

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Гученко Роман Александрович

**ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ
ДЛЯ ДИСКРИМИНАЦИИ
РЕГРЕССИОННЫХ МОДЕЛЕЙ**

Специальность 01.01.07 —
вычислительная математика

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук
профессор Мелас Вячеслав Борисович

Санкт-Петербург — 2017

Оглавление

	Стр.
Введение	5
Глава 1. Постановка задачи планирования дискриминационных экспериментов и обзор известных результатов	14
1.1 Задачи анализа и планирования в контексте восстановления регрессионных зависимостей	14
1.2 Планы эксперимента	16
1.3 Планирование экспериментов для дискриминации регрессионных моделей	18
1.4 Связь задачи дискриминации и задачи по оцениванию параметров	19
1.5 Критерий T -оптимальности и его статистическая интерпретация .	20
1.6 Теорема эквивалентности для критерия T -оптимальности	22
1.7 Связь критерия T -оптимальности с теорией аппроксимации . . .	25
1.8 Алгоритм для численного построения T -оптимальных планов . .	26
1.9 Различные обобщения критерия T -оптимальности	28
Глава 2. T-оптимальные планы для дискриминации дробно-рациональных и полиномиальных моделей. . . .	30
2.1 Простейшая дробь и полином	30
2.2 Другие дробно-рациональные модели	34
2.3 ЕМАХ-модель и квадратичная модель	37
2.3.1 Локально-оптимальный план	37
2.3.2 Робастный план	40
Глава 3. Построение байесовских T_R-оптимальных планов	44
3.1 Различные аналоги критерия T -оптимальности	44
3.2 Сведение байесовских планов к локально оптимальным	47
3.3 Теорема эквивалентности для T_R -оптимальных планов	49
3.4 Численные алгоритмы для построения T_R -оптимальных планов .	51
3.4.1 Квадратичное программирование	56
3.4.2 Градиентный метод	59
3.5 Примеры	60

3.5.1	Байесовские T_P -оптимальные планы для дискриминации экспоненциальных моделей	61
3.5.2	Байесовские T_P -оптимальные планы для дискриминации моделей зависимости эффективности препарата от дозы	65
Глава 4. Построение байесовских KL_P-оптимальных планов		69
4.1	Обобщение критерия T -оптимальности на случай произвольно распределенных ошибок наблюдения	69
4.2	Байесовские KL_P -оптимальные планы и их сведение к локально оптимальным	72
4.3	Теорема эквивалентности для KL_P -оптимальных планов	73
4.4	Численные алгоритмы для построения KL_P -оптимальных планов	75
4.4.1	Квадратичное программирование	76
4.4.2	Градиентный метод	80
4.5	Примеры	81
4.5.1	KL -оптимальные планы для дискриминации ЕМАХ-модели и модели Михаэлиса–Ментен	81
4.5.2	Байесовские KL_P -оптимальные планы для дискриминации экспоненциальных моделей	83
4.5.3	Байесовские KL_P -оптимальные планы для дискриминации моделей зависимости эффективности препарата от дозы	84
Глава 5. Построение полу-параметрических оптимальных планов		90
5.1	Полу-параметрические оптимальные планы	90
5.2	Теоремы эквивалентности для полу-параметрических оптимальных планов	99
5.3	Связь полу-параметрических критериев с критерием T -оптимальности	100
5.4	Примеры	104
5.4.1	Полу-параметрические оптимальные планы для дискриминации ЕМАХ-модели и модели Михаэлиса–Ментен	105
5.4.2	Полу-параметрические оптимальные планы для дискриминации экспоненциальной модели и модели Михаэлиса–Ментен	108

5.4.3 Сравнение подходов	108
Заключение	112
Список литературы	113
Список рисунков	117
Список таблиц	118

Введение

Актуальность темы. Математическая теория планирования эксперимента предоставляет методологию для оптимального выбора условий проведения наблюдений при практических исследованиях, которая позволяет получать результаты, касающиеся числовых характеристик регрессионных моделей для изучаемых явлений, с заданной степенью статистической погрешности при наименьших затратах. Название этому направлению в науке дала книга «The Design of Experiments» знаменитого английского статистика Р. Фишера. Базовый аппарат теории был разработан во второй половине прошлого века в трудах Дж. Элфвинга, Дж. Кифера, Дж. Вольфовица, К. Рао, В. Стаддена, В.В. Федорова, Г. Уинна и других авторов. Целью планирования является нахождение (приближенного) плана эксперимента — дискретной вероятностной меры, заданной на множестве всех возможных условий проведения измерений, оптимальной с точки зрения некоторого заранее заданного критерия.

Большинство работ по оптимальному планированию экспериментов посвящено планам для оценивания неизвестных параметров регрессионных моделей. В этом случае предполагается, что функция регрессии задана с точностью до неизвестных параметров, которые необходимо оценить с некоторой точки зрения оптимально посредством выбора условий для проведения измерений. Примерами оптимальных планов для оценивания параметров являются *D*-оптимальные планы, минимизирующие объем доверительного эллипсоида для оценок неизвестных параметров в предположении о независимости, гомоскедастичности и нормальной распределенности случайных ошибок измерения, или *E*-оптимальные планы, минимизирующие в том же эллипсоиде длину максимальной оси. Однако при проведении прикладных исследований в различных областях знаний нередко возникает ситуация, когда вид регрессионной модели с точностью до параметров не известен а priori до проведения эксперимента, но тем не менее у экспертов есть несколько гипотез о возможном виде модели. В этом случае проводят эксперимент специального вида — дискриминационный эксперимент, планируемый таким образом, чтобы по его результатам можно было оптимально относительно некоторого критерия проверить гипотезу об истинном виде исследуемой регрессионной модели. Самыми популярными критериями, используемыми для решения задач дискриминации, являются критерий

T -оптимальности, введенный в работах Аткинсона и Федорова, и различные его обобщения.

Критерий T -оптимальности для дискриминации двух конкурирующих моделей предполагает наличие априорно заданного фиксированного значения параметров для одной из моделей. Это его свойство называется локальностью. Современная статистическая практика требует нахождения планов, устойчивых относительно некорректного выбора фиксированного вектора параметров, таких как байесовские оптимальные планы, где вместо точечного фиксированного значения параметров берется некоторое распределение. Явный вид опорных точек и весов даже для локального T -оптимального плана получен в литературе только в случае дискриминации полиномиальных моделей, отличающихся на один или на два порядка. Отыскание же байесовских планов в явном виде обычно не представляется возможным. Для численного нахождения дискриминационных планов обычно используют различные варианты алгоритма, предложенного Аткинсоном и Федоровым. В случае байесовских планов применение этого алгоритма становится проблематичным. Данная диссертационная работа посвящена разработке эффективных численных алгоритмов для нахождения дискриминационных планов, а также нахождению T -оптимальных планов в явном виде для некоторых специальных пар регрессионных моделей.

Современное состояние исследований. Первые работы по планированию дискриминационных экспериментов появились в начале 70-х годов прошлого века и были связаны с дискриминацией вложенных моделей, когда одна модель является частным случаем другой при определенных значениях параметров. Так в работе [1] для дискриминации между двумя вложенными полиномиальными моделями в предположении о независимости, нормальности и гомоскедастичности ошибок было предложено искать план, доставляющий минимум объему доверительного эллипсоида для тех параметров более общей модели, которые не входят в менее общую. В рамках этого подхода также написаны следующие работы: [2], где планы для оценивания старших коэффициентов полиномиальной модели были найдены в явном виде; [3], где определяется план, оптимальный для выбора количества слагаемых в модели Фурье; [4], где рассматриваются одновременно задача оценивания степени полинома и задача оценивания его коэффициентов; [5], где рассматривается задача дискриминации для нескольких вложенных нелинейных по параметрам моделей.

Другой критерий оптимальности, T -критерий, применимый для решения задачи о дискриминации двух конкурирующих регрессионных моделей в случае независимых нормально распределенных гомоскедастичных ошибок, не предполагающий вложенности этих моделей, был введен в [6]. В этой же работе был предложен численный алгоритм для нахождения T -оптимальных планов. Критерий T -оптимальности тесно связан с задачей наилучшей чебышёвской аппроксимации. Наиболее полно соответствующие результаты изложены в [7], а в [8] они использованы для получения в явном виде T -оптимальных планов для дискриминации двух полиномиальных моделей, отличающихся на два порядка.

Базовый критерий T -оптимальности имеет ряд существенных ограничений. Одно из ограничений заключается в требовании о нормальности и гомоскедастичности ошибок наблюдения. На случай дискриминации двух моделей при нормальных гетероскедастичных ошибках T -критерий был обобщен в [9]. В [10] был предложен критерий KL -оптимальности, основанный на расстояниях Кульбака–Лейблера и связанный с тестом отношения правдоподобия, не требующий нормальности и пригодный для дискриминации двух произвольных конкурирующих моделей для плотностей ошибок. Критерии из работ [6] и [9] являются частными случаями KL -критерия. В [11] был введен полу-параметрический критерий, в котором функции регрессии для всех конкурирующих моделей и закон распределения ошибок для одной из них считаются известными, а закон распределения для оставшейся модели получается как решение специальной задачи вариационного исчисления. Другое ограничение T -критерия состоит в количестве сравниваемых моделей. Авторы оригинальной работы предложили в [12] вариант его обобщения на случай дискриминации произвольного количества конкурирующих моделей. Симметричная версия T -критерия для дискриминации многих моделей, T_P -критерий, а также численный алгоритм для нахождения оптимальных планов, основанный на многомерной чебышёвской аппроксимации, были введены в [13]. Еще одно ограничение T -критерия — зависимость от априорных значений для параметров одной из моделей, может быть компенсировано с помощью байесовского подхода, обсуждавшегося еще в [6].

Цели и задачи работы. Целью диссертации является исследование различных критериев оптимальности для дискриминации конкурирующих регрессионных моделей, а также разработка эффективных алгоритмов для численно-

го нахождения соответствующих оптимальных планов. Для достижения поставленной цели необходимо было сформулировать и решить следующие задачи:

1. Использовать аппарат теории чебышёвской аппроксимации для нахождения в явном виде T -оптимальных планов в случае дискриминации полиномиальных моделей и моделей, отличающихся от полиномиальных добавлением простого дробно-рационального слагаемого.
2. Предложить метод построения байесовских T_P -оптимальных планов для дискриминации нескольких моделей, преодолевающий недостатки метода Аткинсона и Федорова.
3. Обобщить этот метод на случай байесовских KL_P -оптимальных планов.
4. Исследовать полу-параметрические критерии оптимальности и их связь с классическими критериями, упростить их численное нахождение.

Научная новизна. Все результаты, представленные в диссертационной работе, являются новыми.

Теоретическая и практическая ценность работы. Результаты работы имеют теоретическую ценность и могут быть использованы при планировании реальных экспериментов. В частности, на второй фазе клинических исследований для установления вида зависимости эффекта препарата от дозы, а также для дискриминации моделей аналитической химии, описывающих прошедшую реакцию.

Методология и методы исследования. В работе применяются методы теории аппроксимации, функционального анализа, математической статистики, вариационного исчисления и общие методы теории планирования эксперимента. Численные примеры выполнены с использованием статистического пакета R.

Положения, выносимые на защиту.

1. В явном виде получены T -оптимальные планы для дискриминации полиномиальных моделей и аналогичных моделей, содержащих дополнительное дробно-рациональное слагаемое.
2. Предложен метод построения байесовских T_P -оптимальных планов путем их сведения к локально оптимальным планам. Разработан двухэтапный алгоритм для нахождения локальных T_P -оптимальных планов, состоящий в чередовании обновления носителя плана и оптимизации по его весам. Доказана сходимость этого алгоритма. Для оптимизации по весам предложено две эффективные численные процедуры. Проведено сравнение наиболее ча-

сто используемого в литературе алгоритма с новым, выявившее значительное преимущество последнего.

3. Сформулирован байесовский критерий KL_P -оптимальности и теорема эквивалентности для него. Результаты, касающиеся байесовских T_P -оптимальных планов, обобщены на случай байесовских KL_P -оптимальных планов.
4. Предложен упрощенный метод численного нахождения полу-параметрических оптимальных планов. Доказаны две теоремы, связывающие полу-параметрические критерии с критерием T -оптимальности.

Степень достоверности и апробация результатов. Правильность T -оптимальных планов, выведенных аналитически во второй главе, подтверждается численными результатами. Все планы, полученные численно, были проверены на оптимальность с помощью теорем эквивалентности. Для утверждений, доказанных диссертантом, в работе представлены полные доказательства; для остальных утверждений приведены ссылки на доказательства.

