepsilon(X, X_1) \cdot \frac{1}{N \varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n) \right). \tag{40}
\end{aligned}$$

Знаменатель (40) по вероятности сходится к $(p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta)^2 > 0$. Числитель (40) представлен в виде суммы центрированных независимых (как функции от независимых) одинаково распределенных случайных величин

$$\begin{aligned}
\varsigma_n &= \left(\frac{1}{\varepsilon^{q+d}} K_\varepsilon(X, X_i) F(X, X_i) \right) \cdot \left(\frac{1}{\varepsilon^q} \mathbb{E} K_\varepsilon(X, X_n) \right) - \\
& - \left(\frac{1}{\varepsilon^{q+d}} \mathbb{E} K_\varepsilon(X, X_i) F(X, X_i) \right) \cdot \left(\frac{1}{\varepsilon^q} K_\varepsilon(X, X_n) \right), \quad n = 1, \dots, N
\end{aligned}$$

со вторым моментом из Следствия 5 при $N \rightarrow \infty$:

$$\mathbb{E} \varsigma_n^2 \rightarrow d(X),$$

где

$$\begin{aligned}
& \tilde{d}(X) = \frac{1}{\varepsilon^q} \cdot p_\mu(X)^3 \cdot \\
& \cdot \left(\int_{S^{q-1}} \rho_{V,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta \cdot \left(\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \right)^2 + \right. \\
& + \int_{S^{q-1}} \rho_{V,2d}(X, \theta) d\theta \cdot \left(\int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta \right)^2 - \\
& - 2 \cdot \int_{S^{q-1}} \rho_{V,d}(X, \theta) \varphi(X, \theta, 0) d\theta \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \cdot \\
& \quad \left. \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta \right).
\end{aligned}$$

Также,

$$\mathbb{E}\zeta_n^2 \leq 4 \cdot (1 + C_K)^2 \cdot (1 + C_\varphi + C_{\varphi,1} + C_{\varphi,2}),$$

где $C_\varphi, C_{\varphi,1}, C_{\varphi,2}, C_K$ — константы (30), (31), (32), (28) соответственно. Откуда следует, что

$$N \cdot \varepsilon^q \cdot \text{Var} \left(F_N(X) - \frac{\mathbb{E} \frac{1}{\varepsilon^{q+d}} K_\varepsilon(X, X_1) \cdot F(X, X_1)}{\mathbb{E} \frac{1}{\varepsilon^q} K_\varepsilon(X, X_1)} \right) \rightarrow d(X),$$

где

$$\begin{aligned} d(X) &= N \cdot \varepsilon^q \cdot \tilde{d}(X) / \left(p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \right)^2 = \\ &= p_\mu(X) \cdot \left(\int_{S^{q-1}} \rho_{V,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta \cdot \left(\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \right)^2 + \right. \\ &\quad + \int_{S^{q-1}} \rho_{V,2d}(X, \theta) d\theta \cdot \left(\int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta \right)^2 - \\ &\quad - 2 \cdot \int_{S^{q-1}} \rho_{V,d}(X, \theta) \varphi(X, \theta, 0) d\theta \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \cdot \\ &\quad \left. \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta \right) / \\ &\quad \left. / \left(\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \right)^2. \end{aligned} \quad (41)$$

При $N \rightarrow \infty$:

$$\begin{aligned} &\frac{\mathbb{E} \frac{1}{\varepsilon^{q+d}} K_\varepsilon(X, X_1) \cdot F(X, X_1)}{\mathbb{E} \frac{1}{\varepsilon^q} K_\varepsilon(X, X_1)} - \frac{\int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta}{\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} = \\ &= I(d \neq 2) \cdot \frac{E_1(X)}{p_\mu(X)} \cdot \frac{1}{\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} \cdot \varepsilon + O(\varepsilon^2). \end{aligned}$$

Обозначим

$$m(X) = I(d \neq 2) \cdot \frac{E_1(X)}{p_\mu(X)} \cdot \frac{1}{\int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta}. \quad (42)$$

Так как по P3: $N\varepsilon^4 \rightarrow 0$ при $N \rightarrow \infty$, то из центральной предельной теореме

получаем утверждение Теоремы:

$$\sqrt{N\varepsilon^q} \cdot (F_N(X) - \bar{F}(X)) \rightarrow^d N(m(X), d(X)).$$

Причем $m(X) \equiv 0$ при $d:2$. □

Доказательство теоремы 4. 1. Обозначим

$$N_0 = \frac{\max\{1, (C_\varphi + C_{\varphi,1} + C_{\varphi,2})\}}{p_{\max} \cdot V_q \cdot \min\{1, C_\varphi^2\} \cdot C_K^{2d-1}}, \quad (43)$$

$$C_\sigma = p_{\max} \cdot C_K^2 \cdot (1 + C_\varphi)^2 \cdot V_q. \quad (44)$$

2. Рассмотрим случайные величины $\chi_n = \xi_n - \mathbb{E}\xi_n$, $\xi_n = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X_n) \cdot F(X, X_n)$, $n = 1, \dots, N$, $\chi = \xi - \mathbb{E}\xi$, $\xi = \frac{1}{\varepsilon^{q+d}} \cdot K_\varepsilon(X, X') \cdot F(X, X')$:

$$\text{Var}\chi \leq 2\varepsilon^{-q} \cdot \sigma^2(X);$$

$$\text{Var}\chi \geq 1/2 \cdot \sigma^2(X),$$

из Леммы 10 при

$$\varepsilon \leq \varepsilon_0 \equiv \min\left\{\frac{\sigma^2(X)}{2 \cdot C_{V,2}}, 1\right\} \quad (45)$$

Также, при $\varepsilon < 1$ для $m \geq 2$

$$\begin{aligned} \mathbb{E}|\chi|^m &\leq \text{Var}\chi \cdot \sup_{\omega} |\chi(\omega)|^{m-2} \leq \\ &\leq \text{Var}\chi \cdot \varepsilon^{-q(m-2)} \cdot (C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}))^{m-2} \leq \\ &\leq \frac{m!}{2} \cdot \text{Var}\chi \cdot H^{m-2}, \end{aligned}$$

где $H = \varepsilon^{-q} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})$. Откуда, по Теореме 13 для $z \in \left[0, N \cdot \frac{1}{2} \cdot \frac{\sigma^2(X)}{C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})}\right]$:

$$P\left(\left|\frac{\sum_{n=1}^N \chi_n}{N}\right| > z\right) \leq 2 \cdot \exp\left(-\frac{z^2 \cdot N\varepsilon^q}{2\sigma^2(X)}\right).$$

3. Из (36) следует, что

$$\sigma^2(X) \leq p_{max} \cdot V_q \cdot C_\varphi^2 \cdot C_K^{2d}.$$

Поэтому при $N \geq \frac{(C_\varphi + C_{\varphi,1} + C_{\varphi,2})}{p_{max} \cdot V_q \cdot C_\varphi^2 \cdot C_K^{2d-1}}$ пункт 2 справедлив для $z \in [0, 1]$.

4. Из Леммы 10 следует, что

$$\left| \xi_n - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta \right| \leq \varepsilon \cdot C_{E,1} \cdot I(d/2) + \varepsilon^2 \cdot C_{E,2}.$$

5. Из пунктов 2-4 следует, что при $N \geq \frac{(C_\varphi + C_{\varphi,1} + C_{\varphi,2})}{p_{max} \cdot V_q \cdot C_\varphi^2 \cdot C_K^{2d-1}}$ и $\varepsilon < \min\{1, C_{int}\}$:

$$P \left(\left| \frac{\sum_{n=1}^N \chi_n}{N} \right| > z + \varepsilon \cdot C_{E,1} \cdot I(d/2) + \varepsilon^2 \cdot C_{E,2} \right) \leq 2 \cdot \exp \left(-\frac{z^2 \cdot N \varepsilon^q}{2 \cdot \sigma^2(X)} \right).$$

6. Повторяя пункты 2-5 для случайных величин $K_\varepsilon(X, X_n) \cdot 1$, вместо $K_\varepsilon(X, X_n) \cdot F(X, X_n)$, и используя Следствие 5 для $N \geq \frac{1}{p_{max} \cdot V_q \cdot C_K^{2d-1}}$ и $\varepsilon < \min\{1, C_{int}\}$, получаем

$$\begin{aligned} P \left(\left| \frac{\sum_{n=1}^N K_\varepsilon(X, X_n)}{N \varepsilon^q} - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta \right| > z + \varepsilon^2 \cdot \tilde{C}_{E,2} \right) &\leq \\ &\leq 2 \cdot \exp \left(-\frac{z^2 \cdot N \varepsilon^q}{2 \cdot \sigma^2(X)} \right). \end{aligned}$$

7. При $\varepsilon \leq \sqrt{\frac{1}{2} \cdot \frac{p_{min} \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta}{C_{E,2}}}$

$$|\mathbb{E} \frac{1}{\varepsilon^q} K_\varepsilon(X, X') - p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta| \leq \frac{1}{2} \cdot p_{min} \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta.$$

Также

$$\begin{aligned}
& |\bar{F}(X)| \leq C_\varphi; \\
& \left| \bar{F}(X) - \frac{\mathbb{E}_{\varepsilon^{d+q}} \frac{1}{\varepsilon^{d+q}} K_\varepsilon(X, X') F(X, X')}{\mathbb{E}_{\varepsilon^q} \frac{1}{\varepsilon^q} K_\varepsilon(X, X')} \right| \leq \\
& \leq \bar{F}(X) \cdot \frac{2\varepsilon^2 \cdot \tilde{C}_{E,2}}{p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} + \frac{2\varepsilon \cdot C_{E,1} \cdot I(d \neq 2) + \varepsilon^2 \cdot \tilde{C}_{E,2}}{p_\mu(X) \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} \leq \\
& \leq \frac{2}{p_{\min} \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} \cdot \left(\varepsilon \cdot C_{E,1} \cdot I(d \neq 2) + \varepsilon^2 \cdot \tilde{C}_{E,2} \cdot (1 + C_\varphi) \right).
\end{aligned}$$

8. Из 5-7 получаем

$$\begin{aligned}
P \left(|F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2} \right) &\leq \\
&\leq 4 \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{\sigma^2(X)} \right),
\end{aligned}$$

где

$$C_{LD1} = \frac{2}{p_{\min} \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} \cdot C_{E,1}, \quad (46)$$

$$C_{LD2} = \frac{2}{p_{\min} \cdot \int_{S^{q-1}} \rho_{E,d}(X, \theta) d\theta} \cdot \tilde{C}_{E,2} \cdot (1 + C_\varphi). \quad (47)$$

□

Доказательство Теоремы 5. 1. Из Леммы 11 и свойства конечной сети для подмножества следует, что для $\delta = \varepsilon^3$ найдется конечная δ сеть \mathbb{N}_δ для отдаленных от границы точек \mathbb{M}_ε , состоящая не более чем из $\left(\frac{2a\sqrt{p}}{\varepsilon^3}\right)^p$ элементов (мелкость сети ε^3 выбрана так, чтобы порядок уклонения итоговой оценки совпадал с локальным из Теоремы 4).

2. Для точки $X' = \exp_X(t'\theta')$, где $t' \leq \varepsilon^3$ и $\theta' \in S_{q-1}$ из формулы Тейлора

следует для $\varepsilon \leq 1$:

$$\begin{aligned}
|\bar{F}(X') - \bar{F}(X)| &\leq \varepsilon^3 \cdot \left(\sup_{X,t',\theta} \left| \frac{\partial \varphi(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| + \right. \\
&\quad \left. + \frac{1}{\varepsilon} \cdot V_q \cdot C_\varphi \cdot \sup_{X,t',\theta} \left| \frac{\partial K(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| + \right. \\
&\quad \left. + \frac{1}{\varepsilon} \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot \sup_{X,t',\theta} \left| \frac{\partial K(\exp_X(t'\theta'), \theta, 0)}{\partial t'} \right| \right) / \int_{S_{q-1}} \rho_{E,0}(X, \theta) d\theta \leq \\
&\leq \varepsilon^2 \cdot \left(C_{u,\varphi} + V_q C_\varphi \cdot \tilde{C}_K + V_q \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot \tilde{C}_{K,1} \right) / c_\rho.
\end{aligned}$$

3. Оценим сверху отличие $\frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X', X_n) F(X', X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X', X_n)}$ и $\frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X, X_n) F(X, X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n)}$. Из Теоремы 2 для $z = 1/8$ и К4 с вероятностью не менее $1 - \left(\frac{6a\sqrt{p}}{\varepsilon}\right)^p \cdot \exp(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2 / (9p_{\max}))$

$$\inf_{X' \in \mathbb{M}_\varepsilon} \frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n) \geq C_{K, \frac{1}{2}} \cdot \left(V_q \cdot p_{\min} \cdot 7/8 - \varepsilon^2 \cdot \frac{C_E \cdot p_{\max}}{V_q p_{\min}} \right).$$

При $\varepsilon < \sqrt{\frac{3}{4} \cdot V_q \cdot p_{\min} \cdot \frac{1}{\sqrt{C_E \cdot p_{\max}}}}$:

$$\inf_{X' \in \mathbb{M}_\varepsilon} \frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n) \geq C_{K, \frac{1}{2}} \cdot V_q \cdot p_{\min} \cdot 1/2.$$

Откуда

$$\begin{aligned}
\left| \frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X', X_n) F(X', X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X', X_n)} - \frac{\frac{1}{N\varepsilon^{q+d}} \sum_{n=1}^N K_\varepsilon(X, X_n) F(X, X_n)}{\frac{1}{N\varepsilon^q} \sum_{n=1}^N K_\varepsilon(X, X_n)} \right| &\leq \\
&\leq 2 \cdot \varepsilon^2 \cdot \frac{\tilde{C}_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) + 2C_K \cdot (\tilde{C}_\varphi + \tilde{C}_{\varphi,1} + \tilde{C}_{\varphi,2})}{C_{K, \frac{1}{2}} \cdot V_q \cdot p_{\min}}.
\end{aligned}$$

Обозначим

$$C_{LD,net} = 2 \cdot \frac{\tilde{C}_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) + 2C_K \cdot (\tilde{C}_\varphi + \tilde{C}_{\varphi,1} + \tilde{C}_{\varphi,2})}{C_{K, \frac{1}{2}} \cdot V_q \cdot p_{\min}}. \quad (48)$$

4. Из Теоремы 4 по Лемме 7 для всех элементов конечной сети \mathbb{N}_δ следует

для $z \in [0, 1]$:

$$\begin{aligned} P \left(|F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot C_{LD2} \right) &\leq \\ &\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3} \right)^p \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma} \right). \end{aligned}$$