Основные результаты диссертационной работы докладывались на семинаре кафедры статистического моделирования математико-механического факультета СПбГУ. Исследования по теме диссертации были частично поддержаны грантами СПбГУ (проект 6.38.435.2015) и РФФИ (проект 17-01-00161).

Публикации на тему диссертации. Основные результаты диссертации изложены в трех работах [14], [15], [16], которые опубликованы в журналах, индексируемых системой Scopus.

Вклад диссертанта в совместные работы. Вторая глава основана на работе [14], в которой все результаты и основной текст принадлежат диссертанту. Постановка задачи и текст вводной секции принадлежит соавтору. Третья глава основана на работе [15], в которой метод сведения байесовских планов к локально оптимальным, алгоритм 3.2, секция про метод оптимизации по весам, основанный на квадратичном программировании, теорема 3.3 и все численные результаты принадлежат диссертанту. Постановка задачи, итоговый английский текст и остальные теоремы принадлежат соавторам. Четвертая глава основана на работе [16], в которой адаптация алгоритма из предыдущей работы на случай KL_P -оптимальных планов и все численные результаты принадлежат диссертанту. Постановка задачи, итоговый английский текст и все остальные результаты принадлежат соавторам. Пятая глава основана на работе, выполненной во время визита диссертанта в Рурский университет Бохума

совместно с Х. Детте, В.Б. Меласом и В.К. Вонгом. Теоремы 15, 17 и 18, лемма 4, а также все численные результаты принадлежат диссертанту.

Структура работы. Диссертация состоит из введения, пяти глав и заключения.

Во введении аргументирована актуальность темы диссертации, приведен обзор соответствующей литературы, определены цели и задачи работы, обоснована их научная ценность.

В главе 1 сформулирована задача планирования эксперимента для дискриминации регрессионных моделей, описано ее место в математической статистике и в теории планирования эксперимента, а также дан обзор имеющихся в литературе результатов по данной теме. В § 1.1 приведены постановки задач анализа эмпирических данных и оптимального планирования экспериментов в контексте восстановления регрессионных зависимостей, описана разница между этими двумя постановками. В § 1.2 приведены понятия точного и приближенного планов эксперимента, описано соотношение между ними. Задача планирования эксперимента для дискриминации конкурирующих регрессионных моделей в общем виде сформулирована в § 1.3. В § 1.4 описано то, как связаны планы эксперимента для оценивания неизвестных параметров одной регрессионной модели и планы эксперимента для дискриминации нескольких регрессионных моделей в случае вложенных конкурирующих регрессионных моделей. Наконец, в § 1.5 приведен критерий T -оптимальности для планов дискриминации (предложенный в работах [6; 12]), исследованию различных обобщений которого посвящена содержательная часть работы. В том же параграфе дана статистическая интерпретация T -критерия для случая точных планов эксперимента. В § 1.6 для T -критерия приведена соответствующая теорема эквивалентности, дающая необходимые и достаточные условия для оптимальности плана. В сущности, все последующие теоремы эквивалентности для обобщений T -критерия, которые будут рассматриваться в данной работе, имеют схожую структуру и доказываются аналогично теореме из этого параграфа. В § 1.7 приведены теоремы, связывающие задачу нахождения T -оптимального плана с задачей наилучшей чебышёвской аппроксимации. Используя эти теоремы в главе 2 мы получим в явном виде T -оптимальные планы для дискриминации полиномиальных моделей и моделей, отличающихся от полиномиальных добавлением дробно-рационального слагаемого. В § 1.8 описан самый распространенный

численный алгоритм для нахождения T -оптимальных планов. Альтернативный этому новый численный алгоритм для нахождения байесовских T -оптимальных планов описан в главе 3. В § 1.9 дан обзор критериев, обобщающих критерий T -оптимальности.

В главе 2 в явном виде получены T -оптимальные планы для дискриминации полиномиальных моделей и моделей, отличающихся от полиномиальных добавлением дробно-рационального слагаемого. Нахождение аналитических представлений для точек и весов T -оптимальных планов является трудной минимаксной задачей, поэтому для получения T -оптимальных планов обычно используют численные алгоритмы. До недавнего времени аналитическое решение было известно только для случая дискриминации двух полиномиальных моделей, отличающихся на один порядок. Известно (смотри, например, [7]), что задача нахождения T -оптимального плана тесно связана с задачей наилучшей чебышёвской аппроксимации в том смысле, что опорные точки оптимального плана являются точками максимума функции, представляющей из себя квадрат разности между регрессионной моделью из нулевой гипотезы и ее наилучшей чебышёвской аппроксимации альтернативной моделью. На основе этого факта в работе [8] были аналитически получены T -оптимальные планы для дискриминации полиномиальных моделей, отличающихся на два порядка. В § 2.1 с использованием известных фактов из теории аппроксимации доказано, что опорные точки T -оптимального плана для дискриминации простейшей дробно-рациональной модели и полиномиальной модели являются корнями полинома специального вида. Корни этого полинома, а значит и опорные точки соответствующего плана, могут быть выписаны для небольших степеней полиномиальной модели в явном виде. В § 2.2 аналогичный результат доказан для двух других дробно-рациональных моделей. В § 2.3 при помощи теоремы из § 2.1 аналитически найден T -оптимальный план для дискриминации ЕМАХ-модели и квадратичной модели, а также численно построен план, оптимальный с точки зрения байесовского стандартизированного критерия T -оптимальности. Изложение в главе 2 опирается на материал из работы [14].

В главе 3 предложен метод построения байесовских T_P -оптимальных планов. У стандартного критерия T -оптимальности есть ряд существенных недостатков: он предназначен для сравнения только двух моделей и он является локальным, то есть зависит от фиксированного априорного значения парамет-

ров одной из моделей. В § 3.1 перечислены различные обобщения критерия T -оптимальности, призванные устранить эти недостатки. Для дискриминации нескольких моделей в работе [13] был предложен критерий T_P -оптимальности, представляющий из себя взвешенную сумму из попарных квадратов разностей между конкурирующими моделями, где параметры одной из моделей в паре фиксированы, а параметры другой подбираются по методу наименьших квадратов. Стандартным же способом компенсации локальности любого критерия является введение его байесовской версии, где вместо фиксированного значения параметров используется фиксированное априорное распределение. В § 3.2 вводится байесовский критерий T_P -оптимальности и показывается, что он сводится к обычному критерию T_P -оптимальности с большим числом попарных сравнений в случае, когда априорные распределения для параметров дискретны. В § 3.3 сформулирована теорема эквивалентности для T_P -критерия. В § 3.4 предложен новый двухэтапный численный метод для нахождения T_P -оптимальных планов, суть которого состоит в чередовании обновления носителя промежуточного плана и оптимизации по его весам. Далее доказывается сходимость этого алгоритма, а также рассматриваются две процедуры для выполнения эффективной оптимизации по весам промежуточного плана. В § 3.5 приведены примеры полученных численно байесовских T_P -оптимальных планов с дискретными априорными распределениями. Изложение в главе 3 опирается на материал из работы [15]. Алгоритмы из § 3.4 реализованы в качестве пакета [17] для языка R.

В главе 4 все результаты, полученные для байесовских T_P -оптимальных планов, обобщены на случай байесовских KL_P -оптимальных планов. Критерий T -оптимальности предполагает гомоскедастичность и нормальную распределенность случайных ошибок наблюдения. В [10] был предложен критерий KL -оптимальности, основанный на расстояниях Кульбака–Лейблера, не требующий нормальности и пригодный для дискриминации двух произвольных конкурирующих моделей для плотностей ошибок. Этот критерий и его статистическая интерпретация описаны в § 4.1. Байесовский критерий KL_P -оптимальности вводится в § 4.2 по аналогии с байесовским критерием T_P -оптимальности. Соответствующая теорема эквивалентности сформулирована в § 4.3. Численные алгоритмы из главы 3 обобщены на случай нового критерия и описаны в § 4.4. Заключительный § 4.5 посвящен численным примерам. Изложение в данной

главе опирается на материал из работы [16]. Алгоритмы, описанные в § 4.4, также реализованы в пакете [17] для языка R.

Глава 5 посвящена так называемым полу-параметрическим KL -оптимальным планам. Рассматривается только случай дискриминации двух конкурирующих регрессионных моделей. В работах по планированию дискриминационных экспериментов, как правило, исследуется ситуация, когда конкурирующие модели, определяемые парой из функции регрессии и закона распределения для случайных ошибок, известны с точностью до параметров. В работе [11] был предложен полу-параметрический критерий оптимальности для планирования дискриминационных экспериментов, обобщающий критерий KL -оптимальности из [10], суть которого заключается в том, что функции регрессии для всех конкурирующих моделей и закон распределения ошибок для одной из них считаются известными, а закон распределения для оставшейся модели получается как решение специальной задачи вариационного исчисления с ограничениями. В § 5.1 рассмотрены два варианта полу-параметрического критерия из [11], которые в данной работе обозначаются как $SKL_{(a)}$ - и $SKL_{(b)}$ -критерии, а также доказана теорема, упрощающая нахождение планов, оптимальных с точки зрения этих критериев. Упрощение заключается в том, что предложенный в теореме подход требует решения одного нелинейного уравнения на заданном интервале для вычисления расстояния Кульбака-Лейблера в критерии оптимальности вместо решения системы из двух нелинейных уравнений. В § 5.2 формулируется теорема эквивалентности для $SKL_{(a)}$ - и $SKL_{(b)}$ -критериев. В § 5.3 доказаны две теоремы, связывающие критерии $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальности с критерием T -оптимальности. Первая теорема утверждает, что $SKL_{(a)}$ -оптимальные планы совпадают с T -оптимальными планами в случае, когда плотность случайных ошибок первой модели является усеченной и симметричной относительно условного среднего. Вторая теорема утверждает, что $SKL_{(b)}$ -оптимальные планы совпадают с T -оптимальными планами, если ошибки второй модели распределены нормально. Наконец, в § 5.4 приводятся численные результаты.

В заключении сформулированы основные результаты работы.

Глава 1. Постановка задачи планирования дискриминационных экспериментов и обзор известных результатов

Первая глава является обзорной и имеет следующую структуру. Вначале обсуждается разница между задачами анализа и планирования в контексте восстановления регрессионных зависимостей. Затем формулируются основные положения теории оптимального планирования эксперимента и задача дискриминации регрессионных моделей, описывается связь между задачей дискриминации и задачей оценивания неизвестных параметров в случае вложенных регрессионных моделей. Далее приводятся критерий T -оптимальности для планов дискриминации, его статистическая интерпретация, теорема эквивалентности, теорема о связи с задачей чебышёвской аппроксимации, численный алгоритм построения T -оптимальных планов, а также различные обобщения критерия T -оптимальности, изучению которых посвящены последующие главы работы.

1.1 Задачи анализа и планирования в контексте восстановления регрессионных зависимостей

Пусть зависимость результатов измерений $y_1, \dots, y_N \in \mathbb{R}$ от условий их проведения $x_1, \dots, x_N$ описывается следующим уравнением:

$$y_i = \eta(x_i, \theta) + \varepsilon_i, \quad x_i \in \mathcal{X}, \theta \in \Theta, \quad i = 1, \dots, N, \quad (1.1)$$

где N — общее число проводимых измерений, $\eta(x, \theta)$ — вещественнозначная функция, $\theta = (\theta_1, \dots, \theta_d)^T$ — ее неизвестные параметры, $\varepsilon_1, \dots, \varepsilon_N$ — некоррелированные случайные ошибки с нулевым математическим ожиданием и заданной дисперсией. От требования о некоррелированности случайных ошибок можно отказаться (смотри, например, [18]), но в рамках данного текста этот случай рассматриваться не будет.

Функцию $\eta(x, \theta)$ принято называть регрессионной моделью, а уравнение (1.1) — уравнением регрессии (смотри, например, книгу [19] с.192). Регрессионная модель называется линейной (по параметрам), если она представима в виде

$$\eta(x, \theta) = \theta^T f(x), \quad f(x) = [f_1(x), \dots, f_d(x)]^T, \quad (1.2)$$

где $f_1(x), \dots, f_d(x)$ есть набор вещественнозначных линейно независимых функций. В противном случае модель называется нелинейной.

В этом параграфе рассматриваются две задачи, связанные с уравнением (1.1): задача анализа эмпирических данных и задача оптимального планирования эксперимента. Постановки этих задач связаны с двумя соответствующими типами прикладных исследований: наблюдательными исследованиями (observational studies), где сам исследователь никак не вмешивается в процесс сбора данных и работает с уже готовой выборкой, и экспериментальными исследованиями (experimental studies), где исследователь может непосредственно влиять на процесс сбора данных посредством изменения условий проведения эксперимента.

Задача анализа эмпирических данных в контексте уравнения (1.1) состоит в построении функции регрессии по имеющейся выборке

$$(x_1, y_1), \dots, (x_N, y_N)$$

из совместного распределения x и y , то есть при определенных предположениях о природе случайных ошибок (например, что $\varepsilon_1, \dots, \varepsilon_N$ являются нормальными случайными величинами с постоянной дисперсией) необходимо найти такую функцию $\eta(x, \theta)$ (и оценить при этом ее неизвестные параметры), которая приближала бы имеющиеся выборочные данные достаточно хорошо относительно какого-нибудь заранее заданного критерия качества $\mathcal{R}(\theta)$, например, суммы квадратов разностей реальных наблюдений и предсказаний

$$\mathcal{R}(\theta) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}[y_i, \eta(x_i, \theta)] = \frac{1}{N} \sum_{i=1}^N [y_i - \eta(x_i, \theta)]^2,$$

имея при этом также хорошую обобщающую способность (грубо говоря, функция, которую мы ищем, не должна быть слишком сложной). Более подробное описание этой задачи можно найти, например, в соответствующих главах книг [20] и [21].

Задача оптимального планирования эксперимента в контексте уравнения (1.1) формулируется следующим образом. Пусть выбор точек $x_1, \dots, x_N$, которые соответствуют различным условиям для проведения экспериментальных измерений, находится в руках исследователя-экспериментатора, заинтересован-

ного в оценивании функциональной зависимости $\eta(x, \theta)$ между x и y , которая предполагается известной а priori с точностью до параметров θ . Задача планирования эксперимента состоит в том, чтобы с точки зрения некоторого заранее заданного критерия оптимальности рационально выбрать условия для проведения измерений, то есть точки $x_1, \dots, x_N$. Так, например, объем доверительного эллипсоида для оценок неизвестных параметров θ по методу наименьших квадратов в случае линейной по параметрам регрессионной модели вида (1.2) при нормальных ошибках с одинаковой дисперсией является монотонно возрастающей функцией от величины

$$\Phi(x_1, \dots, x_N) = \det \left\{ \sum_{i=1}^N f(x_i) f^T(x_i) \right\}^{-1},$$

поэтому естественно выбрать $x_1, \dots, x_N$ таким образом, чтобы $\Phi(x_1, \dots, x_N)$ была наименьшей (смотри, например, книгу [22], раздел 2.4, с.51). Это пример задачи планирования эксперимента для оценивания неизвестных параметров регрессионной модели. Отметим, что y в уравнении (1.1) при такой постановке задачи является случайной величиной, а x детерминирован. Настоящая работа посвящена планированию экспериментов для дискриминации регрессионных моделей, о которых речь пойдет в последующих параграфах. При рациональном выборе условий для проведения измерений можно получить результаты заданной степени точности с помощью наименьшего количества экспериментов, что позволяет экономить ресурсы. В случае дорогостоящих экспериментов выгода может быть весьма существенной, поэтому оптимальное планирование экспериментов представляет существенный практический интерес.

1.2 Планы эксперимента

В рамках задач оптимального планирования эксперимента множество $\mathcal{X}$ обычно называют множеством планирования, а любой конечный набор $x_1, \dots, x_N$ элементов $\mathcal{X}$ — планом эксперимента (смотри [23], с.45). В этом параграфе формулируются два стандартных понятия: точного плана эксперимента и приближенного плана эксперимента, которые понадобятся нам в дальнейшем. Предполагается, что в уравнении (1.1) точки $x_1, \dots, x_N$ не обязательно должны

быть различными. Наряду с уравнением (1.1) рассмотрим уравнение:

$$y_{i,k} = \eta(x_i, \theta) + \varepsilon_{i,k}, \quad i = 1, \dots, n, \quad k = 1, \dots, r_i, \quad (1.3)$$

в котором $x_i \neq x_j$ при $i \neq j$, а $n, r_i \in \mathbb{N}$. Дискретную вероятностную меру ξ_N , задаваемую таблицей

$$\xi_N = \begin{bmatrix} x_1 & \dots & x_n \\ q_1 & \dots & q_n \end{bmatrix}, \quad x_i \in \mathcal{X}, \quad q_i = \frac{r_i}{N}, \quad \sum_{i=1}^n r_i = N, \quad (1.4)$$

принято называть (смотри, например, книгу [23], с. 119) точным планом эксперимента (exact experimental design). Точки $x_1, \dots, x_n$ называются опорными точками плана, их совокупность — носителем плана, а $q_1, \dots, q_n$ — весами. Носитель плана ξ_N будем далее обозначать как $\text{supp}(\xi_N)$. Точный план эксперимента определяет, при каких условиях (то есть в каких точках $x_1, \dots, x_n$ из множества планирования $\mathcal{X}$) и в каком количестве (определяемом числами $r_1, \dots, r_n$) экспериментатору необходимо совершить измерения при условии, что общее число доступных измерений равно N . Тут мы предполагаем, что измерения при различных условиях, то есть в различных точках множества $\mathcal{X}$, имеют одинаковую цену. В сущности, мы всегда хотели бы видеть точный план в качестве конечного результата планирования эксперимента, так как точный план реализуем на практике.

Наряду с точными планами эксперимента, Кифером в работе [24] было введено в употребление более общее понятие приближенного плана, где рациональные веса заменяются на положительные вещественные. Дискретную вероятностную меру ξ , задаваемую таблицей

$$\xi = \begin{bmatrix} x_1 & \dots & x_n \\ \omega_1 & \dots & \omega_n \end{bmatrix}, \quad x_i \in \mathcal{X}, \quad \omega_i \geq 0, \quad \sum_{i=1}^n \omega_i = 1, \quad (1.5)$$

принято называть приближенным планом эксперимента (approximate experimental design). Приближенный план удобнее тем, что в отличие от точного он не зависит от числа доступных наблюдений N . Также в классе приближенных планов, как мы впоследствии увидим, задача оптимального планирования эксперимента в некоторых частных случаях может быть решена аналитически. Для тех же случаев, когда аналитическое решение не удастся получить, существуют

достаточно простые численные процедуры, позволяющие найти приближенный план. В отличие от точного плана, приближенный план нельзя реализовать на практике, но имея приближенный план можно получить точный план положив r_i равными $N\omega_i$, округленными предварительно так, чтобы $\sum_{i=1}^n r_i = N$. Более подробное обсуждение того, как округление при переходе от приближенного плана к точному влияет на качество плана, можно найти в главе 12 книги [25].

Отметим, что точные планы также иногда называют дискретными, а приближенные планы — непрерывными, ввиду рациональной/вещественной природы весов (смотри книгу [26], с.53).

1.3 Планирование экспериментов для дискриминации регрессионных моделей

При проведении прикладных исследований в различных областях знаний часто возникает ситуация, когда вид регрессионной модели, описывающей связь между зависимыми и независимыми переменными, не известен а priori до проведения эксперимента, но тем не менее у экспертов имеются несколько гипотез о возможном виде этой модели. Также возможна ситуация, когда данные, полученные в ходе предыдущих исследований, хорошо аппроксимируются сразу несколькими различными регрессионными моделями. В таких случаях проводят эксперимент специального вида (дискриминационный эксперимент), планируемый таким образом, чтобы по его результатам можно было выбрать одну модель среди заданных, наилучшим образом описывающую изучаемое явление.

Для простоты сначала разберем случай, когда конкурирующих моделей рассматривается всего две:

$$\eta_1(x, \theta_1), \quad x \in \mathcal{X}, \theta_1 = (\theta_{1,1}, \dots, \theta_{1,d_1})^T \in \Theta_1; \quad (1.6)$$

$$\eta_2(x, \theta_2), \quad x \in \mathcal{X}, \theta_2 = (\theta_{2,1}, \dots, \theta_{2,d_2})^T \in \Theta_2. \quad (1.7)$$

Также мы будем считать, что одна из этих моделей, которую мы далее будем называть истинной, удовлетворяет в точности уравнению (1.3), то есть $\eta(x, \theta) = \eta_1(x, \theta_1)$ или $\eta(x, \theta) = \eta_2(x, \theta_2)$ при каких-то значениях неизвестных параметров.

1.4 Связь задачи дискриминации и задачи по оцениванию параметров

В первую очередь обсудим один из возможных подходов к планированию дискриминационных экспериментов в случае так называемых вложенных регрессионных моделей. Мы будем говорить, что модель $\eta_2(x, \theta_2)$ вкладывается в модель $\eta_1(x, \theta_1)$, если

1. Любой вектор $\theta_1 \in \Theta_1$ представим в виде

$$\theta_1 = \begin{pmatrix} \theta_2 \\ \lambda \end{pmatrix}, \quad \theta_2 \in \Theta_2, \lambda \in \Lambda,$$

где Λ имеет размерность $d_1 - d_2$ (мы предполагаем, что $d_1 > d_2$), то есть $\Theta_1 = \Theta_2 \times \Lambda$.

2. Существует такой $\lambda_0 \in \Lambda$, что

$$\eta_1 \left[x, \begin{pmatrix} \theta_2 \\ \lambda_0 \end{pmatrix} \right] = \eta_2(x, \theta_2)$$

для любого $\theta_2 \in \Theta_2$.

В случае вложенных регрессионных моделей задачу дискриминации $\eta_1(x, \theta_1)$ и $\eta_2(x, \theta_2)$ можно свести к задаче проверки статистической гипотезы

$$H_0 : \lambda = \lambda_0, \text{ то есть } \eta(x, \theta) = \eta_2(x, \theta_2); \quad (1.8)$$

против альтернативы

$$H_1 : \lambda \neq \lambda_0, \text{ то есть } \eta(x, \theta) = \eta_1(x, \theta_1). \quad (1.9)$$

Первые работы, посвященные планированию дискриминационных экспериментов для вложенных моделей, появились еще в 70-х годах прошлого века. Так в [1] для дискриминации между двумя вложенными полиномиальными моделями было предложено искать план, доставляющий минимум объему доверительного эллипсоида для тех параметров более общей модели, которые не входят в менее общую. В рамках этого подхода также написаны следующие работы: [2], где планы для оценивания старших коэффициентов полиномиальной модели были найдены в явном виде; [3], где определяется план, оптимальный

для выбора количества слагаемых в модели Фурье; [4], где рассматриваются одновременно задача оценивания степени полинома и задача оценивания его коэффициентов; [5], где рассматривается задача дискриминации для нескольких вложенных нелинейных по параметрам моделей.

1.5 Критерий T -оптимальности и его статистическая интерпретация

Другой подход к решению задачи дискриминации, который не предполагает вложенности конкурирующих регрессионных моделей, был предложен Аткинсоном и Федоровым в работах [6; 12]. Этот подход и его обобщения приобрели достаточно широкую популярность.

Мы будем рассматривать, как и в предыдущем параграфе, случай двух конкурирующих моделей $\eta_1(x, \theta_1)$ и $\eta_2(x, \theta_2)$. Пусть случайные ошибки

$$\varepsilon_{i,k}, \quad i = 1, \dots, n; \quad k = 1, \dots, r_i$$

в уравнении (1.3) имеют нормальное распределение с постоянной дисперсией σ^2 . Не умаляя общности можно считать, что $\sigma^2 = 1$. Будем также полагать, что имеются аппроксимации $\bar{\theta}_1$ и $\bar{\theta}_2$ для неизвестных параметров моделей θ_1 и θ_2 , которые были получены, например, из другого эксперимента.

Тогда при фиксированном точном плане ξ_N , который задается таблицей (1.4), случайная величина

$$h = \sum_{i=1}^n \sum_{k=1}^{r_i} [y_{i,k} - \eta_2(x_i, \bar{\theta}_2)]^2$$

будет иметь (как сумма квадратов нормально распределенных случайных величин с соответствующими средними) нецентральное χ^2 -распределение с N степенями свободы и параметром нецентральности

$$\Delta(\xi_N) = N \sum_{i=1}^n q_i [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \bar{\theta}_2)]^2,$$

если верна гипотеза

$$H_0 : \eta(x, \theta) = \eta_1(x, \bar{\theta}_1).$$

Если же верна альтернативная гипотеза

$$H_1 : \eta(x, \theta) = \eta_2(x, \bar{\theta}_2),$$

то h имеет обычное χ^2 -распределение с N степенями свободы.

Мощность χ^2 -теста в задаче проверки гипотезы H_0 против альтернативы H_1 прямо пропорциональна величине параметра нецентральности $\Delta(\xi_N)$. Очевидно, что в такой постановке любой план, сосредоточенный на множестве точек

$$\left\{ x^* : [\eta_1(x^*, \bar{\theta}_1) - \eta_2(x^*, \bar{\theta}_2)]^2 = \sup_{x \in \mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \bar{\theta}_2)]^2 \right\},$$

будет доставлять максимум величине $\Delta(\xi_N)$, а значит и мощности соответствующего теста.