5. Из пунктов 3,4, так как для каждой точки $X \in \mathbb{M}_\varepsilon$ найдется $X' \in \mathbb{N}_\delta$ следует:

$$\begin{aligned} P \left(\sup_{X \in \mathbb{M}_\varepsilon} |F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot (C_{LD2} + C_{LD,net}) \right) &\leq \\ &\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3} \right)^p \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma} \right) + \\ &+ \left(\frac{6a\sqrt{p}}{\varepsilon} \right)^p \cdot \exp \left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2 / (9p_{\max}) \right). \end{aligned}$$

Переобозначив $\tilde{C}_{LD2} = C_{LD2} + C_{LD,net}$, получаем утверждение теоремы. \square

Доказательство Утверждения 5. Обозначим число выборки, удаленных от границы на ε через N_ε . Тогда из Теоремы 5, Леммы 15 и ограниченности $F_N(X)$:

$$F_N = \frac{N_\varepsilon}{N} \sum_{X' \in \mathbb{M} \setminus \mathbb{M}_\varepsilon} F_N(X') + \frac{N - N_\varepsilon}{N} \sum_{X' \in \mathbb{M}_\varepsilon} F_N(X') \rightarrow^p 0 + \bar{F}.$$

\square

Доказательство Теоремы 6. Пользуясь разложением F_N из Утверждения 5, Теоремой 5, Леммами 15 и 16 получаем, что найдутся положительные $N_{U,0}$, $C_{U,1}$, $C_{U,2}$, $C_{U,p}$, что для любого $z \in [0, 1]$ и $N > N_{U,0}$:

$$P \left(|F_N - \bar{F}| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{U,1} + \varepsilon^2 \cdot C_{U,2} \right) \leq \exp \left(-z^2 \cdot N\varepsilon^q \cdot C_{U,p} \right).$$

\square

3 Выборочное оценивание непрерывных интегральных операторов

3.1 Постановка задачи

Пусть $\mathbf{F} = \{\phi\}$ — гильбертово (под)пространство функций определенных на многообразии \mathbb{M} со скалярным произведением $(\phi_1, \phi_2) = \int_{\mathbb{M}} \phi_1(X) \cdot \phi_2(X) \mu(dX)$, μ — вероятностная мера на \mathbb{M} , удовлетворяющие сформулированным в диссертации условиям регулярности. Пусть L — действующий на \mathbf{F} линейный самосопряженный оператор Гильберт-Шмидта с неположительными собственными значениями.

Рассмотрим задачу максимизации функционала

$$\int_{\mathbb{M}} \phi(X) \cdot L\phi(X) d\mu(X), \quad (49)$$

по нормированным функциям $\phi \in \mathbf{F}$ ($\|\phi\|_{\mu}^2 = \int_{\mathbb{M}} |\phi(X)|^2 d\mu(X) = 1$) ортогональным собственному подпространству \mathbf{F}_0 оператора L , например ядру $\text{Ker}(L) = \{\phi \in \mathbf{F} : L\phi = 0\}$ оператора L . Очевидно, что решение этой задачи даётся собственной функцией $\phi^*(X)$ оператора L с наибольшим ненулевым собственным значением λ^* .

Такие задачи естественным образом возникают в моделировании многообразий. Например, задача минимизации функционала $\int_{\mathbb{M}} |\nabla\phi(X)|^2 d\mu(X)$ по нормированным функциям ϕ , естественным образом возникающая при построении низкоразмерной параметризации многообразия данных (снижении размерности)^{19,21}, сводится к решению рассматриваемой оптимизационной задачи с оператором Лапласа-Бельтрами $L = \Delta_{LB}$ (у усеченным спектром), являющимся обобщением стандартного оператора Лапласа на случай многообразия.

В статистической постановке, многообразии \mathbb{M} и мера μ неизвестны, и исследуется задача оценивания ϕ^* по конечной выборке \mathbb{X}_N , состоящей из точек многообразия \mathbb{M} случайно и независимо друг от друга выбранных в \mathbb{M} в соответствии с вероятностной мерой μ .

Оптимизируемый функционал (49) квадратичен по ϕ , поэтому его выборочный аналог естественно также строить в виде квадратичной формы от вектора $\vec{\phi}_N = (\phi(X_1), \dots, \phi(X_N))^T$, состоящего из неизвестных значений функции

ϕ в точках выборки \mathbb{X}_N : интеграл (49) заменяется квадратичной оценкой

$$\int_{\mathbb{M}} \phi(X) \cdot L\phi(X) d\mu(X) \approx \frac{1}{N} \sum_{n=1}^N \phi(X_n) \cdot L\phi(X_n) \approx \vec{\phi}_N^T \cdot W_N \cdot \vec{\phi}_N,$$

в последнем члене вектор, состоящий из величин $\{(L\phi)(X_n), n = 1, 2, \dots, N\}$ заменен оценкой вида $W_N \cdot \vec{\phi}_N$, в котором W_N — некоторая явно построенная по оператору L матрица размера $N \times N$. В диссертации рассматривается конкретный способ^{13,19} построения такой матрицы, обеспечивающий состоятельность этой оценки.

Рассмотрим задачу максимизации квадратичной формы

$$\vec{\phi}_N^T \cdot W_N \cdot \vec{\phi}_N \quad (50)$$

по вектору $\vec{\phi}_N$ при условиях нормировки $\frac{1}{N} |\vec{\phi}_N|^2 = 1$ и ортогональности ядру матрицы $Ker(W_N) = \{\vec{\phi}_N \in \mathbb{F} : W_N \vec{\phi}_N = \mathbf{0}_N\}$, где $\mathbf{0}_N$ — вектор из N нулей. Решение этой задачи даётся собственным вектором $\vec{\phi}_N^* = (\phi_N^*(X_1), \phi_N^*(X_2), \dots, \phi_N^*(X_N))^T$ матрицы W_N с наибольшим ненулевым собственным значением ν_N^* .

По построенным оценкам $(\phi_N^*(X_1), \phi_N^*(X_2), \dots, \phi_N^*(X_N))$ значений неизвестной функции $\phi^*(X)$ в точках выборки, с использованием стандартных методов ядерного непараметрического оценивания²⁹, можно построить оценку $\hat{\phi}_N^*(X)$, функции $\phi^*(X)$ в произвольной точке X . В диссертации рассматривается конкретная оценка

$$\hat{\phi}_N^*(X) = \frac{\sum_{n=1}^N K_\varepsilon(X, X_n) \cdot \phi_N^*(X_n)}{\sum_{n=1}^N K_\varepsilon(X, X_n)}, \quad (51)$$

в которой ядро $K_\varepsilon(X, X')$ описано в Главе 2.

3.2 Иллюстративный пример

Пусть многообразие $\mathbb{M} = [-\pi, \pi] \subset \mathbb{R}^1$. На нем существует естественная параметризация $\mathbb{M} = \{X | X \in [-\pi, \pi]\}$, Однако, любая строго монотонная функция $g: [-\pi, \pi] \rightarrow \mathbb{R}$ также задает параметризацию на \mathbb{M} : $\mathbb{M} = \{X' | X' \in g^{-1}([-\pi, \pi])\}$. Следовательно, параметризаций бесконечно много. Поэтому, что-

бы задача нахождения параметризации имела единственное решение, можно сформулировать оптимизационную задачу. Например,

$$\int_{\mathbb{M}} \phi(X) L \phi(X) dX \rightarrow \max_{\phi \in \mathbf{F}}, \quad (52)$$

при условиях нормировки $\int_{\mathbb{M}} \phi(X)^2 dX = 1$, где \mathbf{F} — пространство бесконечно дифференцируемых функций на \mathbb{M} , $L = \Delta = \frac{\partial^2}{\partial X^2}$ — оператор Лапласа. В пространстве \mathbf{F} оператор Δ имеет базис (ненормированных) собственных функций: $1, X, \sin X, \cos X, \dots, \sin kX, \cos kX, \dots$ соответствующих собственным числам $0, 0, -1, -1, \dots, -k^2, -k^2, \dots$, где $k \in \mathbb{N}$. Если дополнительно потребовать, чтобы ϕ была ортогональна константной функции $\int_{\mathbb{M}} \phi(X) \cdot 1 dX = 0$, то решение (52) будет задаваться единственной допустимой собственной функцией, отвечающей минимальному собственному числу 0 , то есть $\phi^*(X) = \frac{1}{\pi^2} \cdot X$, которая задает выделенную параметризацию на \mathbb{M} (обладающую минимальным квадратом первой производной из теоремы Стокса).

Пусть теперь многообразие \mathbb{M} не известно (но по-прежнему является отрезком $[-\pi, \pi]$ прямой), а задана лишь конечная выборка $\mathbb{X}_N = \{X_1, \dots, X_N\}$ на нём (например, i.i.d. из равномерного распределения). При этом по ней требуется оценить параметризацию $\phi^*(X)$. Оценивание можно разбить на два этапа: оценивание $\phi^*(X)$ для $X \in \mathbb{X}_N$ и оценивание $\phi^*(X)$ для $X \in \mathbb{M}$ по оценкам $\{\phi^*(X_1), \dots, \phi^*(X_N)\}$ с предыдущего шага оценивания. Второй шаг может быть реализован, например, непараметрическим оцениванием (51)

$$\hat{\phi}_N(X) = \frac{\sum_{n=1}^N K_\varepsilon(X, X_n) \cdot \phi_N^*(X_n)}{\sum_{n=1}^N K_\varepsilon(X, X_n)},$$

результаты о качестве которого получены во второй Главе. А для оценивания $\{\phi^*(X_1), \dots, \phi^*(X_N)\}$ на первом шаге возможно сформулировать выборочную оптимизационную задачу. Оптимизируемый функционал (52) квадратичен по ϕ , поэтому его выборочный аналог естественно также строить в виде квадратичной формы от вектора $\vec{\phi}_N = (\phi(X_1), \dots, \phi(X_N))^T$, состоящего из неизвестных значений функции ϕ в точках выборки \mathbb{X}_N : интеграл (52) заменяется квадра-

тичной оценкой

$$\int_{\mathbb{M}} \phi(X) \cdot L\phi(X) d\mu(X) \approx \frac{1}{N} \sum_{n=1}^N \phi(X_n) \cdot L\phi(X_n) \approx \vec{\phi}_N^T \cdot W_N \cdot \vec{\phi}_N,$$

в последнем члене вектор, состоящий из величин $\{(L\phi)(X_n), n = 1, 2, \dots, N\}$ заменен оценкой вида $W_N \cdot \vec{\phi}_N$, в котором W_N — некоторая явно построенная по оператору L матрица размера $N \times N$.

В случае $L = \Delta$ в качестве квадратичной формы можно использовать

$$-\frac{C}{2} \cdot \frac{1}{N} \sum_{n=1}^N \frac{\sum_{j=1}^N K_\varepsilon(X_j, X_n) (\phi(X_j) - \phi(X_n))^2}{\varepsilon^2 \sum_{j=1}^N K_\varepsilon(X_j, X_n)} = -\frac{1}{\varepsilon^2} \cdot \vec{\phi}_N^T \cdot W_N \cdot \vec{\phi}_N,$$

где $W_N = (w_{ij,N})_{i,j=1}^N$, $w_{ii,N} = \left(1 - \frac{K_\varepsilon(X_i, X_i)}{\sum_{j=1}^N K_\varepsilon(X_j, X_i)}\right)$,

$i \neq j$: $w_{ij,N} = \frac{K_\varepsilon(X_i, X_j)}{\sum_{j=1}^N K_\varepsilon(X_j, X_j)}$, $C > 0$ — константа.

Оператор $L = \Delta$ не ограничен (для любого натурального k есть собственное число k^2), поэтому равномерная сходимость спектра W_N к спектру L невозможно. Однако, если рассмотреть лишь ограниченные произвольной константой C собственные значения операторов L и W_N , то, как доказывается в данной главе, собственные числа матрицы W_N будут равномерно сходиться к соответствующим собственным числам L , и собственные векторы W_N будут сходиться к значениям собственной функции ϕ^* в точках выборки.

3.3 Рассматриваемые статистики и основные результаты главы

Предполагается, что

- L1. L — неотрицательный самосопряженный линейный Гильберт-Шмитовский оператор, действующий в пространстве Банаховом пространстве функций $\mathbf{F} = C_b^s(\mathbb{M})$ функций, ограниченных вместе со своей s -ой производной;
- L2. Исключены эффекты границы:
 - или рассматривается компактное подмногообразие, вложенное в него $\mathbb{M}_K \subset \mathbb{M}$. В этой главе оно обозначается \mathbb{M} .

- или, вместо М1, М3 предполагается, что многообразие не имеет границы и покрыто конечным количеством карт.

Гильберт-Шмитовость оператора означает, что для некоторого (следовательно и любого) ортонормированного базиса $\{e_i | i \in I\}$ выполнено $\|L\|_{HS} = \sum_{i \in I} \|Le_i\|^2 < \infty$.

Замечание 6. Вводимые в главе конструкции близки к работе [13], однако, в данной главе рассматривается сходящийся к медленно сходящийся к нулю размер окрестности ε , вместо константного, который был в оригинальной работе, что позволило доказать сходимость выборочного оператора к исходному непрерывному (ранее было доказана сходимость к сглаженному аналогу). Доказательство для случая $\varepsilon \rightarrow 0$ при $N \rightarrow \infty$ стало возможным после получения равномерных результатов главы 2.

Введем обозначения для нормированных ядер $\tilde{K}_\varepsilon(X, X') = \frac{K_\varepsilon(X, X')}{\int_{\mathbb{M}} K_\varepsilon(X, X') \mu(dX')}$ и $\tilde{K}_N(X, X') = \frac{K_\varepsilon(X, X_n)}{\sum_{n=1}^N K_\varepsilon(X, X_n)}$.

Рассматриваются операторы на \mathbf{F}

- L — исходный оператор, действующий на \mathbf{F} ;
- $L_K: (L_K\phi)(X) = L \circ \int_{\mathbb{M}} \tilde{K}_\varepsilon(X, X') \phi(X') \mu(dX')$ — оператор L , сглаженный с весовой функцией \tilde{K}_ε ;
- $L_N: (L_N\phi)(X) = \frac{1}{N} \sum_{n=1}^N \tilde{K}_N(X, X_n) (L\phi)(X_n)$ — эмпирический аналог L_K ;
- $\hat{L}_N: (\hat{L}_N\phi)(X) = \frac{1}{N} \sum_{n=1}^N \tilde{K}_N(X, X_n) (\hat{L}\phi)(X_n)$ — приближение L_N , в котором действие оператора L заменено на его локальную линейную по ϕ оценку.

Связь операторов L и \hat{L}_N описывается Теоремой 5 главы 2.

Рассматриваются вспомогательные операторы

- $R_N: \mathbf{F} \rightarrow \mathbb{R}^N$, где $R_N\phi = \left((\hat{L}\phi)(X_1), \dots, (\hat{L}\phi)(X_N) \right)$ — оператор вычисления значений $L\phi$ в точках выборки;
- $E_N: \mathbb{R}^N \rightarrow \mathbf{F}$, где $E_N(y_1, \dots, y_N)(X) = \frac{1}{N} \sum_{n=1}^N y_n \cdot \tilde{K}_N(X, X_n)$ — оператор оценивания функции по значениям в точках выборки.