В сущности, основная проблема при таком подходе состоит в том, что априорные значения $\bar{\theta}_1$ и $\bar{\theta}_2$ не всегда известны. Один из способов ослабления априорных предположений таков: пусть имеется только $\bar{\theta}_1$. При фиксированном плане ξ_N и фиксированном априорном значении $\bar{\theta}_1$ рассмотрим совокупность всех возможных альтернатив

$$H_1 : \eta(x, \theta) = \eta_2(x, \theta_2)$$

при различных $\theta_2 \in \Theta_2$ и выберем среди них ту, которая имеет наименьший параметр нецентральности, то есть

$$\Delta(\xi_N) = N \sum_{i=1}^n q_i [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \hat{\theta}_2)]^2, \quad (1.10)$$

где оценка $\hat{\theta}_2$ вектора неизвестных параметров второй модели θ_2 получается из уравнения

$$\sum_{i=1}^n q_i [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \hat{\theta}_2)]^2 = \arg \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^n q_i [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \theta_2)]^2. \quad (1.11)$$

Точный план ξ_N , который доставляет максимум $\Delta(\xi_N)$, будем называть T -оптимальным. Таким образом, T -оптимальный план максимизирует мощность

χ^2 -теста (или аналогичного F -теста, если дисперсия σ^2 неизвестна) для наиболее близкой альтернативы.

Задачу поиска точного плана ξ_N , доставляющего максимум $\Delta(\xi_N)$, следуя подходу Дж. Кифера, заменим на задачу поиска приближенного плана, то есть произвольной дискретной вероятностной меры ξ , доставляющей максимум функционалу

$$T_{1,2}(\xi) = \int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \hat{\theta}_2)]^2 \xi(dx), \quad (1.12)$$

где $\hat{\theta}_2 = \hat{\theta}_2(\xi)$ получаются из уравнения

$$\int_{\mathcal{X}} [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \hat{\theta}_2)]^2 \xi(dx) = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} [\eta_1(x_i, \bar{\theta}_1) - \eta_2(x_i, \theta_2)]^2 \xi(dx). \quad (1.13)$$

Решение этой оптимизационной задачи, которое будет иметь вид (1.5), мы также будем называть T -оптимальным планом. Множество приближенных планов является выпуклым, а функционал $T_{1,2}(\xi)$ — вогнутым, то есть

$$T_{1,2}([1 - \alpha]\xi_1 + \alpha\xi_2) \geq [1 - \alpha]T_{1,2}(\xi_1) + \alpha T_{1,2}(\xi_2)$$

поэтому любой локальный максимум $T_{1,2}(\xi)$ также будет и глобальным максимумом.

Так как решение оптимизационной задачи зависит от фиксированного заранее значения $\bar{\theta}_1$ параметров первой модели, T -критерий оптимальности является локальным критерием по Чернову [27]. Иногда, чтобы подчеркнуть эту зависимость, мы будем писать $T_{1,2}(\xi, \bar{\theta}_1)$ вместо $T_{1,2}(\xi)$. В данной работе мы не будем останавливаться на том, откуда на самом деле берутся априорные значения $\bar{\theta}_1$, но мы позднее обсудим более робастные аналоги T -критерия, позволяющие сгладить зависимость от априорных параметров. Отметим также, что в общем случае T -критерий (1.12) не является симметричным в том смысле, что поменяв местами η_1 и η_2 мы получим другой критерий.

1.6 Теорема эквивалентности для критерия T -оптимальности

Основным средством исследования в математической теории планирования экспериментов являются так называемые теоремы эквивалентности, предо-

ставляющие необходимые и достаточные условия для оптимальности планов. В этом параграфе мы сформулируем теорему эквивалентности для T -критерия (смотри работу [6]).

Пусть имеют место

Предположение 1. Множество планирования $\mathcal{X}$ компактно; вдобавок, функции $\eta_1(x, \theta_1)$, $\eta_2(x, \theta_2)$ непрерывны по x на $\mathcal{X}$.

Предположение 2. Множества Θ_1 и Θ_2 компактны. Функции $\eta_1(x, \theta_1)$, $\eta_2(x, \theta_2)$ дифференцируемы по θ_1 , θ_2 на Θ_1 , Θ_2 соответственно.

Предположение 3. Если ξ^* — T -оптимальный план, то существует единственный вектор $\theta_2^* = \widehat{\theta}_2(\xi^*)$, удовлетворяющий условию (1.13).

Введем функцию

$$\psi_{1,2}(x, \xi) = [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \widehat{\theta}_2)]^2, \quad (1.14)$$

где $\widehat{\theta}_2 = \widehat{\theta}_2(\xi)$ есть решение уравнения (1.13).

Теорема 1. Пусть выполнены предположения 1, 2 и 3. Тогда

- (1) План ξ^* является T -оптимальным тогда и только тогда, когда для любой точки $x \in \mathcal{X}$ выполнено неравенство

$$\psi_{1,2}(x, \xi^*) \leq T_{1,2}(\xi^*), \quad (1.15)$$

причем равенство в (1.15) достигается только в опорных точках оптимального плана.

- (2) Если план ξ не является T -оптимальным, то всегда найдется такая точка $\dot{x} \in \mathcal{X}$, что выполнено обратное по отношению к (1.15) неравенство

$$\psi_{1,2}(\dot{x}, \xi) > T_{1,2}(\xi^*). \quad (1.16)$$

Теорема 1 имеет важное значение для дальнейшего изложения, поэтому остановимся на ней подробнее. Обозначим пространство вероятностных мер, заданных на $\mathcal{X}$, через $\mathcal{P}(\mathcal{X})$. Пусть $\xi, \zeta \in \mathcal{P}(\mathcal{X})$ и F — это функционал, заданный

на $\mathcal{P}(\mathcal{X})$. Напомним, что производной по Гато функционала F в точке ξ по направлению ζ называется функционал

$$D_\zeta F(\xi) = \lim_{\alpha \rightarrow 0} \frac{F([1 - \alpha]\xi + \alpha\zeta) - F(\xi)}{\alpha}, \quad (1.17)$$

а функцией влияния для функционала F называется функция

$$I(x) = D_{\delta_x} F(\xi) = \lim_{\alpha \rightarrow 0} \frac{F([1 - \alpha]\xi + \alpha\delta_x) - F(\xi)}{\alpha}, \quad (1.18)$$

где δ_x — это вероятностная мера, сосредоточенная целиком в точке $x \in \mathcal{X}$. Если $F(\xi)$ является линейным функционалом, то есть

$$F(\xi) = \int_{\mathcal{X}} a(x)\xi(dx), \quad (1.19)$$

то функция влияния принимает следующий вид

$$I(x) = a(x) - F(\xi). \quad (1.20)$$

В сущности, функционал

$$T_{1,2}(\xi) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi(dx)$$

почти линейный, за исключением наличия $\inf_{\theta_2 \in \Theta_2}$. Справиться с этим дополнительным затруднением помогает следующий результат:

Лемма 1. Пусть $F(\xi, \theta) = \int_{\mathcal{X}} a(x, \theta)\xi(dx)$, причем функция $a(x, \theta)$ непрерывна по x и дифференцируема по θ . Тогда

$$D_\zeta \inf_{\theta \in \Theta} F(\xi, \theta) = \inf_{\theta \in \Theta^*(\xi)} D_\zeta F(\xi, \theta),$$

где

$$\Theta^*(\xi) = \left\{ \theta^* : \theta^* \in \Theta, F(\xi, \theta^*) = \inf_{\theta \in \Theta} F(\xi, \theta) \right\}$$

Доказательство леммы можно найти в книге [28] на странице 75. Положим $a(x, \theta) = [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2$. Эта функция удовлетворяет условиям леммы 1,

так как выполнены предположения **1** и **2**. Пусть ξ^* — T -оптимальный план. Тогда по лемме

$$D_{\zeta}T_{1,2}(\xi^*) = \inf_{\theta_2 \in \Theta_2^*(\xi^*)} \left(\int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \zeta(dx) - \int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi(dx) \right).$$

По предположению **3** множество $\Theta_2^*(\xi^*)$ состоит из одной точки θ_2^* , поэтому функция влияния в точке ξ^* имеет вид

$$I_{T_{1,2}(\xi^*)}(x) = [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)]^2 - T_{1,2}(\xi^*) = \psi_{1,2}(x, \xi^*) - T_{1,2}(\xi^*).$$

Если ξ^* — T -оптимальный план, то для любого односточечного направления x

$$I_{T_{1,2}(\xi^*)}(x) \leq 0;$$

в противном случае найдется $\dot{x} \in \mathcal{X}$, для которого выполнено противоположное неравенство.

Если отказаться от предположения **3**, то теорема эквивалентности будет иметь несколько более сложный вид. Обозначим как Θ_2^* множество всех решений уравнения (1.13) при $\xi^* = \xi$. План ξ^* является T -оптимальным тогда и только тогда, когда существует мера $\mu(\theta_2)$, такая что

$$\Psi_{1,2}(x, \xi^*) \leq T_{1,2}(\xi^*), \quad (1.21)$$

где

$$\Psi_{1,2}(x, \xi) = \int_{\Theta_2^*} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \hat{\theta}_2)]^2 \mu(d\theta_2), \quad \int_{\Theta_2^*} \mu(d\theta_2) = 1. \quad (1.22)$$

1.7 Связь критерия T -оптимальности с теорией аппроксимации

Задача нахождения T -оптимальных планов тесно связана с задачей чебышевской аппроксимации. В частности, верна следующая

Теорема 2. *Задача нахождения T -оптимального плана для дискриминации моделей $\eta_1(x, \bar{\theta}_1)$ и $\eta_2(x, \theta_2)$, $\theta_2 \in \Theta_2$ эквивалентна задаче о нахождении среди*

функций $\eta_2(x, \theta_2)$, $\theta_2 \in \Theta_2$ наилучшей чебышевской аппроксимации для функции $\eta_1(x, \bar{\theta}_1)$ в следующем смысле:

$$\sup_{\xi} \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi(dx) = \inf_{\theta_2 \in \Theta_2} \sup_{x \in \mathcal{X}} |\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)|^2,$$

причем опорные точки T -оптимального плана принадлежат чебышевскому альтернансу

$$\mathcal{A} = \left\{ x^* \in \mathcal{X} \left| |\eta_1(x^*, \bar{\theta}_1) - \eta_2(x^*, \theta_2^*)| = \sup_{x \in \mathcal{X}} |\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)| \right. \right\},$$

где $\theta_2^* \in \Theta_2$ соответствует наилучшей аппроксимации.

Доказательство теоремы можно найти в работе [7]. В главе 2 используя теорему 2 мы получим аналитические представления точек и весов T -оптимальных планов для дискриминации полиномиальных регрессионных моделей и моделей, отличающихся от полиномиальных добавлением дробно-рациональных слагаемых.

1.8 Алгоритм для численного построения T -оптимальных планов

В работах [6] и [12] также был предложен численный итерационный алгоритм для нахождения T -оптимальных планов:

Алгоритм 1. Пусть ξ_0 — это некоторый начальный план и $(\alpha_s)_{s=0}^\infty$ — последовательность положительных вещественных чисел, удовлетворяющая условиям

$$\lim_{s \rightarrow \infty} \alpha_s = 0, \quad \sum_{s=0}^{\infty} \alpha_s = \infty, \quad \sum_{s=0}^{\infty} \alpha_s^2 < \infty.$$

Тогда для $s = 0, 1, \dots$ поочередно выполняем два шага:

1. Находим точку из множества планирования $\mathcal{X}$, такую что

$$x_{s+1} = \arg \max_{x \in \mathcal{X}} \psi_{1,2}(x, \xi_s),$$

2. Вычисляем новое приближение к оптимальному плану:

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s\xi(x_{s+1}),$$

где $\xi(x_{s+1})$ — это план, сосредоточенный в точке x_{s+1} .

Этот алгоритм является аналогом процедуры Федорова–Уинна, которая была изначально разработана для нахождения D -оптимальных планов. Известна следующая теорема, касающаяся сходимости рассмотренного алгоритма (смотри, например, книгу [19], главу 3, параграф 4, где описан аналогичный результат для процедуры Федорова–Уинна):

Теорема 3. Алгоритм Аткинсона–Федорова сходится в том смысле, что

$$\lim_{s \rightarrow \infty} T_{1,2}(\xi_s) = \sup_{\xi} T_{1,2}(\xi).$$

Основным недостатком алгоритма 1 является большое количество точек в результирующем плане. На каждом шаге число точек увеличивается на одну, поэтому в итоге мы имеем план, сосредоточенный в $n_0 + n_{iter}$ точках, где n_0 — мощность носителя начального плана ξ_0 , а n_{iter} — количество проделанных итераций. Подробнее этот вопрос обсуждается в главе 3, где предлагается альтернативный алгоритм, устраняющий обозначенный недостаток.

Величина

$$\text{Eff}(\xi, \bar{\theta}_1) = \frac{T_{1,2}(\xi, \bar{\theta}_1)}{\sup_{\zeta} T_{1,2}(\zeta, \bar{\theta}_1)},$$

называется эффективностью плана ξ . Она показывает то, насколько произвольный план ξ хуже оптимального при фиксированном $\bar{\theta}_1$. Эффективность — это стандартный показатель качества плана, используемый в литературе. Если истинное значение θ_1^u далеко от выбранного нами приближения $\bar{\theta}_1$, то эффективность $\text{Eff}(\xi^*(\bar{\theta}_1), \theta_1^u)$, где $\xi^*(\bar{\theta}_1)$ есть оптимальный план, посчитанный при фиксированном $\bar{\theta}_1$, может оказаться низкой. Чтобы компенсировать этот эффект обычно используют робастные по отношению к выбору $\bar{\theta}_1$ критерии (например, байесовские или минимаксные).

Если план $\xi^* = \xi^*(\bar{\theta}_1)$ является T -оптимальным, а план ξ — нет, то по теореме эквивалентности 1 существует $\dot{x} \in \mathcal{X}$, такой что $\psi_{1,2}(\dot{x}, \xi) > T_P(\xi^*, \bar{\theta}_1)$.

Из этого соотношения следует, что эффективность можно оценить снизу с помощью величины

$$\widehat{\text{Eff}}_T(\xi, \bar{\theta}_1) = \frac{T_{1,2}(\xi, \bar{\theta}_1)}{\max_{x \in \mathcal{X}} \psi_{1,2}(x, \xi)},$$

которая может быть вычислена на каждой итерации алгоритма. По аналогии эффективность можно ввести для любого другого критерия оптимальности. В качестве условия останова для всех алгоритмов в данной работе используется достижение заданной нижней границы эффективности.

1.9 Различные обобщения критерия T -оптимальности

Критерий T -оптимальности имеет ряд существенных ограничений. В этом параграфе приведен краткий обзор работ, посвященных обобщениям критерия T -оптимальности, которые позволяют их преодолеть.

T -критерий был разработан для сравнения двух конкурирующих моделей. Авторы оригинальной работы [6] предложили в [12] его обобщение на случай дискриминации произвольного количества конкурирующих моделей. Одна модель выбиралась в качестве базовой, ее параметры фиксировались, а оптимальный план выбирался таким образом, чтобы он доставлял максимум минимальному значению среди T -критериев для дискриминации базовой модели и каждой из оставшихся моделей. Симметричная версия T -критерия для дискриминации нескольких моделей, T_P -критерий, который представляет из себя взвешенную сумму T -критериев для всех возможных пар конкурирующих моделей, а также численный алгоритм для нахождения оптимальных планов, основанный на многомерной чебышёвской аппроксимации, были введены в [13]. Другое ограничение T -критерия — зависимость от априорных значений для параметров одной из моделей, может быть компенсировано с помощью байесовского подхода, который обсуждался еще в [6]. В главе 3 предложен новый численный метод для нахождения байесовских T_P -оптимальных планов. Еще одно ограничение заключается в требовании о нормальности и гомоскедастичности ошибок наблюдения. В [9] T -критерий был обобщен на случай дискриминации двух моделей при нормальных гетероскедастичных ошибках. В [10] был предложен критерий KL -оптимальности, основанный на расстояниях Кульбака–Лейблера и связанный с тестом отношения правдоподобия, который не требует нормаль-

ности и позволяет решать задачу дискриминации двух произвольных конкурирующих моделей для плотностей ошибок. Критерии из работ [6] и [9] являются частными случаями KL -критерия. В главе 4 численный метод для нахождения T_P -оптимальных планов обобщен на случай введенных по аналогии KL_P -оптимальных планов. В [11] был предложен полу-параметрический критерий, в котором функции регрессии для всех конкурирующих моделей и закон распределения ошибок для одной из них считаются известными, а закон распределения для оставшейся модели получается как решение специальной задачи вариационного исчисления. Изучению полу-параметрических критериев посвящена глава 5.

Глава 2. T -оптимальные планы для дискриминации дробно-рациональных и полиномиальных моделей.

В настоящей главе рассматривается задача построения T -оптимальных планов для дискриминации между полиномиальными регрессионными моделями и аналогичными моделями, содержащими дополнительное дробно-рациональное слагаемое.

Построение T -оптимальных планов представляет собой весьма сложную математическую задачу минимаксного типа и до недавнего времени (до работы [8]) осуществлялось только численными методами за исключением простейшего случая дискриминации полиномиальных моделей, отличающихся на один порядок (этот случай описан, например, в главе 16 книги [26], с.372). Однако было обнаружено (смотри книгу [26], с.369, или статью [7]), что опорными точками T -оптимальных планов служат экстремальные точки функции, представляющей собой разность модели из нулевой гипотезы и ее наилучшей чебышёвской аппроксимации альтернативной моделью. На основе этого факта в работе [8] были аналитически построены T -оптимальные планы для дискриминации полиномиальных моделей, отличающихся на два порядка. Задача нахождения наилучшей чебышёвской аппроксимации лучше всего изучена в случае приближения непрерывных функций полиномами. В этой главе при аналитическом построении T -оптимальных планов для дискриминации полиномиальных и некоторых дробно-рациональных моделей будут использованы известные результаты из теории аппроксимации. Введенный новый способ построения экстремальных точек упомянутых выше функций, возможно, представляет некоторый независимый интерес. Изложение в данной главе опирается на материал из работы [14].

2.1 Простейшая дробь и полином

Рассмотрим пару регрессионных моделей

$$\begin{aligned}\eta_1(x, \theta_1) &= \sum_{i=0}^m \theta_{1,i} x^i + \frac{1}{x-a}, \\ \eta_2(x, \theta_2) &= \sum_{i=0}^m \theta_{2,i} x^i,\end{aligned}\tag{2.1}$$

заданных на промежутке $x \in [-1, 1]$. Задача нахождения T -оптимального плана для дискриминации моделей (2.1) сводится (в силу теоремы 2 из главы 1) к классической задаче теории чебышёвской аппроксимации — нахождению полинома наилучшего приближения для функции $1/(x - a)$. Естественно предполагать, что $a \notin [-1, 1]$. Без ограничения общности будем считать, что $a > 1$.

Напомним (смотри, например, книгу [29], с.42), что полиномы Чебышёва первого и второго рода, задаваемые тригонометрическими формулами

$$T_k(x) = \cos(k \arccos x), \quad U_k(x) = \frac{\sin([k+1] \arccos(x))}{\sin(\arccos(x))}, \quad x \in [-1, 1],$$

соответственно, имеют также следующие явные представления:

$$\begin{aligned} T_k(x) &= \frac{(x + \sqrt{x^2 - 1})^k + (x - \sqrt{x^2 - 1})^k}{2}, \\ U_k(x) &= \frac{(x + \sqrt{x^2 - 1})^{k+1} - (x - \sqrt{x^2 - 1})^{k+1}}{2\sqrt{x^2 - 1}}. \end{aligned} \quad (2.2)$$

Сформулируем лемму о полиномах Чебышёва первого и второго рода, которая будет использована для доказательства теорем, характеризующих опорные точки T -оптимальных планов.

Лемма 2. *Полиномы Чебышёва первого и второго рода имеют следующие комплексные представления:*

$$T_k(x) = \operatorname{Re}(v^k); \quad U_k(x) = \frac{\operatorname{Im}(v^{k+1})}{\operatorname{Im}(v)},$$

где v — комплексное число, описывающее верхнюю половину единичной окружности, т.е. $v = x + i\sqrt{1 - x^2}$ при $x \in [-1, 1]$.

Доказательство. Этот факт следует из явных формул для полиномов Чебышёва, приведенных в (2.2). $\square$

Следующий результат предлагает решение задачи нахождения опорных точек T -оптимальных планов для дискриминации моделей (2.1) при любой степени $m \geq 1$ полиномиальной части в обеих моделях:

Теорема 4. *T -оптимальный план для дискриминации моделей (2.1) сосредоточен в $(m+2)$ -х точках из отрезка $[-1, 1]$, причем точки ± 1 принадлежат*

плану при любом m , а опорные точки из интервала $(-1, 1)$ являются корнями полиномов

$$\Psi_1(x) = U_m(x) - 2\alpha U_{m-1}(x) + \alpha^2 U_{m-2}(x), \quad m \geq 2, \quad (2.3)$$

$$\Psi_2(x) = (\alpha^2 - 1) T_m(x) + (2\alpha - x - \alpha^2 x) U_{m-1}(x), \quad m \geq 1, \quad (2.4)$$

где $\alpha = a - \sqrt{a^2 - 1}$.

Доказательство. Пусть $a > 1$. Введем обозначение

$$\Phi(x) = \frac{1}{x - a} - P_m(x),$$

где $P_m(x)$ — это полином степени m с вещественными коэффициентами, наименее уклоняющийся от $1/(x - a)$. В главе 2 книги [30] было доказано, что

$$\Phi(x) = \frac{M}{2} \left[v^m \frac{\alpha - v}{1 - \alpha v} + v^{-m} \frac{1 - \alpha v}{\alpha - v} \right],$$

где зависимости между v и x , α и a выражаются уравнениями

$$x = \frac{1}{2} \left[v + \frac{1}{v} \right], \quad a = \frac{1}{2} \left[\alpha + \frac{1}{\alpha} \right], \quad \alpha = a - \sqrt{a^2 - 1}, \quad v = x + i\sqrt{1 - x^2},$$

а константа

$$M = \frac{4\alpha^{m+2}}{[1 - \alpha^2]^2} \quad (2.5)$$

есть величина наилучшего приближения.

По теореме Чебышёва (книга [30], страница 66) число точек в альтернансе функции $\Phi(x)$, а значит и число точек в оптимальном плане, не меньше, чем $m + 2$. Найдём экстремальные точки функции $\Phi(x)$. Обозначим

$$z = v^m \frac{\alpha - v}{1 - \alpha v}.$$

Так как $|v| = 1$, то и $|z| = 1$, поэтому $\Phi(x)$ можно переписать в виде:

$$\Phi(x) = \frac{M}{2} \left[z + \frac{1}{z} \right] = \frac{M}{2} [z + \bar{z}] = M \operatorname{Re}(z).$$

Функция $\Phi(x)$ достигает наибольших по модулю значений в точках, где $\operatorname{Re}(z) = \pm 1$, или, что то же самое, $\operatorname{Im}(z) = 0$. В итоге имеем уравнение для экстремальных точек

$$\operatorname{Im}(z) = \operatorname{Im} \left(v^m \frac{\alpha - v}{1 - \alpha v} \right) = \operatorname{Im} \left(\frac{v^m [\alpha - v][1 - \alpha \bar{v}]}{|1 - \alpha v|^2} \right) = 0, \quad (2.6)$$

которое эквивалентно уравнению

$$\operatorname{Im}(2\alpha v^m - v^{m+1} - \alpha^2 v^{m-1}) = 2\alpha \operatorname{Im}(v^m) - \operatorname{Im}(v^{m+1}) - \alpha^2 \operatorname{Im}(v^{m-1}) = 0. \quad (2.7)$$

Заметим, что тождество (2.7) выполнено при $x = \pm 1$, так как в этом случае $\operatorname{Im}(v) = 0$. Из этого следует, что точки $x = \pm 1$ принадлежат носителю T -оптимального плана. Если $x \in (-1, 1)$, то $\operatorname{Im}(v) > 0$, поэтому можно умножить выражение (2.7) на $1/\operatorname{Im}(v)$:

$$2\alpha \frac{\operatorname{Im}(v^m)}{\operatorname{Im}(v)} - \frac{\operatorname{Im}(v^{m+1})}{\operatorname{Im}(v)} - \alpha^2 \frac{\operatorname{Im}(v^{m-1})}{\operatorname{Im}(v)} = 0.$$

Пользуясь представлением полиномов Чебышёва второго рода через комплексные числа из леммы 2, получим (2.3):

$$2\alpha U_{m-1}(x) - U_m(x) - \alpha^2 U_{m-2}(x) = 0.$$

Число корней в последнем уравнении не превосходит m . Заметим, что последняя формула верна при $m \geq 2$.