Также, $\hat{L}_N = E_N \circ R_N$ и $W_N = R_N \circ E_N$. Поэтому операторы \hat{L}_N и W_N связаны аналогом сингулярного разложения для Гильбертовых пространств — их

спектры совпадают с точностью до кратности нулевого собственного значения, собственные векторы являются значениями собственных функций в точках выборки, а собственные функции — образами собственных векторов при действии на них оператором оценивания E_N .

Теорема 7. Для собственных чисел $\lambda_m, m > 0$ оператора L , упорядоченных в порядке неубывания, и собственных чисел $\nu_{1,N} \leq \dots \nu_{N,N}$ матрицы W_N , имеет место сходимость по вероятности при $N \rightarrow \infty$:

$$\sup_{m \leq N} |\lambda_m - \nu_{m,N}| \xrightarrow{P} 0.$$

Теорема 8. Для произвольного ненулевого собственной функции $\phi^*(X)$ оператора L с собственным числом $\lambda_m, m > 0$ и функций $\hat{\phi}_N(X)$ (51), построенных по собственным векторам матрицы W_N , соответствующим собственным числам $\nu_{m,N}$, при $N \rightarrow \infty$ равномерно по m имеет место сходимость по вероятности:

$$\sup_{X \in \mathbb{M}} |\hat{\phi}_N(X) - \phi^*(X)| \xrightarrow{P} 0.$$

Теорема 9. Для собственных чисел $\lambda_m, m > 0$ оператора L , упорядоченных в порядке неубывания, и собственных чисел $\nu_{1,N} \geq \dots \nu_{N,N}$ матрицы W_N , существуют числа N_0 и $\delta_0 > 0$ такие, что для всех $N > N_0$ и $\alpha \in [0, \delta_0]$ имеет место неравенство:

$$P \left(\sup_m |\lambda_m - \nu_m| > \alpha \cdot C_6 + C_7 \cdot \varepsilon^2 \right) \leq Q(N) \cdot \exp(-C_8 \cdot \alpha^2 \cdot N\varepsilon^q),$$

где C_6, C_7, C_8 — положительные константы, а $Q(N)$ — полином от N .

Теорема 10. Для всех ненулевых собственных чисел λ_m оператора L и соответствующих им собственных функций $\phi_m^*(X)$ и соответствующих последовательностей $\nu_{m,N}$ собственных чисел матрицы W_N и функций $\hat{\phi}_{m,N}(X)$ (51), для всех отделенных на ε от границы многообразия X найдутся числа N_0 и $\delta_0 > 0$ такие, что для всех $N > N_0$:

$$\begin{aligned} P \left(\sup_m \sup_{X \in \mathbb{M}_\varepsilon} |\hat{\phi}_{m,N}(X) - \phi_m^*(X)| > C_9 \alpha + C_{10} \cdot \varepsilon^2 \right) &\leq \\ &\leq Q(N) \cdot \exp(-C_{11} \cdot \alpha^2 \cdot N\varepsilon^q), \end{aligned}$$

где C_9, C_{10}, C_{11} — положительные константы, а $Q(N)$ — полином от N .

3.4 Доказательство основных теорем главы

Доказательство Теорем 7, 8, 9 и 10. В данной теореме C — положительные константы, в общем случае разные в разных выражениях и вхождениях в одно выражение, однако равномерные по точкам многообразия.

По Теореме 5 для отделенных точек многообразия выполнено

$$\begin{aligned} P \left(\sup_{X \in \mathbb{M}_\varepsilon} |F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot I(d \neq 2) \cdot C_{LD1} + \varepsilon^2 \cdot \tilde{C}_{LD2} \right) &\leq \\ &\leq 4 \cdot \left(\frac{2a\sqrt{p}}{\varepsilon^3} \right)^p \cdot \exp \left(-\frac{z^2 \cdot N\varepsilon^q}{C_\sigma} \right) + \\ &+ \left(\frac{6a\sqrt{p}}{\varepsilon} \right)^p \cdot \exp \left(-1/16 \cdot N\varepsilon^q V_q p_{\min}^2 / (9p_{\max}) \right). \end{aligned}$$

То есть, из Гильберт-Шмитовости оператора L , $\|L - \hat{L}_N\|_{HS} \leq z + \varepsilon^2 \cdot \varepsilon^2$ с вероятностью не меньше $Q(N) \cdot \exp(-C \cdot z^2 \cdot N\varepsilon^q)$, где $Q(N)$ — полином от N .

Откуда, выбрав малое $z \sim \frac{1}{\sqrt{N\varepsilon^{q+\delta}}}$, для $4 > \delta > 0$ получаем из Утверждения 21 из [13] получаем сходимость спектров и собственных функций операторов \hat{L}_N к спектрам и собственным функциям оператора L . Откуда, по Утверждению 22 из [13] получаем сходимость собственных чисел W_N к собственным числам L и собственных векторов W_N к значениям собственных векторов L , взятым в точках выборки. В завершение, применяя Теорему 5 непараметрическому оцениванию 51 получаем утверждения Теорем 7, 8, 9 и 10. \square

4 Свойства процедур моделирования многообразий

В данной главе рассмотрены примеры конкретных статистических процедур моделирования многообразий [10, 46], в которых были использованы полученные в диссертации результаты.

4.1 Оценивание многообразий как задача вложения многообразий

Рассмотрим неизвестное q -мерное многообразие данных

$$\mathbb{M} = \{X = f(b) \in \mathbb{R}^p : b \in \mathbb{B} \subset \mathbb{R}^q\}, \quad (53)$$

покрытое одной картой (\mathbb{B}, f) и вложенное в p -мерное пространство \mathbb{R}^p , $q < p$. Отображение f является взаимнооднозначным из открытого ограниченного множества $\mathbb{B} \subset \mathbb{R}^q$ на многообразии $\mathbb{M} = f(\mathbb{B})$ с гладким обратным отображением $f^{-1}: \mathbb{M} \rightarrow \mathbb{B}$. Внутренняя размерность многообразия q также является неизвестной.

Предполагается, что для многообразия данных \mathbb{M} существует положительное число обусловленности $c(\mathbb{M})$, то есть для каждой точки $X \in \mathbb{R}^p$, удаленной от \mathbb{M} не более чем на $1/c(\mathbb{M})$, существует единственная проекция на \mathbb{M} [40].

Существует обратная функция $h_f(X) = f^{-1}(X)$, чьи значения $b = h_f(X) \in \mathbb{R}^q$, можно рассматривать как низкоразмерные координаты на многообразии \mathbb{M} , задающие низкоразмерные представления (признаки) $b = h_f(X)$ многомерных данных X с многообразия.

Если отображения $h_f(X)$ и $f(b)$ гладкие, и матрица Якоби $J_f(b)$ размера $p \times q$ отображения $f(b)$ ранга q , тогда q -мерное линейное пространство

$$L(X) = \text{Span}(J_f(h_f(X))) \quad (54)$$

пространства \mathbb{R}^p является касательным пространством многообразия \mathbb{X} в точке $X \in \mathbb{M}$. Здесь и далее $\text{Span}(H)$ — это линейное пространство, построенное на столбцах матрицы H .

Пусть $\mathbb{X}_N = \{X_1, \dots, X_N\} \subset \mathbb{M}$ — случайная выборка на многообразии, сгенерированная из какой-то (неизвестной) вероятностной меры, чей носитель совпадает с \mathbb{M} . Общая задача обучения на многообразиях ставится следующим образом [8]: по заданной выборке \mathbb{X}_N построить низкоразмерную параметризацию многообразия данных, которая задает отображение

$$H: \mathbb{M} \subset \mathbb{R}^p \rightarrow \mathbb{Y}_h = h(\mathbb{M}) \subset \mathbb{R}^q \quad (55)$$

из многообразия данных \mathbb{M} в пространство признаков $\mathbb{Y}_h \subset \mathbb{R}^q$, $q < p$, которое сохраняет свойства многообразия данных.

4.2 Типичная схема алгоритмов вложения многообразий

Следуя [47], рассмотрим класс типичных алгоритмов обучения на многообразиях, который восстанавливает внутреннюю структуру многообразия данных по выборке; этот класс включает так называемые алгоритмы с “нормированными входами”. Общая схема рассматриваемых алгоритмов состоит из четырех шагов.

4.2.1 Шаг первый: построение окрестности

Для каждой точки выборки X_n ищутся окрестности из соседей $U_{X_n} = \{X_n, X_{n,1}, \dots, X_{n,k(n)}\} \subset \mathbb{X}_N$, то есть подмножество элементов выборки, состоящее из близких к X_n точек. Например, $U_{X_n} = U_{X_n,\varepsilon}$ состоит из точек \mathbb{X}_N , попавших в шар радиуса ε с центром в X_n в пространстве \mathbb{R}^p , или $U_{X_n,k}$ состоит из $k(N) = k$ ближайших к X_n элементов \mathbb{X}_N .

Построенные окрестности определяют граф выборки $\Gamma(\mathbb{X}_N)$, состоящий из вершин $\{X_1, \dots, X_N\}$, и между вершинами X_i и X_j есть ребро тогда и только тогда, когда $X_i \in U_N(X_j)$ или $X_j \in U_{X_i}$.

4.2.2 Шаг второй: описание окрестностей

Вычисляются описания выбранных локальных свойств многообразия. Примерами таких описаний служат:

- барицентрические координаты $\{w_{n,1}, \dots, w_{n,k}\}$ “центральной” точки X_i по системе ее ближайших соседей $\{X_{n,1}, \dots, X_{n,k}\}$, то есть решение оптимизационной задачи $\|X_n - \sum_j w_{n,j} X_{n,j}\|^2$ [17];
- результатом применения метода главных компонент к окрестности $U_{X_n,\varepsilon}$ является ортогональная матрица $Q_{PCA}(X_n)$ размера $p \times q$, чьи столбцы являются главными собственными векторами, соответствующими q наибольшим собственным значениям [18, 48]. Такие матрицы определяют q -мерные линейные подпространства $L_{PCA}(X_n) = \text{Span}(Q_{PCA}(X_n))$ в \mathbb{R}^p , которые, при определенных условиях хорошо аппроксимируют касательные пространства $T_{X_n}(\mathbb{M})$ к \mathbb{M} в точке X_n [14].

4.2.3 Шаг третий: глобальное описание

Вычисляется глобальное описание многообразия данных. Для этого решаются выпуклые оптимизационные задачи при некоторых нормировочных ограничениях. Обычно, низкоразмерные признаки $\mathbb{Y}_N = h(\mathbb{X}_N) = \{y_1, \dots, y_N\} \subset \mathbb{Y}_h$ находятся как решения задач минимизации выбранной функции цены $L(\mathbb{Y}_N | \mathbb{X}_N)$ по \mathbb{Y}_N . Примерами таких функций могут служить

$$\begin{aligned} L_{LLE}(\mathbb{Y}_N | \mathbb{X}_N) &= \sum_{n=1}^N \|y_n - \sum_{j=1}^N w_{n,j} y_{n,j}\|_F^2; \\ L_{LE}(\mathbb{Y}_N | \mathbb{X}_N) &= \sum_{n,j=1}^N K(X_n, X_j) \cdot \|y_n - y_j\|_2^2; \\ L_{LTSA}(\mathbb{Y}_N | \mathbb{X}_N) &= \\ &= \sum_{j=1}^N \|(I_q - Q_{PCA}(X_n) \cdot Q_{PCA}^T(X_n)) \cdot H_n \cdot Y_{(n)}\|_F^2; \end{aligned}$$

используемые в алгоритмах LLE [17], LE [19] и LTSA [49] соответственно; здесь $Y_{(n)}$ — матрицы размера $q \times (k(i) + 1)$, состоящие из q -мерных столбцов $\{y_n, y_{n,1}, \dots, y_{n,k(n)}\}$, где пары индексов используются как в U_{X_n} ; $H_n = I_q - (1/k(n)) \cdot \vec{1} \cdot \vec{1}^T$ — матрица размера $q \times q$, I_q — единичная матрица порядка q , $\vec{1}$ — вектор из единиц размера q . Некоторые ограничения в задачах используются, чтобы избежать вырожденных решений.

4.2.4 Шаг четвертый: расширение задачи для точек вне выборки

Выборка признаков \mathbb{Y}_n задает значения отображения вложения $h(X)$ (55) только в точках выборки. Обычно, интерес представляет нахождение признаков $h(X)$ для точек вне выборки $X \in \mathbb{M} \setminus \mathbb{X}_N$. Популярное расширение для алгоритмов LLE, LE основаны на общем подходе ядерного метода главных компонент [50] и предложены в [51, 52]. Концепция функций цены также позволяет строить вложения для $X \in \mathbb{M} \setminus \mathbb{X}_N$ [53].

4.3 Основной результат главы

Множество $\mathbb{M} \setminus \mathbb{X}_N$ неизвестно. Поэтому для отображения $h(X)$ (55) необходимо задать область определения. Так как алгоритмы могут оценивать раз-

личные свойства многообразий, то, чтобы гарантировать хорошие свойства оценок, область определения задается различным образом в зависимости от рассматриваемого алгоритма. В данной работе рассматривается область определения алгоритма спектральных вложений Грассмана-Штифеля [?, 46, 46, 48, 54, 55].

Для ее определения введем дополнительные обозначения. Для точки $X \in \mathbb{X}$, гладкой неотрицательной монотонно убывающей функции одной скалярной переменной $k(\rho)$, $\rho \in [0, \infty)$ и скалярного параметра $\varepsilon = \varepsilon(N)$ определим выборочную матрицу ковариационную матрицу:

$$\Sigma(X) = \frac{\sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon) \cdot (X_n - X) \cdot (X_n - X)^T}{\sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon)}, \quad (56)$$

где $U_{X, \varepsilon}$ — шар радиуса ε с центром в X в \mathbb{R}^p , $\|X - X_n\|$ — вторая норма в пространстве \mathbb{R}^p .

Область определения \mathbb{M}_h отображения сжатия $h(X)$ определяется как множество точек $X \in \mathbb{R}^p$ таких, что ровно q наибольших собственных чисел матрицы $\Sigma(X)$ больше или равны заданному скалярному параметру λ_{\min} .

Обозначим

$$C_K = \frac{\int_{Z \in \mathbb{R}^q, \|Z\| \leq 1} k(\|Z\|) \|Z\|^2 dZ}{q \cdot \int_{Z \in \mathbb{R}^q, \|Z\| \leq 1} k(\|Z\|) dZ}. \quad (57)$$

Обозначим расстояние Хаусдорфа между множествами $A, B \subset \mathbb{R}^p$ через $d_H(A, B)$:

$$d_H(A, B) = \max\left\{\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\|\right\}. \quad (58)$$

Основной результат работы состоит в

Теорема 11. *Если относительно многообразия \mathbb{X} , выборки, параметра $\varepsilon(N)$ предполагается M1-M8, S1-S3, P1-P3 соответственно, функция $k(\rho)$, $\rho \in [0, \infty)$ является неотрицательной монотонно убывающей функцией одной скалярной переменной и параметры $0 < \alpha < \beta < 1$, то для $\alpha \cdot C_K \leq \lambda_{\min} \leq \beta \cdot C_K$ выполнено:*

$$d_H(\mathbb{X}_h, \mathbb{X}) \xrightarrow{p} 0.$$

Для доказательства теоремы нам потребуются дополнительные обозначения.

Пусть для точки $X \in \mathbb{X}$ векторы V_1, \dots, V_q задают ортонормированный базис в касательном пространстве $L(X)$, и векторы W_1, \dots, W_{p-q} задают ортонормированный базис в ортогональном дополнении $L(X)$ до \mathbb{R}^p . Обозначим элементы матрицы ковариации в базисе $V_1, \dots, V_q, W_1, \dots, W_{p-q}$: для $E_1, E_2 \in \{V_1, \dots, V_q, W_1, \dots, W_{p-q}\}$:

$$\begin{aligned} \Sigma(X|E_1, E_2) &= \\ &= \sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon) \cdot \\ &\cdot ((X_n - X)^T \cdot E_1) \cdot ((X_n - X)^T \cdot E_2) / \\ &\sum_{n \in U_N(X, \varepsilon)} k(\|X - X_n\|/\varepsilon), \end{aligned} \quad (59)$$

Теорема 12 (Вейля). Пусть A и E — симметричные матрицы порядка p . Тогда для упорядоченных по возрастанию собственных $\lambda_1 \geq \dots \geq \lambda_p$ и $\mu_1 \geq \dots \geq \mu_p$ чисел матриц A и $A + E$ выполнено для $i = 1, \dots, p$

$$\lambda_i - \|E\|_F \leq \mu_i \leq \lambda_i + \|E\|_F,$$

где $\|B\|_F$ — норма Фробениуса матрицы B .

Доказательство. Доказана в книге [56]. □

Доказательство теоремы 11. Для матрицы $A = (a_{ij})_{i,j=1}^p$, где $a_{i,i} = C_K$ для $i \leq p$, а остальные элементы равны нулю, и матрицы $E = (e_{ij})_{i,j=1}^p$, где $|e_{ij}| \leq \varepsilon^2 \cdot C_\Sigma$, по Теореме 12 для собственных чисел матрицы $A + E$:

$$\mu_i \geq C_K - p^2 \cdot \varepsilon^2 \cdot C_\Sigma, \quad i \leq q;$$

$$\mu_i \leq p^2 \cdot \varepsilon^2 \cdot C_\Sigma, \quad i > p.$$

Следовательно, при больших N :

$$\mu_i \geq \lambda_{\min}, \quad i \leq q;$$

$$\mu_i \leq \lambda_{\min}, \quad i > p.$$

Применяя Теорему 5к каждому из $p \cdot (p - 1)/2$ уникальных элементов матрицы ковариации, получаем, что $\mathbb{X}_\varepsilon \subset \mathbb{X}_h$ с вероятностью не менее

$$C \cdot p^{p/2+2} \cdot \varepsilon^{-p} \cdot \exp(-C_e \cdot N\varepsilon^q V_q)$$

для некоторых положительных констант C, C_e . Далее, точки удаленные от \mathbb{X} больше, чем на ε не входят в \mathbb{X}_h . Поэтому

$$\sup_{X \in \mathbb{X}} \inf_{X' \in \mathbb{X}_h} \|X - X'\| \leq \varepsilon$$

с вероятностью не меньше

$$C \cdot p^{p/2+2} \cdot \varepsilon^{-p} \cdot \exp(-C_e \cdot N\varepsilon^q V_q) \rightarrow 0,$$

при $N \rightarrow \infty$.

□

Заключение

В диссертации получены следующие основные результаты:

1. Исследованы асимптотические свойства числа точек $N_\varepsilon(X)$ выборки, попавших в асимптотически малые ε -окрестности точек X многообразия (состоятельность, асимптотическое разложение и большие отклонения биномиального случайного поля $N_\varepsilon(X)$), а также доказана асимптотическая равномерность этих точек в рассматриваемой асимптотически малой окрестности. Доказана равномерность по многообразию результатов о больших отклонениях биномиального случайного поля $N_\varepsilon(X)$ от своего предельного значения;
2. Доказана состоятельность, найдено асимптотическое разложение и получены теоремы о больших отклонениях, для рассмотренного класса локальных статистик $F_N(X)$, доказана равномерность этих результатов по точкам многообразия;
3. Рассмотрен класс глобальных статистик F_N , являющихся усреднениями локальных статистик по точкам выборки. Доказана их состоятельность и

получены теоремы о больших уклонениях;

4. Для рассмотренного класса непрерывных функционалов от определенных на многообразии функций и их выборочных конечномерных матричных аналогов. Доказана сходимость собственных чисел и собственных функций матричных аналогов к собственным числам и собственным функциям предельных функционалов, получены равномерные теоремы о больших уклонениях.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Bellman R. E.* Dynamic programming. — Princeton University Press, 1957.
2. *Donoho D. L.* High-dimensional data analysis: The curses and blessings of dimensionality // *AMS conference on math challenges of 21st century.* — 2000. — Pp. 1–31.
3. *Hinton G., Salakhutdinov R.* Reducing the dimensionality of data with neural networks // *Science.* — 2006. — Vol. 313, no. 5786. — Pp. 504–507.
4. *Cox M. A. A., Cox T. F.* Multidimensional Scaling // *Handbook of Data Visualization.* — Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. — Pp. 315–347.
5. *Бухштабер В. М., Маслов В. К.* Факторный анализ и экстремальные задачи на многообразиях Грассмана // *Математические методы решения экономических задач.* — 1977. — Т. 7. — С. 85–102.
6. *Buchstaber V. M.* Time series analysis and Grassmannians // *Applied problems of Radon transform.* — 1994. — Vol. 162, no. 2. — Pp. 1–17.
7. *Seung H. S., Lee D. D.* COGNITION: The Manifold Ways of Perception // *Science.* — 2000. — dec. — Vol. 290, no. 5500. — Pp. 2268–2269.
8. *Ma Y., Fu Y.* Manifold Learning Theory and Applications. — London: CRC Press, 2011.
9. *Huo X., Smith A.* A Survey of Manifold-Based Learning Methods // *Journal of Machine Learning Research.* — 2008. — Pp. 1–34.
10. *Kuleshov A., Bernstein A.* Nonlinear multi-output regression on unknown input manifold // *Annals of Mathematics and Artificial Intelligence.* — 2017. — may. — Pp. 1–32.
11. *Henry G., Mucoz A., Rodriguez D.* Locally adaptive density estimation on Riemannian manifolds // *Statistics and Operations Research Transactions.* — 2013. — Vol. 37, no. 2. — Pp. 111–130.
12. *Smith A., Zha H., Wu X.-m.* Convergence and Rate of Convergence of a Manifold-Based Dimension Reduction Algorithm // *Advances in Neural Information Processing Systems 21.* — 2009. — Pp. 1529–1536.

13. *Rosasco L., Belkin M., Vito E. D.* On learning with integral operators // *The Journal of Machine Learning Research*. — 2010. — Vol. 11. — Pp. 905–934.
14. *Singer A., Wu H.-T.* Vector diffusion maps and the connection Laplacian // *Communications on Pure and Applied Mathematics*. — 2012. — aug. — Vol. 65, no. 8. — Pp. 1067–1144.
15. *Perrault-Joncas D., Meila M.* Non-linear Dimensionality Reduction: Riemannian Metric Estimation and the Problem of Geometric Recovery // *arXiv*. — 2013. — Vol. 1305.7255v. — Pp. 1–32.
16. *Tenenbaum J. B., de Silva V., Langford J.* A Global Geometric Framework for Nonlinear Dimensionality Reduction // *Science*. — 2000. — dec. — Vol. 290, no. 5500. — Pp. 2319–2323.
17. *Roweis S. T., Saul L. K.* Nonlinear dimensionality reduction by locally linear embedding // *Science*. — 2000. — Vol. 290. — Pp. 2323–2326. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.111.3313>.
18. *Zhang Z., Zha H.* Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment // *SIAM Journal on Scientific Computing*. — 2004. — Vol. 26, no. 1. — Pp. 313–338.
19. *Belkin M., Niyogi P.* Laplacian Eigenmaps for dimensionality reduction and data representation // *Journal Neural Computation*. — 2003. — Vol. 15, no. 6. — Pp. 1373–1396.
20. *Donoho D. L., Grimes C.* Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data // *Proceedings of the National Academy of Sciences*. — 2003. — may. — Vol. 100, no. 10. — Pp. 5591–5596.
21. *Bernstein A., Kuleshov A. P.* Manifold Learning: generalizing ability and tangent proximity // *International Journal of Software and Informatics*. — 2013. — Vol. 7, no. 3. — Pp. 359–390.
22. Spectral methods for dimensionality reduction / L. K. Saul, K. Q. Weinberger, J. H. Ham et al. // *Semisupervised Learning*. — 2006. — Pp. 1–18.
23. *Gine E., Koltchinskii V.* Empirical graph Laplacian approximation of Laplace-Beltrami operators: large sample results // *High Dimension Probability*. — 2006. — Vol. 51. — Pp. 238–259.

24. *Yanovich Y.* Asymptotic Properties of Local Sampling on Manifold // *Journal of Mathematics and Statistics*. — 2016. — Vol. 12, no. 3. — Pp. 157–175.
25. *Yanovich Y.* Asymptotic Properties of Nonparametric Estimation on Manifold // *JMLR Workshop and Conference Proceedings*. — 2017. — Vol. 60. — Pp. 1–21.
26. *Yanovich Y.* Asymptotic Properties of Eigenvalues and Eigenfunctions Estimates of Linear Operators on Manifolds // *Lobachevskii Journal of Mathematics*. — 2017.
27. *Bernstein A., Kuleshov A., Yanovich Y.* Information preserving and locally isometric&conformal embedding via Tangent Manifold Learning // *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on*. — Paris: IEEE, 2015. — oct. — Pp. 1–9.
28. *Bernstein A. V., Kuleshov A. P., Yanovich Y. A.* Locally isometric and conformal parameterization of image manifold // *Proceedings of SPIE 9875, Eighth International Conference on Machine Vision (ICMV 2015) / Ed. by A. Verikas, P. Radeva, D. Nikolaev*. — 2015. — Pp. 1–7.
29. *Bernstein A., Kuleshov A., Yanovich Y.* Manifold Learning in Regression Tasks // *Lecture Notes in Computer Science*. — 2015. — Vol. 9047. — Pp. 414–423.
30. *Bernstein A., Kuleshov A., Yanovich Y.* Statistical Learning via Manifold Learning // *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. — IEEE, 2015. — Pp. 64–69.
31. *Bernstein A., Kuleshov A., Yanovich Y.* Locally Isometric and Conformal Low-dimensional Data Representation in Data Analysis // *Abstracts of Yandex School of Data Analysis Conference “Machine Learning: Prospects and Applications”*. — 2015. — Pp. 87–88.
32. *Bernstein A., Kuleshov A., Yanovich Y.* Nonparametric algorithm for Tangent Bundle Manifold Learning problem // *Twenty-Seventh Annual Conference on Neural Information Processing Systems (NIPS-2013), Workshop ‘Modern Non-parametric Methods in Machine Learning’*. — 2013. — Pp. 1–4.

33. *Bernstein A., Kuleshov A. P., Yanovich Y.* Asymptotically Optimal Method for Manifold Estimation Problem // Abstracts of the 29-th European Meeting of Statistician. — 2013. — P. 325.
34. *Bernstein A., Kuleshov A. P., Yanovich Y.* Asymptotically Optimal Method for Manifold Estimation Problem // Abstracts of the International Workshop on Statistical Learning. — 2013. — Pp. 8–9.
35. *Янович Ю.* Состоятельность оценки области определения алгоритмом спектральных вложений Грассмана-Штифеля // Сборник статей конференции "Информационные технологии и системы"(ИТиС'16). — 2016. — Pp. 191–197.
36. *Янович Ю. А., Киселиус .* Генерация последовательностей случайных точек с заданной плотностью на многообразиях // Сборник статей конференции "Информационные технологии и системы"(ИТиС'15). — 2015. — Pp. 1036–1040.
37. *Янович Ю. А.* Равномерное оценивание касательного к многообразию пространства // Сборник статей конференции "Информационные технологии и системы"(ИТиС'13). — 2013. — Pp. 371–375.
38. *Levina E., Bickel P. J.* Maximum Likelihood Estimation of Intrinsic Dimension // Advances in Neural Information Processing Systems. — MIT Press, 2005. — Pp. 777–784.
39. *Deheuvels P., Puri M. L., Ralescu S. S.* Asymptotic expansions for sums of nonidentically distributed Bernoulli random variables // *Journal of Multivariate Analysis.* — 1989. — Vol. 28, no. 2. — Pp. 282–303.
40. *Niyogi P., Smale S., Weinberger S.* Finding the Homology of Submanifolds with High Confidence from Random Samples // *Discrete & Computational Geometry.* — 2008. — mar. — Vol. 39, no. 1-3. — Pp. 419–441.
41. *Pennec X.* Probabilities And Statistics On Riemannian Manifolds: Basic Tools For Geometric Measurements // *In IEEE workshop on nonlinear signal and image processing.* — 1999. — Pp. 194–198.
42. *Petersen P.* Riemannian Geometry. — Springer New York, 2006. — Vol. 171 of *Graduate Texts in Mathematics.*

43. *Coifman R. R., Lafon S.* Diffusion maps // *Applied and Computational Harmonic Analysis*. — 2006. — jul. — Vol. 21, no. 1. — Pp. 5–30.
44. *Bishop C. M.* Pattern Recognition and Machine Learning. — New York: Springer-Verlag, 2006.
45. *Wasserman L.* All of Nonparametric Statistics. — Springer-Verlag, 2006. — P. 242.
46. *Bernstein A., Kuleshov A.* Manifold Learning: Generalization Ability and Tangent Proximity // *International Journal of Software and Informatics*. — 2013. — Vol. 7, no. 3. — Pp. 359–390.
47. Manifold Learning: The Price of Normalization / Y. Goldberg, A. Zakai, D. Kushnir, Y. Ritov // *J. Mach. Learn. Re.* — 2008. — Vol. 9. — Pp. 1909–1939.
48. *Bernstein A., Kuleshov A.* Manifold Learning: Generalization Ability and Tangent Proximity // *International Journal of Software and Informatics*. — 2013. — Vol. 7, no. 3. — Pp. 359–390.
49. *Zhang Z., Zha H.* Principal Manifolds and Nonlinear Dimensionality Reduction via Local Tangent Space Alignment // *SIAM Journal on Scientific Computing*. — 2004. — Vol. 26, no. 1. — Pp. 313–338.
50. *Schölkopf B., Smola A., Müller K.-R.* Nonlinear Component Analysis as a Kernel Eigenvalue Problem // *Neural Computation*. — 1998. — jul. — Vol. 10, no. 5. — Pp. 1299–1319.
51. Learning Eigenfunctions Links Spectral Embedding and Kernel PCA / Y. Bengio, O. Delalleau, N. L. Roux et al. // *Neural Computation*. — 2004. — oct. — Vol. 16, no. 10. — Pp. 2197–2219.
52. *Bengio Y., Paiement J.-F., Vincent P.* Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering // *In Advances in Neural Information Processing Systems*. — 2003. — Pp. 177–184.
53. *Bunte K., Biehl M., Hammer B.* A General Framework for Dimensionality-Reducing Data Visualization Mapping // *Neural Computation*. — 2012. — mar. — Vol. 24, no. 3. — Pp. 771–804.
54. *Bernstein A., Kuleshov A.* Low-Dimensional Data Representation in Data Analysis. — 2014. — Pp. 47–58.