Теперь выведем уравнение (2.4), верное при $m \geq 1$. Запишем v^m в тригонометрической форме

$$v^m = \cos(m \arccos(x)) + i \sin(m \arccos(x)). \quad (2.8)$$

Подставим выражение (2.8) для v^m в уравнение (2.6) и получим

$$\begin{aligned} \operatorname{Im}(z) = \sqrt{1-x^2} \frac{\alpha^2 - 1}{1 - 2\alpha x + \alpha^2} \cos(m \arccos(x)) \\ + \frac{2\alpha - x - \alpha^2 x}{1 - 2\alpha x + \alpha^2} \sin(m \arccos(x)) = 0. \end{aligned}$$

Таблица 1 — Опорные точки T -оптимального плана для дискриминации моделей (2.1)

m	$x_1, \dots, x_n$
1	$-1, \alpha, 1$
2	$-1, -\frac{1}{2} + \frac{\alpha}{2}, \frac{1}{2} + \frac{\alpha}{2}, 1$
3	$-1, \frac{\alpha - \sqrt{\alpha^2 + 8}}{4}, \frac{\alpha}{2}, \frac{\alpha + \sqrt{\alpha^2 + 8}}{4}, 1$
4	$-1, \frac{\alpha - 1 - \sqrt{\alpha^2 + 2\alpha + 5}}{4}, \frac{\alpha + 1 - \sqrt{\alpha^2 - 2\alpha + 5}}{4}, \frac{\alpha - 1 + \sqrt{\alpha^2 + 2\alpha + 5}}{4}, \frac{\alpha + 1 + \sqrt{\alpha^2 - 2\alpha + 5}}{4}, 1$

Умножим последнее выражение на $1/\sqrt{1-x^2}$, полагая, что $x \neq \pm 1$, и получим формулу для нахождения точек плана внутри промежутка $(-1, 1)$, верную при $m \geq 1$:

$$(\alpha^2 - 1) T_m(x) + (2\alpha - x - \alpha^2 x) U_{m-1}(x) = 0.$$

□

В таблице 1 приведены опорные точки T -оптимального плана для дискриминации моделей (2.1), полученные как корни полиномов (2.3) и (2.4) из теоремы 4, при небольших значениях m .

2.2 Другие дробно-рациональные модели

В этом параграфе рассматриваются две другие пары конкурирующих регрессионных моделей. Первая модель в каждой паре имеет дробно-рациональное слагаемое со знаменателем второй степени, вторая модель является полиномиальной. Решения задач T -оптимального планирования для этих пар связаны. Итак, пусть у нас имеются две следующие пары регрессионных моделей

$$\begin{aligned} \eta_1(x, \theta_1) &= \sum_{i=0}^m \theta_{1,i} x^i + \frac{1}{x^2 - a}; & \eta_1(x, \theta_1) &= \sum_{i=0}^m \theta_{1,i} x^i + \frac{x}{x^2 - a}; \\ \eta_2(x, \theta_2) &= \sum_{i=0}^m \theta_{2,i} x^i; & \eta_2(x, \theta_2) &= \sum_{i=0}^m \theta_{2,i} x^i; \end{aligned} \tag{2.9}$$

заданные на промежутке $x \in [-1, 1]$. Будем, как и ранее, полагать, что $a > 1$.

Сформулируем результат, который позволяет аналитически решать задачу нахождения опорных точек T -оптимального плана для моделей (2.9) в некоторых частных случаях.

Теорема 5. *T -оптимальные планы для дискриминации моделей (2.9) сосредоточены в $(m+2)$ -х точках. Если m — нечетное, то опорные точки из интервала $(-1,1)$ плана для левой пары моделей совпадают с корнями полиномов*

$$\Psi_1(x) = U_m(x) - 2\alpha^2 U_{m-2}(x) + \alpha^4 U_{m-4}(x), \quad m \geq 4, \quad (2.10)$$

$$\Psi_2(x) = 2x(\alpha^4 - 1)T_{m-1}(x) + (\alpha^4 + 2\alpha^2 + 1 - 2x^2[\alpha^4 + 1])U_{m-2}(x), \quad m \geq 2, \quad (2.11)$$

а если m — четное, то это верно для правой пары. Точки ± 1 принадлежат оптимальному плану при любом m ; $\alpha = a - \sqrt{a^2 - 1}$.

К сожалению, для четных m не удалось получить аналитического решения для левой пары моделей, а для нечетного m — для правой пары, так как мы не нашли соответствующих результатов в работах по теории аппроксимации.

Доказательство. Доказательство данной теоремы также опирается на известный результат из теории аппроксимации. Пусть $a > 1$. Рассмотрим функцию:

$$\Phi(x) = v^{m-1} \frac{v^2 - \alpha^2}{\alpha^2 v^2 - 1} + v^{1-m} \frac{\alpha^2 v^2 - 1}{v^2 - \alpha^2},$$

где зависимости между v и x , α и a выражаются уравнениями

$$x = \frac{1}{2} \left[v + \frac{1}{v} \right], \quad a = \frac{1}{2} \left[\alpha + \frac{1}{\alpha} \right], \quad \alpha = a - \sqrt{a^2 - 1}, \quad v = x + i\sqrt{1 - x^2}.$$

В [31] доказано, что точки экстремума этой функции являются точками альтернанса для задачи наилучшего приближения функции $1/(x^2 - a^2)$ полиномом степени m , если m является нечетным, и функции $x/(x^2 - a^2)$, если m является четным.

Введем обозначение

$$z = v^{m-1} \frac{v^2 - \alpha^2}{\alpha^2 v^2 - 1}.$$

Заметим, что $|z| = 1$, так как $|v| = 1$, поэтому

$$\Phi(x) = z + \frac{1}{z} = z + \bar{z} = 2\operatorname{Re}(z).$$

Функция $\Phi(x)$ достигает своего наибольшего значения при $\operatorname{Re}(z) = \pm 1$ или, что то же самое, при $\operatorname{Im}(z) = 0$. Проведем ряд эквивалентных преобразований:

$$\begin{aligned} \operatorname{Im}(z) &= \operatorname{Im}\left(v^{m-1} \frac{v^2 - \alpha^2}{\alpha^2 v^2 - 1}\right) = \operatorname{Im}\left(v^{m-1} \frac{[v^2 - \alpha^2] [\overline{\alpha^2 v^2 - 1}]}{|\alpha^2 v^2 - 1|^2}\right) = 0, \\ \operatorname{Im}(v^{m-1} [v^2 - \alpha^2] [\overline{\alpha^2 v^2 - 1}]) &= 0, \end{aligned} \quad (2.12)$$

$$\operatorname{Im}(2\alpha^2 v^{m-1} - \alpha^4 v^{m-3} - v^{m-1}) = 0. \quad (2.13)$$

Из уравнения (2.13) видно, что точки $x = \pm 1$ принадлежат альтернансу. Умножим уравнение (2.13) на $1/\operatorname{Im}(v)$, считая, что $x \neq \pm 1$, и получим следующее уравнение для нахождения опорных точек оптимального плана при $x \in (-1, 1)$:

$$U_m(x) - 2\alpha^2 U_{m-2}(x) + \alpha^4 U_{m-4}(x) = 0.$$

Последнее уравнение позволяет найти опорные точки T -оптимального плана при $m \geq 4$.

Выведем теперь другое уравнение, позволяющее найти опорные точки при $m \geq 2$. Вернемся к уравнению (2.12). Так как

$$v^{m-1} = \cos((m-1) \arccos(x)) + i \sin((m-1) \arccos(x)),$$

уравнение (2.12) переписывается в виде:

$$\begin{aligned} 2x\sqrt{1-x^2}(\alpha^4 - 1) \cos((m-1) \arccos(x)) \\ + (\alpha^4 + 2\alpha^2 + 1 - 2x^2[\alpha^4 + 1]) \sin((m-1) \arccos(x)) = 0. \end{aligned}$$

Умножим полученное выражение на $1/\sqrt{1-x^2}$, считая, что $x \neq \pm 1$. В итоге:

$$2x(\alpha^4 - 1) T_{m-1}(x) + (\alpha^4 + 2\alpha^2 + 1 - 2x^2[\alpha^4 + 1]) U_{m-2}(x) = 0.$$

Последняя формула работает для $m \geq 2$. □

Таблица 2 — Опорные точки T -оптимального плана для дискриминации моделей (2.9)

m	$x_1, \dots, x_n$
2	$-1, -\frac{\alpha^2+1}{2}, \frac{\alpha^2+1}{2}, 1$
3	$-1, -\sqrt{\frac{\alpha^2+1}{2}}, 0, \sqrt{\frac{\alpha^2+1}{2}}, 1$
4	$-1, -\frac{\sqrt{4\alpha^2+5}+1}{4}, -\frac{\sqrt{4\alpha^2+5}-1}{4}, \frac{\sqrt{4\alpha^2+5}-1}{4}, \frac{\sqrt{4\alpha^2+5}+1}{4}, 1$
5	$-1, -\frac{\sqrt{\alpha^2+3}}{2}, -\frac{\sqrt{\alpha^2+1}}{2}, 0, \frac{\sqrt{\alpha^2+1}}{2}, \frac{\sqrt{\alpha^2+3}}{2}, 1$

В таблице 2 приведены опорные точки T -оптимального плана для дискриминации моделей (2.9), полученные как корни полиномов (2.10) и (2.11) из теоремы 5, при небольших значениях m .

2.3 ЕМАХ-модель и квадратичная модель

2.3.1 Локально-оптимальный план

В этом параграфе приводится пример реальной прикладной задачи, для которой, благодаря теореме 4, можно получить аналитическое решение. Рассмотрим пару конкурирующих регрессионных моделей:

$$\begin{aligned}\eta_1(x, \theta_1) &= \theta_{1,1} + \frac{\theta_{1,2}x}{\theta_{1,3} + x}; \\ \eta_2(x, \theta_2) &= \theta_{2,1} + \theta_{2,2}x(\theta_{2,3} - x),\end{aligned}\tag{2.14}$$

заданных на промежутке $x \in [0, 500]$. Вектор априорных значений для параметров первой модели θ_1 задается следующим образом

$$\bar{\theta}_1 = (\bar{\theta}_{1,1}, \bar{\theta}_{1,2}, \bar{\theta}_{1,3}) = (60, 294, 25).$$

Это часть задачи 4.2 из работы [15]. Введем функцию

$$\begin{aligned}\eta(x, b, \gamma) &= \eta_1(x, \theta_1) - \eta_2(x, \theta_2) = \theta_{1,1} + \frac{\theta_{1,2}x}{\theta_{1,3} + x} - \theta_{2,1} - \theta_{2,2}x(\theta_{2,3} - x) \\ &= \theta_{1,1} + \theta_{1,2} - \frac{\theta_{1,3}}{\theta_{1,3} + x} - \theta_{2,1} - \theta_{2,2}\theta_{2,3}x + \theta_{2,2}x^2 \\ &= b_1 + b_2x + b_3x^2 - \frac{\gamma}{x + \gamma},\end{aligned}$$

где $b_1 = \theta_{1,1} + \theta_{1,2} - \theta_{2,1}$, $b_2 = -\theta_{2,2}\theta_{2,3}$, $b_3 = \theta_{2,2}$, $\gamma = \theta_{1,3}$. Заметим, что

$$\sup_{\xi} \inf_b \int_{\mathcal{X}} \eta(x, b, \gamma)^2 \xi(dx) = \sup_{\xi} \inf_b \frac{1}{\gamma^2} \int_{\mathcal{X}} \eta(x, b, \gamma)^2 \xi(dx).$$

Мы получили, что задача нахождения T -оптимального плана для пары моделей (2.14) эквивалентна задаче для пары

$$\begin{aligned} \eta_1(x, \tilde{\theta}_1) &= \frac{1}{x + \gamma}; \\ \eta_2(x, \tilde{\theta}_2) &= b_1 + b_2x + b_3x^2. \end{aligned} \quad (2.15)$$

при $x \in [0, 500]$. При этом оптимальный план зависит только от γ (в нашем случае $\gamma = \bar{\theta}_{1,3} = 25$) и умножение $\eta_1(x, \tilde{\theta}_1)$ на константу не повлияет на оптимальный план. Чтобы воспользоваться теоремой 4, необходимо перевести задачу на промежуток $[-1, 1]$. Заметим, что имеет место взаимно однозначное соответствие

$$g(y) = -f(x) = -f(250 - 250y),$$

где $g(y) = 1/(y - a)$ и $f(x) = 250/(x + \gamma)$. Если y пробегает отрезок $[-1, 1]$ от большего значения к меньшему, то $x = 250 - 250y$ пробегает отрезок $[0, 500]$. В итоге имеем

$$g(y) = -f(250 - 250y) = -\frac{250}{250 - 250y + \gamma} = \frac{1}{y - \frac{\gamma+250}{250}} = \frac{1}{y - a}. \quad (2.16)$$