55. *Kuleshov A., Bernstein A.* Extended Regression on Manifolds Estimation // Lecture Notes in Computer Science. — 2016. — Pp. 208–228.
56. *Franklin J. N.* Matrix Theory. — Dover Publications, 2000.
57. *Шуряев А. Н.* Вероятность. — Москва: Академиздатцентр "Наука 1989. — С. 640.
58. *Петров В. В.* Предельные теоремы для сумм независимых случайных величин. — Наука, 1987. — С. 320.

А Определения и леммы из дифференциальной геометрии

В данной главе вводятся необходимые для дальнейшего рассмотрения сведения, относящиеся к дифференциальной геометрии. Раздел А.1 формализует понятие дифференцирования на многообразиях. Далее, в разделе А.2 вводятся понятия второй фундаментальной формы и кривизны Риччи, которые используются в разделе А.3 для связывания евклидовых расстояний и объемов с соответствующими объектами на многообразии.

А.1 Ковариантное дифференцирование

Дифференцирование многообразий понимается стандартно. Напомним, что это означает. В p -мерном действительном пространстве \mathbb{R}^p для любой точки $Z \in \mathbb{R}^p$ и любого направления $V \in \mathbb{R}^p \setminus \{0\}$ легко определить производную функции по заданному направлению в заданной точке. Однако, для $X \in \mathbb{X}$ малое смещение не во все направления оставляет результат переноса на многообразии. Точнее, для многообразия \mathbb{X} , покрытого картой (\mathbb{B}, f) с полноранговой матрицей Якоби $J_f(f^{-1}(X))$, то в точке $X_0 = f(b_0)$ локально многообразие ведет себя почти как q -мерное линейное пространство — касательное пространство $T_{X_0}(\mathbb{X})$, которое можно определить как линейное пространство, натянутое на p -мерные вектор-столбцы матрицы

$$J_f(b_0) = \left(\frac{\partial f}{\partial b^1}(b_0) \mid \dots \mid \frac{\partial f}{\partial b^q}(b_0) \right),$$

где верхний индекс обозначает компоненту вектора. То есть, дифференцировать можно только по направлениям $V(X_0) \in T_{X_0}(\mathbb{X})$.

Как видно из записи, касательное пространство $T_{X_0}(\mathbb{X})$ зависит от точки многообразия X_0 , в которой оно строится. Поэтому, в общем случае, на многообразии нельзя определить производную скалярной функции $\varphi(X)$ как предел отношений изменения функции φ при переходе от точки X_0 к точке $X_0 + tV(X_0)$ к длине $tV(X_0)$, так как $X_0 + tV(X_0) \notin \mathbb{X}$ и значения функции ε в этой точке не определена в общем случае. Поэтому, вместо $X_0 + tV(X_0)$ рассматривают кривую $\gamma(t)$, $t \in (-\varepsilon, \varepsilon)$ на многообразии такую, что $\gamma(0) = X_0$ и $\frac{\partial \gamma}{\partial t}(0) = V(X_0)$.

Производная скалярной функции определяется как

$$\nabla_{V(X_0)}\varphi(X_0) = \lim_{t \rightarrow 0} \frac{\varphi(\gamma(t)) - \varphi(\gamma(0))}{t} = \lim_{t \rightarrow 0} \frac{\varphi(\gamma(t)) - \varphi(X_0)}{t}.$$

Замечание 7. В случае евклидова пространства \mathbb{R}^m существует его тождественное накрытие: $\mathbb{B} = \mathbb{R}^m$ и $f(b) = b$, то есть $X = b$. Поэтому в качестве кривой $\gamma(t)$ можно выбрать $\gamma(t) = X_0 + tV(X_0)$ и ковариантная производная функции φ совпадает с обычной производной по направлению.

Рассмотрим ограничение векторного поля $V(X) \in T_X(\mathbb{X})$, $X \in \mathbb{X}$ на кривую $\gamma(t)$: $\tilde{V}(t) = V(\gamma(t))$, $t \in (-\varepsilon, \varepsilon)$. Производная $\tilde{V}(t)$ может быть определена как обычно:

$$\frac{\partial \tilde{V}}{\partial t}(t) = \lim_{h \rightarrow 0} \frac{\tilde{V}(t+h) - \tilde{V}(t)}{h},$$

где $t \in (-\varepsilon, \varepsilon)$. Однако, производная $\frac{\partial \tilde{V}}{\partial t}(t)$ может не лежать в $T_{\gamma(t)}(\mathbb{X})$. Поэтому определяют ковариантную производную $\frac{D\tilde{V}}{dt}(h)$ как проекцию $\frac{\partial \tilde{V}}{\partial t}(t)$ на $T_{\gamma(t)}(\mathbb{X})$. Далее, рассмотрим уравнение

$$\begin{cases} \frac{D\tilde{W}}{dt}(t) = 0 \\ \tilde{W}(0) = \tilde{W} \end{cases}, \quad (60)$$

где $\tilde{W} \in T_{\gamma(0)}(\mathbb{X})$. Решение $\tilde{W}(t)$ уравнения (60) существует и называется параллельным переносом вектора $\tilde{W}(\gamma(0))$ вдоль кривой $\gamma(t)$ и обозначается $\tilde{W}(t) = P_{\gamma(t), \gamma(0)}\tilde{W}$.

Теперь, чтобы определить производную векторного поля $W(X)$, $X \in \mathbb{X}$, вдоль кривой $\gamma(t)$, $t \in (-\varepsilon, \varepsilon)$ будем параллельно переносить $W(\gamma(t))$ в точку $X_0 = \gamma(0)$. Результат параллельного переноса $P_{\gamma(0), \gamma(t)}W(\gamma(t)) \in T_{\gamma(0)}(\mathbb{X})$, поэтому в $T_{\gamma(0)}(\mathbb{X})$ определена разность $P_{\gamma(0), \gamma(t)}W(\gamma(t)) - W(\gamma(0))$ и определена ковариантная производная векторного поля $W(X)$ на \mathbb{X} :

$$\nabla_{V(X_0)}W(X_0) = \lim_{t \rightarrow 0} \frac{P_{\gamma(0), \gamma(t)}W(\gamma(t)) - W(\gamma(0))}{t},$$

где $\gamma: (-\varepsilon, \varepsilon) \rightarrow \mathbb{X}$ и $\gamma(0) = X_0 \in \mathbb{X}$, $\gamma'(0) = V(X_0) \in T_{X_0}(\mathbb{X})$.

Замечание 8. В случае евклидова пространства \mathbb{R}^m решением уравнения

(60) является $P_{\gamma(0),\gamma(t)}W(\gamma(t)) = W(\gamma(t)) = W(X_0 + tV(X_0))$ и ковариантная производная векторного совпадает с обычной производной по направлению:

$$\nabla_{V(X_0)}W(X_0) = \lim_{t \rightarrow 0} \frac{W(X_0 + tV(X_0)) - W(X_0)}{t}.$$

Будем обозначать ковариантную производную в направлении $V \in T_X(\mathbb{X}) \setminus \{0\}$ в точке X как ∇_V . Ковариантная производная обладает привычными свойствами производной по направлению, например линейностью $V_1, V_2 \in T_X(\mathbb{X}) \setminus \{0\}$, $\alpha_1, \alpha_2 \in \mathbb{R} \setminus 0$ при $\alpha_1 V_1 + \alpha_2 V_2 \neq 0$:

$$\nabla_{\alpha_1 V_1 + \alpha_2 V_2} = \alpha_1 \nabla_{V_1} + \alpha_2 \nabla_{V_2}.$$

А.2 Квадратичные формы многообразия и кривизна Риччи

В работе предполагается, что метрический тензор многообразия \mathbb{X} порожден его вложением в многомерное пространство \mathbb{R}^p . Это означает, что скалярное произведение для векторов $V, W \in T_X(\mathbb{X})$ является ограничением скалярного произведения \mathbb{R}^p на $T_X(\mathbb{X})$, то есть в ортонормированном базисе выражается билинейной формой $I_X(V, W) = V^T W$.

Первая квадратичная форма многообразия — это форма длины векторов касательного пространства (и позволяет вычислять длины кривых):

$$I_X(V, V) = V^T V.$$

Запишем координаты вектора V в базисе, заданном параметризацией (\mathbb{B}, f) :

$$\begin{aligned} b &= f^{-1}(X); \\ V &= J_f(b)\alpha_V, \quad \alpha_V \in \mathbb{R}^q, \\ \alpha_V &= (J_f(b)^T J_f(b))^{-1} J_f(b)^T V; \\ I_X^c(\alpha_v) &= \alpha_v^T (J_f(b)^T J_f(b)) \alpha_v. \end{aligned}$$

Вторая билинейная форма многообразия показывает ортогональную касательному пространству составляющую изменения векторов касательного про-

странства вдоль разных направлений касательного пространства $V, W \in T_X(\mathbb{X})$. Запишем выражение второй билинейной формы $II_X(V, W)$ в базисе, заданном параметризацией (\mathbb{B}, f) , в точке $X = f(b)$:

$$II_X(V, W) = \pi(X)^\perp \cdot \sum_{i,j=1}^q \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{V,i} \alpha_{W,j}, \quad (61)$$

где $\alpha_V = (\alpha_{V,1}, \dots, \alpha_{V,q})^T$ и $\alpha_W = (\alpha_{W,1}, \dots, \alpha_{W,q})^T$ — соответственно координаты векторов V и W в параметрическом базисе, $\frac{\partial^2 f(b)}{\partial b_i \partial b_j} \in \mathbb{R}^p$,

$$\pi(X)^\perp = \left(I - J_f(b) (J_f^T(b) J_f(b))^{-1} J_f(b)^T \right) \quad (62)$$

— проектор на кокасательное пространство $(T_X(\mathbb{X}))^\perp$.

Вторая квадратичная форма $II_X(V, V)$ выражает нормальную кривизну многообразия.

Для определения кривизны Риччи нам введем обозначения для символов Кристоффеля. Определим

$$g_{ij} = \frac{\partial f_j(b)}{\partial b_i};$$

$$\Gamma_{jk,l} = \frac{1}{2} \left(\frac{\partial g_{jl}}{\partial b_k} + \frac{\partial g_{kl}}{\partial b_j} - \frac{\partial g_{jk}}{\partial b_l} \right). \quad (63)$$

$\Gamma_{jk,l}$ — символы Кристоффеля первого рода. Обозначим g^{ij} элементы обратной матрицы $J_f^T(b) J_f(b)$, то есть элементы матрицы $(J_f^T(b) J_f(b))^{-1}$. Символы Кристоффеля второго рода

$$\Gamma_{jk}^i = \sum_{l=1}^q \Gamma_{jk,l} \cdot g^{li}. \quad (64)$$

Элементы тензора кривизны Риччи определяются как

$$R_{ij} = \sum_{k=1}^q \left(\frac{\partial \Gamma_{ji}^k}{\partial b_k} - \frac{\partial \Gamma_{ki}^j}{\partial b_j} + \sum_{l=1}^q \Gamma_{kl}^k \Gamma_{ji}^l - \Gamma_{jl}^k \Gamma_{ki}^l \right), \quad i, j = 1, \dots, q. \quad (65)$$

Обозначим $R = (R_{ij})_{i,j=1}^q$ — матрица порядка q . Для $\theta \in T_X(\mathbb{X})$:

$\alpha_\theta = (J_f(b)^T J_f(b))^{-1} J_f(b)^T \theta$. Кривизна Риччи в направлении θ в точке X :

$$Ric_X(\theta, \theta) = \alpha_\theta^T \cdot R \cdot \alpha_\theta. \quad (66)$$

Кривизна Риччи характеризует отличие Евклидового элемента объема от элемента объема многообразия.

А.3 Локально-римановы координаты и экспоненциальное отображение

Для малой окрестности точки $X_0 \in \mathbb{X}$ отображение $X \mapsto V \in T_X(\mathbb{X})$: $X = f(\gamma(1))$, где $\gamma: \gamma(0) = X_0$ и $\frac{\partial \gamma}{\partial t}(0) = V$, определяет взаимнооднозначное соответствие между окрестностью точки X_0 и окрестностью точки 0 в $T_{X_0}(\mathbb{X})$. Обратное отображение называют экспоненциальным и обозначают

$$X = \exp_{X_0}(V). \quad (67)$$

В данной окрестности векторы $V \in T_{X_0}(\mathbb{X})$ задают координаты. Эти координаты называются локально римановыми координатами.

Замечание 9. В случае евклидового пространства $\mathbb{M} = \mathbb{R}^m$, касательное пространство $T_X(\mathbb{X}) = \mathbb{R}^m$ и $\exp_{X_0}(V) = X_0 + V$ для любого $V \in T_X(\mathbb{X})$ (то есть, результат справедлив для произвольной окрестности X_0).

Свяжем теперь расстояние между близкими точками многообразия в p -мерном пространстве и расстояние в между их q мерными римановыми координатами связаны Леммами 1 и 2. При этом Леммы 3 и 4 (грубо) оценивают сверху вторую фундаментальную форму и кривизну Риччи через достаточно гладкую параметризацию многообразия. Докажем их:

Доказательство леммы 3. Для произвольной точки $b \in \mathbb{B}$ обозначим V_1, \dots, V_q — ортонормированный базис собственных векторов матрицы $J_f(b)^T J_f(b)$. Обозначим $\lambda_1, \dots, \lambda_q$ соответствующие V_1, \dots, V_q собственные значения.

Запишем координаты $\theta \in T_{f(b)}(\mathbb{X})$: $|\theta| = 1$ в базисе V_1, \dots, V_q : $\theta =$

$\sum_{i=1}^q \beta_i V_i$, где $\sum_{i=1}^q \beta_i^2 = 1$. Учитывая (7), преобразуем

$$\begin{aligned} \alpha_\theta &= (J_f(b)^T J_f(b))^{-1} J_f(b)^T \theta = (J_f(b)^T J_f(b))^{-1} J_f(b)^T \sum_{i=1}^q \beta_i V_i; \\ |\alpha_\theta|^2 &= \sum_{i=1}^q \frac{1}{\lambda_i} \beta_i^2; \\ |\alpha_\theta|^2 &\leq \frac{1}{c_J}. \end{aligned} \quad (68)$$

Учитывая $X = f(b)$, (9), (68), оценим

$$\begin{aligned} \|\Pi_X(\theta, \theta)\| &= \left\| \pi^\perp(X) \cdot \sum_{i,j=1}^q \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{\theta,i} \alpha_{\theta,j} \right\| \leq \\ &\leq \left\| \sum_{i,j=1}^q \frac{\partial^2 f(b)}{\partial b_i \partial b_j} \alpha_{\theta,i} \alpha_{\theta,j} \right\| \leq C_H \cdot \|\alpha_\theta\|_1^2 \leq C_H \cdot q \cdot \frac{1}{c_J}. \end{aligned}$$

Лемма доказана. □

Доказательство леммы 4. Учитывая (63) и (64), оценим элемент матрицы кривизны (65) по неравенству треугольника:

$$\begin{aligned} |R_{ij}| &= \left| \sum_{k=1}^q \left(\frac{\partial \Gamma_{ji}^k}{\partial b_k} - \frac{\partial \Gamma_{ki}^k}{\partial b_j} + \sum_{l=1}^q \Gamma_{kl}^k \Gamma_{ji}^l - \Gamma_{jl}^k \Gamma_{ki}^l \right) \right| \leq \\ &\leq 2q \cdot \max_{i,j,k=1,\dots,q} \left| \frac{\partial \Gamma_{ji}^k}{\partial b_k} \right| + 2q^2 \cdot \left(\max_{j,i,k=1,\dots,q} |\Gamma_{ji}^k| \right)^2; \\ |\Gamma_{jk}^i| &= \left| \sum_{l=1}^q \Gamma_{jk,l} g^{li} \right| = \left| \sum_{l=1}^q \frac{1}{2} \left(\frac{\partial g_{jl}}{\partial b_k} + \frac{\partial g_{kl}}{\partial b_j} - \frac{\partial g_{jk}}{\partial b_l} \right) g^{li} \right| \leq \\ &\leq \frac{3q}{2} \cdot \max_{k,l,j=1,\dots,q} \left| \frac{\partial g_{jl}}{\partial b_k} \right| \cdot \max_{k,l=1,\dots,q} |g^{kl}|; \\ \left| \frac{\partial g_{jl}}{\partial b_k} \right| &= \left| \frac{\partial \left(\frac{\partial f(b)}{\partial b_j}, \frac{\partial f(b)}{\partial b_l} \right)}{\partial b_k} \right| \leq \\ &\leq 2 \cdot \max_{j,k=1,\dots,q} \left\| \frac{\partial^2 f(b)}{\partial b_j \partial b_k} \right\| \cdot \max_{j=1,\dots,q} \left\| \frac{\partial f(b)}{\partial b_j} \right\| \leq 2 \cdot C_H \cdot \sqrt{C_J}; \\ \max_{k,l=1,\dots,q} |g^{kl}| &\leq 1/\sqrt{C_J}; \end{aligned}$$

$$\begin{aligned}
\left| \frac{\partial \Gamma_{ji}^k}{\partial b_k} \right| &= \left| \sum_{l=1}^q \frac{\partial \Gamma_{jk,l}}{\partial b_k} g^{li} - \sum_{l_1, i_1=1}^q \Gamma_{jk,l} \cdot g^{ll_1} \frac{\partial g_{l_1 i_1}}{\partial b_k} g^{i_1 i} \right| \leq \\
&\leq q \cdot \max_{j,k,l=1,\dots,q} \left| \frac{\partial \Gamma_{jk,l}}{\partial b_k} \right| \cdot \max_{k,l=1,\dots,q} |g^{kl}| + \\
&+ q^3 \cdot \max_{j,k,l=1,\dots,q} |\Gamma_{jk,l}| \cdot \left(\max_{k,l=1,\dots,q} |g^{kl}| \right)^2 \cdot \max_{j,l,k=1,\dots,q} \left| \frac{\partial g_{jl}}{\partial b_k} \right| \leq \\
&\leq q/\sqrt{c_J} \cdot \max_{j,k,l=1,\dots,q} \left| \frac{\partial \Gamma_{jk,l}}{\partial b_k} \right| + 2q^3 \cdot C_H \cdot \sqrt{C_J}/c_J; \\
\left| \frac{\partial \Gamma_{jk,l}}{\partial b_k} \right| &\leq 3/2 \cdot \max_{i,j,k,m=1,\dots,q} \left| \frac{\partial^2 \left(\frac{\partial f(b)}{\partial b_j}, \frac{\partial f(b)}{\partial b_l} \right)}{\partial b_k \partial b_m} \right| \leq 3 \cdot \max_{i,j,k,m=1,\dots,q} \left| \frac{\partial \left(\frac{\partial^2 f(b)}{\partial b_j \partial b_k}, \frac{\partial f(b)}{\partial b_l} \right)}{\partial b_m} \right| \leq \\
&\leq 3 \cdot \left(C_H^2 + C_T \cdot \sqrt{C_J} \right).
\end{aligned}$$

Таким образом,

$$|R_{ij}| \leq 2q^2 \cdot \frac{C_H^2 + C_T \sqrt{C_J}}{\sqrt{c_J}} + q^4 \cdot \left(18 \cdot C_H^2 \frac{C_J}{c_J} + 4 \cdot C_H \frac{\sqrt{C_J}}{c_J} \right).$$

Подставляя в (66) и учитывая (68), получаем:

$$\begin{aligned}
\|Ric_X(\theta, \theta)\| &= \|\alpha_\theta^T \cdot R \cdot \alpha_\theta\| \leq \max_{i,j=1,\dots,q} |R_{ij}| \cdot \|\alpha_\theta\|_1^2 \leq \\
&\leq 2q^3 \cdot \frac{C_H^2 + C_T \sqrt{C_J}}{c_J^{3/2}} + q^5 \cdot \left(18 \cdot C_H^2 \frac{C_J}{c_J} + 4 \cdot C_H \frac{\sqrt{C_J}}{c_J} \right) / c_J.
\end{aligned}$$

Лемма доказана. □

В Доказательство лемм

В этой главе перечислены леммы, используемые при доказательстве основных теорем, и доказаны наиболее важные из них.

В.1 Леммы Муавра-Лапласа для медленно убывающей вероятности

Докажем локальную и интегральные леммы Муавра-Лапласа для параметра успеха, медленно стремящегося к нулю.

Доказательство леммы 5. Доказательство во многом повторяет доказательство локальной предельной теоремы из [57] и существенно использует формулу Стирлинга

$$N! = \sqrt{2\pi N} \cdot e^{-N} N^N \cdot (1 + R(N)),$$

где $R(N) = \frac{1}{12N} + \frac{1}{288N^2} - \frac{139}{51840N^3} + O(\frac{1}{N^4})$, и $\frac{1}{12N} > R(N) > \frac{1}{13N}$, $N \geq 1$ может быть выражена через числа Бернулли. Для удобства будем опускать индекс N у p_N и q_N .

Откуда, если $N \rightarrow \infty, k \rightarrow \infty, N - k \rightarrow \infty$:

$$C_N^k = \frac{N!}{k!(N-k)!} = \frac{\sqrt{2\pi N} \cdot e^{-N} N^N}{\sqrt{2\pi k} \cdot 2\pi(N-k) \cdot e^{-k} k^k \cdot e^{-(N-k)} (N-k)^{N-k}} \cdot \frac{1 + R(N)}{(1 + R(k))(1 + R(N-k))} = \frac{1}{\sqrt{2\pi N \frac{k}{N}} \cdot (1 - \frac{k}{N})} \cdot \frac{1 + r}{(\frac{k}{N})^k (1 - \frac{k}{N})^{N-k}},$$

где $r = r(N, k, N - k) = \frac{1 + R(N)}{(1 + R(k))(1 + R(N-k))}$. Оценим r для $k > 1, N - k > 1$:

$$1 > \frac{1 + R(N)}{(1 + R(k))(1 + R(N-k))} > \left(1 + \frac{1}{13N}\right) \left(1 - \frac{1}{12k}\right) \cdot \left(1 - \frac{1}{12(N-k)}\right) > 1 - \frac{1}{12k} - \frac{1}{12(N-k)};$$

$$|r| \leq \frac{1}{12k} + \frac{1}{12(N-k)}.$$

Поэтому

$$p(k) = C_N^k p^k q^{N-k} = \frac{1}{\sqrt{2\pi N \frac{k}{N}} (1 - \frac{k}{N})} \cdot \frac{p^k (1-p)^{N-k}}{(\frac{k}{N})^k (1 - \frac{k}{N})^{N-k}} \cdot (1 + r).$$

Обозначим $\hat{p} = \frac{k}{N}$, $\hat{q} = 1 - \hat{p}$. Заметим, что $p - \hat{p} = \hat{q} - q$, так как

$$p + q = \hat{p} + \hat{q} = 1.$$

Также, $\frac{p-\hat{p}}{p}$ и $\frac{q-\hat{q}}{q} = -\frac{p-\hat{p}}{q}$ являются малыми параметрами, так как $\frac{p-\hat{p}}{pq} = \frac{Np-k}{Npq}$ мало по предположению леммы и $0 < p, q < 1$. Тогда

$$\begin{aligned} p(k) &= \frac{1}{\sqrt{2\pi N\hat{p}(1-\hat{p})}} \left(\frac{p}{\hat{p}}\right)^k \left(\frac{1-p}{1-\hat{p}}\right)^{N-k} \cdot (1+r) = \\ &= \frac{1}{\sqrt{2\pi N\hat{p}(1-\hat{p})}} \exp\left(k \ln \frac{p}{\hat{p}} + (N-k) \ln \frac{1-p}{1-\hat{p}}\right)^{N-k} \cdot (1+r) = \\ &= \frac{1}{\sqrt{2\pi N\hat{p}(1-\hat{p})}} \exp\left(N \cdot \left(\hat{p} \ln \frac{p}{\hat{p}} + \hat{q} \ln \frac{q}{\hat{q}}\right)\right) \cdot (1+r) = \\ &= \frac{1}{\sqrt{2\pi N\hat{p}(1-\hat{p})}} \exp\left(-N \cdot \left(p \cdot \left(1 + \frac{\hat{p}-p}{p}\right) \ln \left(1 + \frac{\hat{p}-p}{p}\right) + \right. \right. \\ &\quad \left. \left. + q \cdot \left(1 - \frac{\hat{p}-p}{q}\right) \ln \left(1 - \frac{\hat{p}-p}{q}\right)\right)\right) \cdot (1+r). \end{aligned}$$

Воспользуемся разложением по формуле Тейлора с остаточным членом в форме Лагранжа

$$\begin{aligned} \left(1 + \frac{\hat{p}-p}{p}\right) \ln \left(1 + \frac{\hat{p}-p}{p}\right) &= -\frac{\hat{p}-p}{p} + \frac{1}{2} \cdot \left(\frac{\hat{p}-p}{p}\right)^2 + \\ &\quad + \left(\frac{1}{2} - \frac{1}{3} \cdot \frac{1}{(1+\delta_p)^3}\right) \cdot \left(\frac{\hat{p}-p}{p}\right)^3; \\ \left(1 - \frac{\hat{p}-p}{q}\right) \ln \left(1 - \frac{\hat{p}-p}{q}\right) &= \frac{\hat{p}-p}{q} + \frac{1}{2} \cdot \left(\frac{\hat{p}-p}{q}\right)^2 + \\ &\quad + \left(-\frac{1}{2} + \frac{1}{3} \cdot \frac{1}{(1-\delta_q)^3}\right) \cdot \left(\frac{\hat{p}-p}{q}\right)^3; \\ p \left(1 + \frac{\hat{p}-p}{p}\right) \ln \left(1 + \frac{\hat{p}-p}{p}\right) + q \left(1 - \frac{\hat{p}-p}{q}\right) \ln \left(1 - \frac{\hat{p}-p}{q}\right) &= \\ &= \frac{1}{2} \left(\frac{1}{p} + \frac{1}{q}\right) (\hat{p}-p)^2 + \left(\frac{1}{2} - \frac{1}{3} \cdot \frac{1}{(1+\delta_p)^2}\right) \cdot \frac{(\hat{p}-p)^3}{p^2} + \\ &\quad + \left(-\frac{1}{2} + \frac{1}{3} \cdot \frac{1}{(1-\delta_q)^2}\right) \cdot \frac{(\hat{p}-p)^3}{q^2}, \end{aligned}$$

где $\delta_p \in \left[\min\{0, \frac{\hat{p}-p}{p}\}, \max\{0, \frac{\hat{p}-p}{p}\} \right]$ и $\delta_q \in \left[\min\{0, \frac{\hat{p}-p}{q}\}, \max\{0, \frac{\hat{p}-p}{q}\} \right]$.
 Также, $\frac{1}{p} + \frac{1}{q} = \frac{p+q}{pq} = \frac{1}{pq}$, откуда

$$\frac{n}{2} \left(\frac{1}{p} + \frac{1}{q} \right) (\hat{p} - p)^2 = \frac{N}{2pq} \left(\frac{k}{N} - p \right)^2 = \frac{(k - Np)^2}{2Npq}.$$

Откуда

$$\frac{N}{2pq} \left(\frac{k}{N} - p \right)^2 = \frac{(k - Np)^2}{2Npq}.$$

Получаем

$$p(k) = \frac{1}{\sqrt{2\pi Npq}} \exp \left(-\frac{(k - Np)^2}{2Npq} \right) \cdot (1 + \tilde{r}(N, k, N - k)),$$

где

$$1 + \tilde{r}(N, k, N - k) = (1 + r(N, k, N - k)) \exp \left(-N \cdot \left(\left(\frac{1}{2} - \frac{1}{3} \cdot \frac{1}{(1 + \delta_p)^2} \right) \cdot \frac{(\hat{p} - p)^3}{p^2} + \left(-\frac{1}{2} + \frac{1}{3} \cdot \frac{1}{(1 - \delta_q)^2} \right) \cdot \frac{(\hat{p} - p)^3}{q^2} \right) \sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}} \right).$$

Откуда

$$\sup |\tilde{r}(N, k, N - k)| \rightarrow 0, \quad N \rightarrow \infty$$

для $k: |k - Np|/(Npq)^{2/3} \rightarrow 0$. □

Доказательство леммы 6. Пусть $\infty < a \leq b < \infty$

$$P_N(a, b) = \sum_{a < z \leq b} P_N(Np_N + z\sqrt{Np_Nq_N}),$$

где суммирование ведется по таким z , что $Np_N + z\sqrt{Np_Nq_N}$ — целые числа.