То есть мы получили, что опорные точки T -оптимального плана ξ^* для дискриминации моделей (2.15) совпадают с опорными точками для пары моделей (2.1) при $m = 2$ и $a = \frac{\gamma+250}{250} = 1.1$, переведенными на промежуток $[0, 500]$ с помощью соответствия $x = 250 - 250y$ и записанными в обратном порядке. Используя таблицу 1, получаем

$$x_1^* = 0, \quad x_2^* = 250 - 250 \left[\frac{1}{2} + \frac{\alpha}{2} \right], \quad x_3^* = 250 - 250 \left[-\frac{1}{2} + \frac{\alpha}{2} \right], \quad x_4^* = 500. \quad (2.17)$$

Для того, чтобы получить веса, воспользуемся тем, что

$$\sum_{i=1}^n \omega_i^* [\eta_1(x_i^*, \bar{\theta}_1) - \eta_2(x_i^*, \theta_2^*)] \frac{\partial \eta_2(x_i^*, \theta_2)}{\partial \theta_2} \bigg|_{\theta_2 = \theta_2^*} = 0 \quad (2.18)$$

для любого T -оптимального плана $\xi^* = \{x_1^*, \dots, x_n^*; \omega_1^*, \dots, \omega_n^*\}$, где θ_2^* — решение оптимизационной задачи $\inf_{\theta_2 \in \Theta_2}$ в (1.12) при $\xi = \xi^*$. Этот факт можно найти, например, в [7] (теорема 2.2). Учитывая, что

$$\eta_1(x_i^*, \bar{\theta}_1) - \eta_2(x_i^*, \theta_2^*) = - [\eta_1(x_{i+1}^*, \bar{\theta}_1) - \eta_2(x_{i+1}^*, \theta_2^*)], \quad i = 1, \dots, n-1, \quad (2.19)$$

так как опорные точки принадлежат чебышёвскому альтернансу, получаем систему линейных уравнений относительно ω_i^*

$$\begin{cases} \omega_1^* - \omega_2^* + \omega_3^* - \omega_4^* = 0 \\ \omega_1^* x_1 - \omega_2^* x_2 + \omega_3^* x_3 - \omega_4^* x_4 = 0 \\ \omega_1^* x_1^2 - \omega_2^* x_2^2 + \omega_3^* x_3^2 - \omega_4^* x_4^2 = 0 \\ \omega_1^* + \omega_2^* + \omega_3^* + \omega_4^* = 1 \end{cases}$$

Первые три уравнения в системе получаются из (2.18) и (2.19), а последнее — это естественное ограничение на веса. Решая систему, получаем

$$\begin{aligned} \omega_1^* &= \frac{\omega_2^*(x_2^2 - x_4^2) - \frac{1}{2}(x_3^2 - x_4^2)}{x_1^2 - x_3^2}, \\ \omega_2^* &= \left[\frac{1}{2} \frac{x_3^2 - x_4^2}{x_1^2 - x_3^2} - \frac{1}{2} \frac{x_3 - x_4}{x_1 - x_3} \right] / \left[\frac{x_2^2 - x_4^2}{x_1^2 - x_3^2} - \frac{x_2 - x_4}{x_1 - x_3} \right], \\ \omega_3^* &= \frac{1}{2} - \omega_1^*, \\ \omega_4^* &= \frac{1}{2} - \omega_2^*. \end{aligned} \quad (2.20)$$

Таким образом, формулы (2.17) и (2.20) задают искомое аналитическое представление точек и весов T -оптимального плана для дискриминации моделей (2.14). Подставляя числовые значения, получаем

$$\xi_T^* = \begin{bmatrix} 0 & 44.782 & 294.782 & 500 \\ 0.348 & 0.452 & 0.152 & 0.048 \end{bmatrix}.$$

2.3.2 Робастный план

Обсудим теперь задачу построения робастных планов для пары моделей (2.14). Напомним, что величина

$$\text{Eff}(\xi, \bar{\theta}_1) = \frac{T_{1,2}(\xi, \bar{\theta}_1)}{\sup_{\zeta} T_{1,2}(\zeta, \bar{\theta}_1)},$$

называется эффективностью плана ξ при фиксированном $\bar{\theta}_1$. Чтобы компенсировать зависимость оптимального плана от выбора параметров $\bar{\theta}_1$ обычно используют робастные критерии, один из которых мы сейчас рассмотрим. Будем считать, что вместо одного фиксированного значения $\bar{\theta}_1$ у нас имеется априорное распределение на множестве Θ_1 параметров первой модели. Стандартизированным байесовским T -оптимальным называется план, доставляющий максимум величине $T_{\text{SB}}(\xi)$, равной интегралу эффективности по априорному распределению (смотри, например, статью [32]).

Пусть априорное распределение сосредоточено в конечном количестве точек h и является равномерным. Тогда

$$T_{\text{SB}}(\xi) = \frac{1}{h} \sum_{i=1}^h \text{Eff}(\xi, \theta_1^{(i)}). \quad (2.21)$$

Заметим, что критерий (2.21) эквивалентен критерию

$$T_P(\xi) = \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \xi(dx), \quad (2.22)$$

где $\nu = h + 1$, $\eta_i(x, \theta_i) = \eta_1(x, \theta_1)$, при $i = 1, \dots, h$, $\eta_{h+1}(x, \theta_{h+1}) = \eta_2(x, \theta_2)$, $\bar{\theta}_i = \theta_1^{(i)}$ при $i = 1, \dots, h$ и матрица $P = \{p\}_{i,j=1}^{h+1}$ имеет вид

$$P = \frac{1}{h} \begin{bmatrix} 0 & 0 & \dots & 0 & 1/\sup_{\zeta} T_{1,2}(\zeta, \theta_1^{(1)}) \\ 0 & 0 & \dots & 0 & 1/\sup_{\zeta} T_{1,2}(\zeta, \theta_1^{(2)}) \\ & & \dots & & \\ 0 & 0 & \dots & 0 & 1/\sup_{\zeta} T_{1,2}(\zeta, \theta_1^{(h)}) \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}. \quad (2.23)$$

Заметим также, что стандартный T -критерий (1.12) эквивалентен (2.22) при $\nu = 2$, $p_{1,2} = 1$, $p_{1,1} = p_{2,1} = p_{2,2} = 0$. Численным алгоритмам для поиска оптимальных относительно критерия (2.22) планов (T_P -оптимальных планов) посвящена глава 3 (результаты опубликованы в [15]). Основная проблема, мешающая применить методологию из главы 3, состоит в том, чтобы вычислить матрицу P , которая в данном случае зависит от h обычных T -оптимальных планов. В сущности, $\sup_{\zeta} T_{1,2}(\zeta, \theta_1^{(i)})$, $i = 1, \dots, h$ можно вычислить напрямую, но в нашем примере мы можем воспользоваться тем, что эти величины равны квадратам величин наилучшего приближения для соответствующих задач чебышёвской аппроксимации. Посчитаем стандартизированный байесовский план для дискриминации моделей (2.15) с равномерным априорным распределением для параметра γ , сосредоточенным на

$$\Theta_{1,\gamma} = (\theta_{1,\gamma}^{(1)}, \dots, \theta_{1,\gamma}^{(10)}) = (25, 50, 75, 100, 125, 150, 175, 200, 225, 250).$$

Эта задача соответствует (см. формулу (2.16)) нахождению стандартизированного плана для моделей (2.1) с равномерным априорным распределением для a , заданным на

$$\Theta_{1,a} = (\theta_{1,a}^{(1)}, \dots, \theta_{1,a}^{(10)}) = (1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0).$$

В этом случае по формуле (2.5) для величины наилучшего приближения имеем

$$\sup_{\zeta} T_{1,2}(\zeta, \theta_{1,a}^{(i)}) = M_i^2, \quad M_i = \frac{4\alpha_i^{m+2}}{[1 - \alpha_i^2]^2}, \quad \alpha_i = \theta_{1,a}^{(i)} - \sqrt{(\theta_{1,a}^{(i)})^2 - 1}, \quad i = 1, \dots, 10. \quad (2.24)$$

Подставим (2.24) в (2.23) и воспользуемся алгоритмом из [15], реализованным в пакете [17] для R, который будет описан далее по тексту в главе 3, чтобы найти оптимальный относительно критерия (2.22) план для пары (2.1). Воспользуемся описанным выше способом перевода опорных точек с отрезка $[-1, 1]$ на отрезок $[0, 500]$ и перепишем точки и веса оптимального плана в обратном порядке, чтобы получить план для (2.15). В итоге имеем

$$\xi_{SB}^* = \begin{bmatrix} 0 & 75.1663 & 327.8767 & 500 \\ 0.268 & 0.410 & 0.232 & 0.090 \end{bmatrix}.$$

Таблица 3 — Эффективность байесовского плана ξ_{SB}^* относительно локально оптимальных критериев с фиксированными параметрами $\theta_{1,\gamma}^{(i)}$

$\theta_{1,\gamma}^{(i)}$	25	50	75	100	125	150	175	200	225	250
Eff	0.825	0.942	0.984	0.998	0.999	0.995	0.987	0.979	0.970	0.960

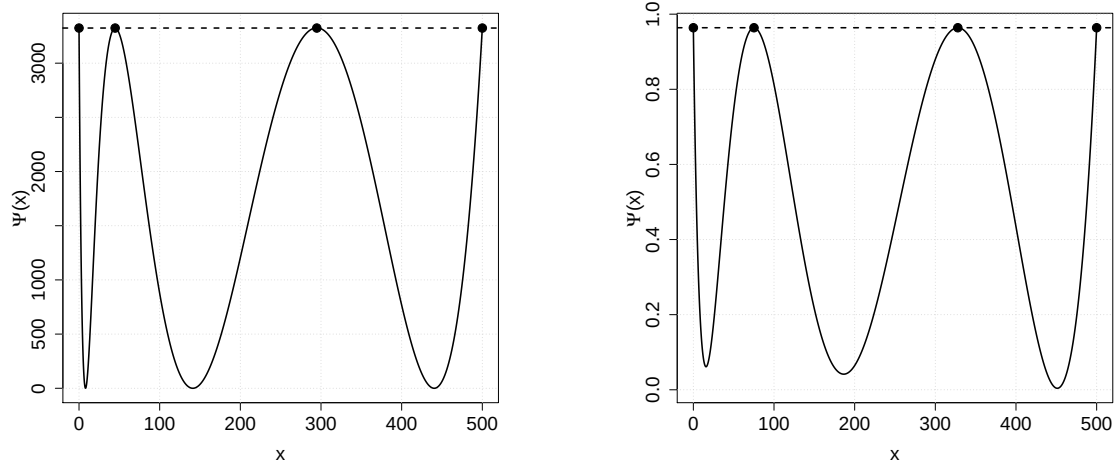
Эффективности стандартизированного байесовского плана ξ_{SB}^* с равномерным априорным распределением на $\Theta_{1,\gamma}$ относительно локально оптимальных планов, посчитанных в каждой отдельной точке $\Theta_{1,\gamma}$, представлены в таблице 3, где $\text{Eff} = \text{Eff}(\xi_{SB}^*, \theta_{1,\gamma}^{(i)})$. Проверку того, что найденные планы ξ_T^* и ξ_{SB}^* являются оптимальными относительно (1.12) и (2.21) соответственно, можно провести с помощью теоремы эквивалентности (см. [15], теоремы 2.1 и 2.2). Теорема эквивалентности для критерия (2.22), к которому сводятся критерии (1.12) и (2.21), может быть сформулирована следующим образом: план ξ^* является T_P -оптимальным тогда и только тогда, когда для функции

$$\Psi(x, \xi) = \sum_{i,j=1}^{\nu} p_{i,j} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \hat{\theta}_{i,j})]^2, \quad (2.25)$$

где $\hat{\theta} = \hat{\theta}(\xi)$ — это значения параметров, на которых достигаются $\inf_{\theta_{i,j} \in \Theta_j}$ в (2.22), для всех точек $x \in \mathcal{X}$ выполнено неравенство

$$\Psi(x, \xi^*) \leq T_P(\xi^*), \quad (2.26)$$

причем в опорных точках оптимального плана в (2.26) достигается равенство. На рис. 2.1 приводятся иллюстрации к теоремам эквивалентности. Из иллюстраций видно, что найденные планы удовлетворяют необходимым и достаточным условиям оптимальности.