Из локальной леммы 5 следует, что для всех $t_k: k = Np_N + t_k\sqrt{Np_Nq_N}$ и удовлетворяющих условию $|t_k| \leq T < \infty$,

$$P_N(Np_N + t_k\sqrt{Np_Nq_N}) = \frac{\delta_k}{\sqrt{2\pi}} \exp(-t_k^2/2) (1 + \varepsilon(t_k, N)),$$

где

$$\sup_{|t_k| \leq T} |\varepsilon(t_k, N)| \rightarrow 0, \quad N \rightarrow \infty$$

$$\text{и } \delta_k = \frac{1}{\sqrt{Np_Nq_N}}.$$

Следовательно, для фиксированных a и b таких, что $-T \leq a \leq b \leq T$, $T \leq \infty$,

$$\begin{aligned} & \sum_{a < t_k \leq b} P_N \left(Np + t_k \sqrt{Npq} \right) = \\ &= \sum_{a < t_k \leq b} \frac{\delta_k}{\sqrt{2\pi}} \exp(-t_k^2/2) + \sum_{a < t_k \leq b} \varepsilon(t_k, N) \frac{\delta_k}{\sqrt{2\pi}} \exp(-t_k^2/2) = \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-z^2/2) dx + R_n^{(1)}(a, b) + R_N^{(2)}(a, b), \end{aligned}$$

где

$$\begin{aligned} R_N^{(1)}(a, b) &= \sum_{a < t_k \leq b} \frac{\delta_k}{\sqrt{2\pi}} \exp(-t_k^2/2) - \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-z^2/2) dx; \\ R_N^{(2)}(a, b) &= \sum_{a < t_k \leq b} \varepsilon(t_k, N) \frac{\delta_k}{\sqrt{2\pi}} \exp(-t_k^2/2). \end{aligned}$$

Из свойств интегральных сумм

$$\sup_{-T \leq a \leq b \leq T} |R_N^{(1)}(a, b)| \rightarrow 0, \quad N \rightarrow \infty.$$

Учитывая неотрицательность подынтегральной функции:

$$\frac{1}{\sqrt{2\pi}} \int_{-T}^T \exp(-z^2/2) dz \leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-z^2/2) dx.$$

То есть,

$$\begin{aligned} & \sup_{-T \leq a \leq b \leq T} |R_n^{(2)}(a, b)| \leq \sup_{|t_k| \leq T} |\varepsilon(t_k, n)| \cdot \\ & \cdot \left(\frac{1}{\sqrt{2\pi}} \int_{-T}^T \exp(-z^2/2) dz + \sup_{-T \leq a \leq b \leq T} |R_n^{(1)}(a, b)| \right) \rightarrow 0. \end{aligned} \quad (69)$$

Обозначим

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp(-t^2/2) dt.$$

Получаем

$$\sup_{-T \leq a \leq b \leq T} |P_N(a, b) - (\Phi(b) - \Phi(a))| \rightarrow 0, \quad N \rightarrow \infty.$$

Докажем, что результат справедлив и для $T = \infty$: $\forall \varepsilon > 0$
 $\exists T = T(\varepsilon) > 0$:

$$\frac{1}{\sqrt{2\pi}} \int_{-T}^T \exp(-z^2/2) dz > 1 - \frac{\varepsilon}{4}$$

Из 69:

$$\exists N' : \forall N > N' : \sup_{-T \leq a \leq b \leq T} |P_N(a, b) - (\Phi(b) - \Phi(a))| < \frac{\varepsilon}{4}.$$

Откуда

$$\begin{aligned} P_N(-T, T] &> 1 - \frac{\varepsilon}{2}, \\ P_N(-\infty, -T] + P_N[T, \infty) &< \frac{\varepsilon}{2}, \end{aligned}$$

где $P_N(-\infty, -T] = \lim_{S \rightarrow -\infty} P_N(S, -T]$ и $P_N(T, \infty) = \lim_{S \rightarrow -\infty} P_N[T, S)$.

То есть, $\forall T : -\infty \leq -T \leq a \leq b \leq \infty$

$$\begin{aligned}
& \left| P_N(a, b) - \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-z^2/2) dz \right| \leq \\
& \leq \left| P_N(-T, T] - \frac{1}{\sqrt{2\pi}} \int_{-T}^T \exp(-z^2/2) dz \right| + \\
& + \left| P_N(a, -T] - \frac{1}{\sqrt{2\pi}} \int_a^{-T} \exp(-z^2/2) dz \right| + \\
& + \left| P_N(T, b) - \frac{1}{\sqrt{2\pi}} \int_T^b \exp(-z^2/2) dz \right| \leq \\
& \leq \frac{\varepsilon}{4} + P_N(-\infty, -T] + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-T} \exp(-z^2/2) dz + \\
& + P_N(T, \infty) + \frac{1}{\sqrt{2\pi}} \int_T^{\infty} \exp(-z^2/2) dz \leq \\
& \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{8} + \frac{\varepsilon}{8} = \varepsilon.
\end{aligned}$$

То есть, $P_N(a, b]$ стремится к $\Phi(b) - \Phi(a)$ по всем $-\infty \leq a < b \leq \infty$. \square

В.2 Вероятность больших уклонений для ограниченных случайных величин

Теорема 13. (из [58]) Пусть χ_1, \dots, χ_n — независимые случайные величины, $\mathbb{E}\chi_k = 0$, $\sigma_k^2 = \mathbb{E}\chi_k^2 < \infty$, $k = 1, \dots, n$, $B = \sum_{k=1}^n \sigma_k^2$. Пусть существует такая константа $H > 0$, что

$$|\mathbb{E}\chi_k^m| \leq \frac{m!}{2} \cdot \sigma_k^2 \cdot H^{m-2}, \quad k = 1, \dots, n$$

для всех целых $m \geq 2$. Тогда для $S = \sum_{k=1}^n \chi_k$ при $0 \leq x \leq B/H$:

$$P(S \geq x) \leq \exp(-x^2/4B),$$

$$P(S \leq -x) \leq \exp(-x^2/4B)$$

и при $x \geq B/H$:

$$P(S \geq x) \leq \exp(-x/4H),$$

$$P(S \leq -x) \leq \exp(-x/4H).$$

Доказательство. приведено в [58] в главе 3, теорема 18. □

Докажем лемму о вероятности больших уклонений для ограниченных случайных величин:

Доказательство леммы 7. Преобразуем

$$\begin{aligned}\bar{\chi} &= \frac{1}{n} \sum_{k=1}^n \chi_k - m_k = \frac{1}{n} \sum_{k=1}^n (\chi_k - \mathbb{E}\chi_k) + (\mathbb{E}\chi_k - m_k) = \\ &= \frac{1}{n} \sum_{k=1}^n (\chi_k - \mathbb{E}\chi_k) + \frac{1}{n} \sum_{k=1}^n (\mathbb{E}\chi_k - m_k).\end{aligned}$$

Для первой суммы, умноженной на n , воспользуемся теоремой 13. Для этого проверим условия теоремы:

$$\begin{aligned}\mathbb{E}(\chi_k - \mathbb{E}\chi_k) &= 0; \\ \mathbb{E}(\chi_k - \mathbb{E}\chi_k)^2 &= \sigma^2 < \infty; \\ |\mathbb{E}(\chi_k - \mathbb{E}\chi_k)^m| &\leq \mathbb{E} |\chi_k - \mathbb{E}\chi_k|^m \leq \mathbb{E} |\chi_k - \mathbb{E}\chi_k|^2 \cdot C^{m-2} \leq \\ &\leq \sigma^2 \cdot C^{m-2} \leq \frac{m!}{2} \sigma^2 C^{m-2}.\end{aligned}$$

Таким образом, условия теоремы 13 выполнены для $H \geq 2C$.

Применяя теорему 13 для $\chi_1 - \mathbb{E}\chi_1, \dots, \chi_n - \mathbb{E}\chi_n$ и учитывая, что $|\frac{1}{n} \sum_{k=1}^n (\chi_k - \mathbb{E}\chi_k) - \bar{\chi}| \leq m$ получаем требуемые в лемме неравенства. □

В.3 Замена области интегрирования

Доказательство леммы 8. Преобразуем разность рассматриваемых в лемме интегралов для $\varepsilon < \frac{1}{c(\mathbb{X})}$:

$$\begin{aligned}& \int_{B_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) = \\ &= \int_{B_\varepsilon(X) \setminus \tilde{B}_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X) \setminus B_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}).\end{aligned}$$

Из леммы 2 следует, что каждый элемент симметрической разности

$B_\varepsilon(X) \Delta \tilde{B}_\varepsilon(X)$ отстоит от поверхности шара $\tilde{B}_\varepsilon(X)$ не более чем на

$$\tilde{\varepsilon} = \frac{1}{24} \|\text{II}_{\tilde{X}}(\tilde{\theta}, \tilde{\theta})\| \varepsilon^3$$

для некоторых $\tilde{X} \in \mathbb{X}$ и $\tilde{\theta} \in T_{\tilde{X}}(\mathbb{X}) : \|\tilde{\theta}\| = 1$. По лемме 3: $\tilde{\varepsilon} \leq \frac{1}{24} C_{II} \varepsilon^3$.

$$\text{При } \varepsilon \leq \sqrt{\frac{24 \cdot (\sqrt[3]{2}-1)}{\max\{1, C_{II}\}}}:$$

$$(\varepsilon + \tilde{\varepsilon})^q \leq 2 \cdot \varepsilon^q.$$

$$\text{При } \varepsilon \leq \frac{1}{2\sqrt{C_{Ric}}}:$$

$$1 + C_{Ric} (\varepsilon + \tilde{\varepsilon})^2 \leq 2.$$

Учитывая леммы 1 и 4:

$$\begin{aligned} & \left| \int_{B_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_\varepsilon(X)} g(X, \tilde{X}) dV(\tilde{X}) \right| \leq \\ & \leq \left| \int_{\tilde{B}_{\varepsilon+\tilde{\varepsilon}}(X)} g(X, \tilde{X}) dV(\tilde{X}) - \int_{\tilde{B}_{\varepsilon-\tilde{\varepsilon}}(X)} g(X, \tilde{X}) dV(\tilde{X}) \right| \leq \\ & \leq ((\varepsilon + \tilde{\varepsilon})^q - (\varepsilon - \tilde{\varepsilon})^q) \cdot V_q \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \cdot (1 + (\varepsilon + \tilde{\varepsilon})^2 \cdot C_{Ric}) \leq \\ & = V_q \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \cdot 2\varepsilon \cdot (1 + 4 \cdot \varepsilon^2 \cdot C_{Ric}) \cdot \sum_{m=0}^{q-1} (\varepsilon + \tilde{\varepsilon})^m \cdot (\varepsilon - \tilde{\varepsilon})^{q-m-1} \leq \\ & \leq 2 \cdot V_q \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \cdot \tilde{\varepsilon} \cdot q \cdot (\varepsilon + \tilde{\varepsilon})^{q-1} (1 + 4 \cdot \varepsilon^2 \cdot C_{Ric}) \leq \\ & \leq \varepsilon^{q+2} \cdot 2 \cdot V_q \cdot \left(1 + \frac{1}{24} C_{II} \varepsilon^2\right)^q \cdot (1 + 4 \cdot \varepsilon^2 \cdot C_{Ric}) \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \leq \\ & \leq 8 \cdot V_q \cdot \sup_{X, \tilde{X}} |g(X, \tilde{X})| \cdot \varepsilon^{q+2}. \end{aligned}$$

Лемма доказана. □

В.4 Конечные сети

Для доказательства равномерности оценок будем использовать дополнительное построение. δ -сеть метрического пространства \mathbb{Z} есть множество $\mathbb{Z}_{net}(\delta) \subset \mathbb{Z}$ такое, что для любой точки $Z \in \mathbb{Z}$ найдётся точка $Z_{net} \in \mathbb{Z}_{net}(\delta)$, удалённая от Z не более чем δ .

Лемма 11. Для каждого $\delta > 0$ найдётся конечная δ -сеть многообразия \mathbb{X} ,

содержащая не более $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ элементов, где $a > 0$ — ребро описанного вокруг \mathbb{X} p -мерного гиперкуба.

Доказательство. Так как собственные числа матрицы Якоби параметризации (\mathbb{B}, f) ограничены по предположению М4 и множество $\mathbb{B} \subset \mathbb{R}^q$ ограничено по предположению М2, то множество $\mathbb{M} \subset \mathbb{R}^p$ ограничено. Так как $\mathbb{M} \subset \mathbb{R}^p$ ограничено, оно может быть помещено в гиперкуб S_a с ребром $a > 0$. Пусть $\delta > 0$ — некоторое число. Рассмотрим равномерную сетку $G(\delta)$ для куба с расстояниями между точками вдоль каждого ребра не превосходящим $\frac{\delta}{\sqrt{p}}$. Число точек в сетке не превосходит $\left(\frac{a\sqrt{p}}{\delta}\right)^p$. Куб S_a разбит сеткой $G(\delta)$ на малые кубы с ребрами не превосходящими $\frac{\delta}{\sqrt{p}}$. Поэтому для каждой точки Z из куба S_a расстояние между Z и $G(\delta)$ не превосходит длины диагонали малого, которому она принадлежит:

$$d(Z, G(\delta)) \leq \sqrt{\sum_{k=1}^p \frac{\delta^2}{p}} = \delta,$$

где для точки $Z \in \mathbb{R}^p$ и множества $A \subset \mathbb{R}^p$, $d(Z, A) = \inf_{Z' \in A} \|Z - Z'\|$ — расстояние от точки до множества. Поэтому во множестве $G(\delta)$ для каждой точки $X \in \mathbb{X}$ содержится точка, удаленная от нее не более чем на δ . Но $G(\delta) \not\subset \mathbb{X}$.

Обозначим $B_\delta(X)$ шар с центром X и радиусом δ . Определим $\tilde{G}(\delta)$: Для каждой точки $X \in G(\delta/2)$, если $\tilde{B}_{\delta/2}(X) = B_\delta(X) \cap \mathbb{M} \neq \emptyset$, то выберем точку \tilde{X} из \tilde{B}_δ и поместим ее в $\tilde{G}(\delta)$. Множество $\tilde{G}(\delta)$ является δ -сетью для \mathbb{X} и оно состоит не более чем из $\left(\frac{2a\sqrt{p}}{\delta}\right)^p$ точек. \square

Докажем лемму о выполнении системы событий

Лемма 12. Пусть каждое из событий A_1, \dots, A_M наступает с вероятностью не меньше, чем p . Тогда все они наступают одновременно с вероятностью не меньше $P(\cap_{m=1}^M A_m) \geq 1 - M \cdot (1 - p)$.

Доказательство. Обозначим \bar{A} — дополнение борелевского множества A до

множества элементарных исходов. Преобразуем

$$\begin{aligned} \bigcap_{m=1}^M A_m &\subset \overline{\bigcap_{m=1}^M \overline{A_m}}; \\ P(\bigcap_{m=1}^M A_m) &\geq P(\overline{\bigcap_{m=1}^M \overline{A_m}}) = \\ &= 1 - P(\bigcap_{m=1}^M \overline{A_m}) \geq 1 - \sum_{m=1}^M P(\overline{A_m}) \geq 1 - M \cdot (1 - p). \end{aligned}$$

Лемма доказана. □

В.5 Моменты случайных величин

Лемма 13 (четность главного члена $F(X, X')$). При $d \geq 2$ функция $\varphi(X, \theta, 0)$ — четная по θ и функция $\varphi_1(X, \theta, 0)$ — нечетная по θ . При $d \neq 2$ функция $\varphi(X, \theta, 0)$ — нечетная по θ и функция $\varphi_1(X, \theta, 0)$ — четная по θ .

Доказательство. Заметим, что

$$\begin{aligned} \varphi(X, \theta, 0) &= d! \cdot (\nabla_{\theta})^d F(X, \exp_X(t\theta))|_{t=0}; \\ \varphi_1(X, \theta, 0) &= (d+1)! \cdot (\nabla_{\theta})^{d+1} F(X, \exp_X(t\theta))|_{t=0}. \end{aligned}$$

Порядки по θ доказывают лемму. □

Лемма 14 (четность главного члена $F(X, X')$). Из предположения КЗ следует существование $\rho_{E,m}(X, \theta)$ (35) и $\rho_{V,m}(X, \theta)$ (39) для $m \in \{0\} \cup \mathbb{N}$.

Доказательство. Действительно

$$\begin{aligned} 0 \leq \rho_{E,m}(X, \theta) &= \int_0^1 K(X, \theta, t) t^{m+q-1} dt \leq \\ &\leq \int_0^1 K(X, \theta, t) 1^m \cdot t^{q-1} dt = \rho_{E,0}(X, \theta); \\ 0 \leq \rho_{V,m}(X, \theta) &= \int_0^1 K(X, \theta, t)^2 t^{m+q-1} dt \leq \\ &\leq \int_0^1 K(X, \theta, t)^2 1^m \cdot t^{q-1} dt = \rho_{V,0}(X, \theta). \end{aligned}$$

□

Доказательство леммы 10. 0. По Лемме 14 $\rho_{E,m}(X, \theta)$ и $\rho_{E,m}(X, \theta)$ существуют.

1. По Лемме 8 для $\xi \cdot \varepsilon^q$ при $\varepsilon \leq C_{int}$:

$$\begin{aligned} |\mathbb{E}\xi \cdot \varepsilon^q - \int_{\tilde{B}_\varepsilon(X)} K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \cdot p_\mu(X') \cdot dV(X')| &\leq \\ &\leq 8 \cdot V_q \cdot p_{max} \cdot \sup_{X, X'} \left| K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \right| \cdot \varepsilon^{q+2}. \end{aligned}$$

Следовательно

$$\begin{aligned} \left| \mathbb{E}\xi - \int_{\tilde{B}_\varepsilon(X)} K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^{q+d}} \cdot p_\mu(X') \cdot dV(X') \right| &\leq \\ &\leq 8 \cdot V_q \cdot p_{max} \cdot \sup_{X, X'} \left| K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \right| \cdot \varepsilon^2. \end{aligned}$$

2. Далее перейдем в функции $K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^{d+q}} \cdot p_\mu(X') dV(X')$ к локально римановым координатам (θ, t) : $X' = \exp_X(\varepsilon \cdot t\theta)$, $t \in [0, 1]$ и разложим в окрестности $\varepsilon = 0$. Для этого выразим по формуле Тейлора с остаточным членом в форме Лагранжа $K_\varepsilon(X, X') = K(X, \theta, \frac{|X'-X|}{\varepsilon})$ через $K(X, \theta, t)$, используя Лемму 2:

$$K(X, \theta, \frac{|X' - X|}{\varepsilon}) = K(X, \theta, t) + t^2 \cdot \varepsilon^2 \cdot K'(X, \theta, t_1),$$

где $\tilde{t}_1 \in [0, t]$. Также

$$\begin{aligned} \frac{F(X, X')}{\varepsilon^d} &= t^d \cdot \varphi(X, \theta, 0) + \varepsilon \cdot t^{d+1} \cdot \varphi_1(X, \theta, 0) + \varepsilon^2 \cdot \frac{t^{d+2}}{2} \cdot \varphi_2(X, \theta, t_2); \\ p_\mu(X') &= p_\mu(X) + \varepsilon \cdot t \cdot \nabla_\theta p_\mu(X) + \varepsilon^2 \cdot t^2 \cdot (\nabla_{\tilde{\theta}})^2 p_\mu(\tilde{X}), \end{aligned}$$

где $t_2 \in [0, t]$, $\tilde{X} = \exp_X(\tilde{t}\theta\varepsilon)$, $\tilde{\theta} \in T_X(\mathbb{M})$. Поэтому, учитывая Лемму 1

$$\begin{aligned}
& K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \cdot p_\mu(X') dV(X') = \\
& = (K(X, \theta, t) + t^2 \cdot \varepsilon^2 \cdot K'(X, \theta, t_1)) \cdot \\
& \cdot \left(t^d \cdot \varphi(X, \theta, 0) + \varepsilon \cdot t^{d+1} \cdot \varphi_1(X, \theta, 0) + \varepsilon^2 \cdot \frac{t^{d+2}}{2} \cdot \varphi_2(X, \theta, t_2) \right) \cdot \\
& \cdot \left(p_\mu(X) + \varepsilon \cdot t \cdot \nabla_\theta p_\mu(X) + \varepsilon^2 \cdot t^2 \cdot (\nabla_\theta)^2 p_\mu(\tilde{X}) \right) \cdot \\
& \cdot \left(t^{q-1} + t^{q+1} Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta}) \right) dt d\theta. \tag{70}
\end{aligned}$$

Выпишем первые три члена по порядку ε при $dt d\theta$:

$$\varepsilon^0 : K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X);$$

$$\begin{aligned}
\varepsilon^1 : & K(X, \theta, t) \cdot t^{d+q} \cdot \varphi_1(X, \theta, 0) \cdot p_\mu(X) + \\
& + K(X, \theta, t) \cdot t^{d+q} \cdot \varphi(X, \theta, 0) \cdot \nabla_\theta p_\mu(X);
\end{aligned}$$

$$\begin{aligned}
\varepsilon^2 : & K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi_2(X, \theta, t_2) \cdot p_\mu(X) + \\
& + K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi(X, \theta, 0) \cdot (\nabla_\theta)^2 p_\mu(X) + \\
& + t^{d+q+1} \cdot K'(X, \theta, t_1) \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) + \\
& + t^{d+q+1} \cdot K(X, \theta, t) \cdot \varphi_1(X, \theta, 0) \cdot \nabla_\theta p_\mu(X) + \\
& + K(X, \theta, t) \cdot t^{d+q+1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) \cdot Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta}).
\end{aligned}$$

3. Заметим, что единичный шар по (t, θ) равный $[0, 1] \times S_{q-1}$ является симметричной областью. Поэтому при интегралы от нечетных функций по нему обращаются в нуль по Лемме 13. Так, при $d \not\equiv 2$

$$\int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) dt d\theta = 0,$$

а при $d \equiv 2$ интеграл от членов первого порядка по ε обращается в нуль.

4. Проинтегрируем главный и следующий члены по ε из пункта 2 по t

внутри единичной сферы:

$$\begin{aligned}
& \int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q-1} \cdot \varphi(X, \theta, 0) \cdot p_\mu(X) dt d\theta = \\
& \quad = p_\mu(X) \cdot \int_{S_{q-1}} \rho_{E,d}(X, \theta) \varphi(X, \theta, 0) d\theta; \\
E_1(X) & \equiv \int_{S_{q-1}} \int_0^1 K(X, \theta, t) \cdot t^{d+q} \cdot \varphi_1(X, \theta, 0) \cdot p_\mu(X) + \\
& \quad + K(X, \theta, t) \cdot t^{d+q} \cdot \varphi(X, \theta, 0) \cdot \nabla_{\theta} p_\mu(X) dt d\theta = \\
& \quad = p_\mu(X) \cdot \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi_1(X, \theta, 0) d\theta + \\
& \quad + \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi(X, \theta, 0) \nabla_{\theta} p_\mu(X) d\theta \tag{71}
\end{aligned}$$

5. Ограничим $E_1(X)$:

$$\begin{aligned}
|E_1(X)| & \leq |p_\mu(X) \cdot \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi_1(X, \theta, 0) d\theta| + \\
& + \left| \int_{S_{q-1}} \rho_{E,d+1}(X, \theta) \varphi(X, \theta, 0) \nabla_{\theta} p_\mu(X) d\theta \right| \leq V_q \cdot C_\rho \cdot (p_{max} \cdot C_{\phi,1} + C_{p,1} \cdot C_\phi).
\end{aligned}$$

Обозначим

$$C_{E,1} = V_q \cdot C_\rho \cdot (p_{max} \cdot C_{\phi,1} + C_{p,1} \cdot C_\phi). \tag{72}$$

Ограничим интеграл от членов порядка выше первого по ε . Для воспользуемся (70) и учтем $\varepsilon \leq 1$ и $t \leq 1$:

$$|K(X, \theta, t) + t^2 \cdot K'(X, \theta, t_1)| \leq C_K + C_{K,1};$$

$$\begin{aligned}
& |t^d \cdot \varphi(X, \theta, 0) + t^{d+1} \cdot \varphi_1(X, \theta, 0) + \frac{t^{d+2}}{2} \cdot \varphi_2(X, \theta, t_2)| \leq \\
& \leq |\varphi(X, \theta, 0) + \varphi_1(X, \theta, 0) + \varphi_2(X, \theta, t_2)| \leq C_\varphi + C_{\varphi,1} + C_{\varphi,2};
\end{aligned}$$

$$\begin{aligned}
& |p_\mu(X) + t \cdot \nabla_\theta p_\mu(X) + t^2 \cdot (\nabla_\theta)^2 p_\mu(\tilde{X})| \leq \\
& \leq |p_\mu(X) + \nabla_\theta p_\mu(X) + (\nabla_\theta)^2 p_\mu(\tilde{X})| \leq \\
& \leq p_{max} + C_{p,1} + C_{p,2};
\end{aligned}$$

$$\begin{aligned}
& |1 + t^2 Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta})| \leq \\
& \leq |1 + Ric_{\tilde{X}}(\tilde{\theta}, \tilde{\theta})| \leq 1 + C_{Ric}.
\end{aligned}$$

Учитывая пункт 1 и

$$\begin{aligned}
& 8 \cdot V_q \cdot p_{max} \cdot \sup_{X, X'} \left| K_\varepsilon(X, X') \cdot \frac{F(X, X')}{\varepsilon^d} \right| \leq \\
& \leq 8 \cdot V_q \cdot p_{max} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})
\end{aligned}$$

получаем

$$\begin{aligned}
C_{E,2} = & V_q \cdot (C_K + C_{K,1}) \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2}) \cdot \\
& \cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + \\
& + 8 \cdot V_q \cdot p_{max} \cdot C_K \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})
\end{aligned} \tag{73}$$

6. Повторяя пункты 1-6 для ξ^2 , учитывая четность $2d$ получаем равенство нулю невязок порядка ε^{-q+1} и выражение

$$\mathbb{E}\xi^2 = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta + \varepsilon^{-q+2} \cdot C_{E2,2},$$

где

$$\begin{aligned}
C_{E2,2} = & V_q \cdot (C_K + C_{K,1})^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2 \cdot \\
& \cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) \\
& + 8 \cdot V_q \cdot p_{max} \cdot C_K^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2.
\end{aligned}$$

В завершение, огрубляя оценки по степеням ε и заменяя t на 1:

$$\begin{aligned}
& \text{Var}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \\
& = \varepsilon^{-q} \cdot p_\mu(X) \cdot \int_{S^{q-1}} \rho_{D,2d}(X, \theta) \varphi(X, \theta, 0)^2 d\theta + \varepsilon^{\min\{-q+2,0\}} \cdot C_{V,2},
\end{aligned}$$

где

$$\begin{aligned}
C_{V,2} = & V_q \cdot (C_K + C_{K,1})^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2 \cdot \\
& \cdot (p_{max} + C_{p,1} + C_{p,2}) \cdot (1 + C_{Ric}) + \\
& + 8 \cdot V_q \cdot p_{max} \cdot C_K^2 \cdot (C_\varphi + C_{\varphi,1} + C_{\varphi,2})^2 + C_{E2,2}.
\end{aligned} \tag{74}$$

□

В.6 Случайные величины у границы

Лемма 15 (число точек у границы). Для достаточно малого $\varepsilon \leq \min\{\varepsilon_B, \varepsilon_c\}$, число точек $\mathbb{X}_{N,\varepsilon}$ выборки \mathbb{X}_N , попавших близко к границе $\mathbb{M} \setminus \mathbb{M}_\varepsilon$ с большой вероятностью относительно мало:

$$P(|\mathbb{X}_{N,\varepsilon}| \geq 2 \cdot \varepsilon \cdot N \cdot p_{max} \cdot dV(\mathbb{M}) \cdot C_B) \leq \exp(-\tilde{C}_B \cdot N\varepsilon),$$

где \tilde{C}_B — константа, p_{max} — константа из (12), C_B — константа из M10.

Доказательство. Попадание точки в окрестность границы — Бернуллиевская случайная величина. Вероятность ее успеха стремится к нулю не медленнее $\varepsilon \cdot N \cdot p_{max} \cdot dV(\mathbb{M}) \cdot C_B$ из Предположения M9. Откуда, аналогично Теореме 2 получаем условие Теоремы для некоторого $\tilde{C}_B > 0$. □

Лемма 16 ($F_N(X)$ у границы). Для всех точек многообразия $X \in \mathbb{M}$, для любого $N > N_{0,B}$ и $\varepsilon < \tilde{\varepsilon}_{0,B}$ и некоторых положительных констант $C_{LD1,B}, C_{LDP,B}$ выполнено и для $z \in [0, 1]$:

$$\begin{aligned}
P\left(\sup_{X \in \mathbb{M}} |F_N(X) - \bar{F}(X)| \geq z + \varepsilon \cdot C_{LD1,B}\right) & \leq \\
& \leq \exp(-z^2 \cdot C_{LDP,B} \cdot N\varepsilon^q).
\end{aligned}$$

Доказательство. Аналогично доказательству Теоремы 5 без использования симметрии области интегрирования, с применением M10 и учетом превосходства экспоненциального затухания над полиномиальным ростом. □