а) $\psi_{1,2}(x, \xi_T^*)$ б) $\Psi(x, \xi_{SB}^*)$

Непрерывными линиями обозначены графики функций $\psi_{1,2}(x, \xi_T^*)$ и $\Psi(x, \xi_{SB}^*)$, задаваемых формулами (1.14) и (2.25), пунктирной — значения функционалов $T_{1,2}(\xi_T^*)$ и $T_P(\xi_{SB}^*)$, определенных в (1.12) и (2.22), а заштрихованными точками — опорные точки оптимальных планов ξ_T^* и ξ_{SB}^* .

Рисунок 2.1 — Иллюстрации к теоремам эквивалентности для оптимальных планов ξ_T^* и ξ_{SB}^*

Глава 3. Построение байесовских T_P -оптимальных планов

Обычный T -критерий оптимальности для планов дискриминации имеет ряд существенных недостатков: он является локальным, то есть зависит от фиксированных параметров одной из моделей, подходит только для сравнения двух конкурирующих моделей и не является симметричным относительно выбора модели, параметры которой мы фиксируем. Из литературы известны несколько обобщений T -критерия, компенсирующие перечисленные выше недостатки. Так для дискриминации нескольких регрессионных моделей в работе [13] был предложен критерий T_P -оптимальности, который представляет из себя взвешенную сумму попарных квадратов разностей между конкурирующими моделями, причем для каждой пары параметры одной из моделей фиксированы, а параметры другой находятся по методу наименьших квадратов. T_P -критерий также является локальным, так как зависит от фиксированных заранее параметров. Для частичного преодоления локальности любого критерия оптимальности можно использовать его байесовский аналог, в котором вместо фиксированного значения параметров используется фиксированное априорное распределение.

В этой главе предложен метод построения байесовских T_P -оптимальных планов для дискриминации произвольного количества регрессионных моделей. Доказано, что построение байесовских планов в случае дискретных априорных распределений может быть сведено к построению локально-оптимальных планов с большим количеством конкурирующих моделей. Предложен новый двухэтапный численный алгоритм для нахождения T_P -оптимальных планов, суть которого состоит в чередовании обновления носителя промежуточного плана и оптимизации по его весам, доказана его сходимости, а также предложены две процедуры для выполнения оптимизации по весам. Продемонстрировано превосходство предлагаемых алгоритмов над наиболее часто используемым в литературе. Изложение в данной главе опирается на материал из работы [15]. Алгоритмы, описанные в параграфе 3.4, реализованы в качестве пакета [17] для языка R и доступны через официальный репозиторий CRAN.

3.1 Различные аналоги критерия T -оптимальности

В этом параграфе перечислены различные обобщения критерия T -оптимальности на случай дискриминации произвольного количества конкурирую-

щих регрессионных моделей. Также приведена байесовская версия стандартного T -критерия для дискриминации двух моделей. Для упрощения изложения в лаконичной форме повторяется часть материала главы 1.

На протяжении этой главы будем полагать, что в рамках стандартного уравнения регрессии

$$y = \eta(x, \theta) + \varepsilon \quad (3.1)$$

имеется ν конкурирующих регрессионных моделей

$$\eta_i(x, \theta_i), \quad x \in \mathcal{X}, \quad \theta_i \in \Theta_i, \quad i = 1, \dots, \nu. \quad (3.2)$$

Будем считать, что множества $\mathcal{X}$ и Θ_i компактны, функции $\eta_i(x, \theta_i)$ непрерывны по x при всех $i = 1, \dots, \nu$, а ошибки в (3.1) независимы и имеют нормальное распределение с нулевым средним и постоянной дисперсией. Мы хотим спланировать такой эксперимент, по результатам которого можно было бы определить “истинную” модель среди (3.2). При этом мы подразумеваем, что “истинная” модель, то есть та модель, которой в точности соответствуют экспериментальные данные, находится среди (3.2). В случае, когда конкурирующих моделей всего две, то есть $\nu = 2$, Аткинсон и Федоров в работе [6] предложили искать план эксперимента, доставляющий максимум величине

$$T_{1,2}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \left[\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2) \right]^2 \xi(dx), \quad (3.3)$$

где $\bar{\theta}_1$ есть некоторый априорно фиксированный вектор параметров. Такой план называется (локальным) T -оптимальным планом для дискриминации двух моделей η_1 и η_2 . Его статистическая интерпретация в случае точного плана, то есть дискретной вероятностной меры с положительными рациональными весами на $\mathcal{X}$, такова: в случае линейной по параметрам модели $\eta_2(x, \theta_2)$ он позволяет максимизировать нижнюю границу мощности χ^2 -теста (если дисперсия ошибок (3.1) известна, или F -теста, если она неизвестна) при проверке гипотезы о том, что истинной является модель η_1 против альтернативы, где истинной является модель η_2 . Максимум величины (3.3) проще искать в классе приближенных планов, то есть дискретных вероятностных мер с положительными вещественными весами, а затем переходить от них к точным планам. Далее каждый раз,

когда мы будем говорить о плане эксперимента, будет подразумеваться приближенный план. В отличие от подхода, основанного на проверке гипотезы о том, что часть параметров самой общей из двух моделей равна нулю, подход, предложенный в [6], не требует того, чтобы модели были вложенными. Поскольку план зависит от $\bar{\theta}_1$ он является локальным в смысле [27]. Априорное значение $\bar{\theta}_1$ берется, как правило, из другого эксперимента (например, из специализированного эксперимента для оценки параметров) или исходя из опыта эксперта. Заметим, что критерий (3.3) является несимметричным в том смысле, что поменяв местами модели η_1 и η_2 местами мы, в сущности, получим другой критерий. Если модели не вложены, то возникает вопрос, параметры какой из моделей нам следует зафиксировать, а по каким параметрам следует искать $\inf$ (в случае вложенных моделей фиксируются параметры самой общей модели, потому что иначе $\inf$ в (3.3) будет равен нулю).

В последующей работе [12] те же авторы предложили обобщение критерия (3.3) на случай произвольного количества конкурирующих моделей

$$\begin{aligned} T(\xi, \bar{\theta}_1) &= \min_{2 \leq j \leq \nu} T_{1,j}(\xi, \bar{\theta}_1) \\ &= \min_{2 \leq j \leq \nu} \inf_{\theta_j \in \Theta_j} \int_{\mathcal{X}} \left[\eta_1(x, \bar{\theta}_1) - \eta_j(x, \theta_j) \right]^2 \xi(dx), \end{aligned} \quad (3.4)$$

где параметры $\bar{\theta}_1$ имеют тот же смысл, что и раньше. План, доставляющий максимум (3.4), называется (локальным) Т-оптимальным планом для дискриминации нескольких моделей η_i , $i = 1, \dots, \nu$. Здесь мы снова фиксируем параметры одной из моделей и максимизируем мощность теста при сравнении пар моделей η_1 и η_i , для $i = 2, \dots, \nu$ в наименее благоприятном случае (то есть в том случае, когда модели из пары больше всего похожи). Опять возникает вопрос, какую из конкурирующих моделей следует взять в качестве базовой модели η_1 .

В недавней работе [13] была предложена другая версия критерия (3.4):

$$\begin{aligned} T_P(\xi, \bar{\theta}_1, \dots, \bar{\theta}_\nu) &= \sum_{i,j=1}^{\nu} p_{i,j} T_{i,j}(\xi, \bar{\theta}_i) \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \left[\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j}) \right]^2 \xi(dx), \end{aligned} \quad (3.5)$$

где $\{p_{i,j}\}_{i,j=1}^\nu$ — это набор неотрицательных весов, при этом $p_{i,i} = 0$ для всех $i = 1, \dots, \nu$. План, максимизирующий (3.5), называется (локальным) T_P -оптимальным планом для дискриминации нескольких моделей η_i , $i = 1, \dots, \nu$. Здесь несимметричность исходного критерия Т-оптимальности компенсируется за счет возможности проведения двух сравнений для каждой пары (i, j) , где мы поочередно фиксируем параметры обеих сравниваемых моделей.

Для решения проблемы локальности критериев (3.3), (3.4), (3.5), то есть для уменьшения зависимости этих критериев от априорных параметров, можно использовать байесовский подход, предложенный еще в пионерской работе [6] для критерия (3.3). Суть подхода заключается в том, что вместо фиксированного априорного вектора параметров $\bar{\theta}_1$ берется фиксированное априорное вероятностное распределение на Θ_1 и критерий (3.3) интегрируется по этому распределению. Байесовская версия критерия (3.3) имеет вид

$$T_{1,2}^B(\xi) = \int_{\Theta_1} \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \left[\eta_1(x, \lambda_1) - \eta_2(x, \theta_2) \right]^2 \xi(dx) \mathcal{P}_1(d\lambda_1), \quad (3.6)$$

где $\mathcal{P}_1$ есть априорное распределение для параметров $\theta_1 \in \Theta_1$. План, максимизирующий (3.6), называется байесовским Т-оптимальным планом для дискриминации двух моделей η_1 и η_2 . Обобщение байесовского подхода на случай критериев (3.4) и (3.5) не составляет труда. В этой главе обсуждаются эффективные численные алгоритмы для построения планов, оптимальных с точки зрения байесовской версии критерия (3.5).

3.2 Сведение байесовских планов к локально оптимальным

Сформулируем байесовскую версию критерия T_P -оптимальности (3.5), изучению которой посвящена текущая глава. Байесовским T_P -оптимальным планом мы будем называть дискретную вероятностную меру, доставляющую максимум величине

$$T_P^B(\xi) = \sum_{i,j=1}^\nu p_{i,j} \int_{\Theta_i} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \left[\eta_i(x, \lambda_i) - \eta_j(x, \theta_{i,j}) \right]^2 \xi(dx) \mathcal{P}_i(d\lambda_i). \quad (3.7)$$

Здесь $p_{i,j} \geq 0$, $p_{i,i} = 0$ при любом i , а мера $\mathcal{P}_i$ для каждого $i = 1, \dots, \nu$ задает априорное распределение для параметров θ_i модели η_i такое, что все интегралы в формуле (3.7) определены.

В настоящем параграфе мы покажем, что байесовский критерий (3.7), в сущности, сводится к локальному критерию (3.5) с большим количеством участвующих конкурирующих моделей. Заметим, что в подавляющем большинстве интересных с практической точки зрения случаев аналитическое вычисление интегралов по Θ_i в формуле (3.7) не представляется возможным. Вместо этого естественно использовать процедуру численного интегрирования на конечной сетке $\lambda_{i,1}, \dots, \lambda_{i,\ell_i}$ значений из Θ_i , что соответствует случаю, когда априорные распределения $\mathcal{P}_i$ дискретны:

$$\mathcal{P}_i = \begin{bmatrix} \lambda_{i,1} & \dots & \lambda_{i,\ell_i} \\ \tau_{i,1} & \dots & \tau_{i,\ell_i} \end{bmatrix}, \quad i = 1, \dots, \nu.$$

Тогда критерий (3.7) переписывается в виде

$$T_P^B(\xi) = \sum_{i,j=1}^{\nu} \sum_{k=1}^{\ell_i} p_{i,j} \tau_{i,k} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} [\eta_i(x, \lambda_{i,k}) - \eta_j(x, \theta_{i,j})]^2 \xi(dx), \quad (3.8)$$

откуда видно, что это есть локальный T_P -критерий вида (3.5) с той лишь разницей, что критерий, полученный с помощью байесовского подхода, предусматривает гораздо большее количество сравнений между различными функциями η_i и η_j , чем критерий (3.5). В качестве иллюстрации рассмотрим простейшую ситуацию, когда мы применяем байесовский подход в случае двух конкурирующих моделей. Байесовский T -критерий для дискриминации двух моделей имеет вид (3.6), а его дискретная версия записывается как

$$T_{12}^B(\xi) = \sum_{k=1}^{\ell} \tau_k \inf_{\theta_{2,k} \in \Theta_2} \int_{\mathcal{X}} [\eta_1(x, \lambda_k) - \eta_2(x, \theta_{2,k})]^2 \xi(dx). \quad (3.9)$$

Критерий (3.9) — это в точности критерий (3.5) при $\nu = \ell + 1$, $p_{i,\ell+1} = \tau_i$ ($i = 1, \dots, \ell$) и $p_{i,j} = 0$ во всех других случаях. То есть вместо того, чтобы совершить одно сравнение, как в случае с локальным критерием Т-оптимальности (3.3), при использовании байесовского подхода (с дискретным априорным распределением $\mathcal{P}_1$) необходимо совершить ℓ сравнений, где ℓ — это число опорных точек априорного распределения $\mathcal{P}_1$. Заметим, что здесь для каждой точки $\lambda_{i,k}$ из априорных распределений в критерии (3.8) (или для каждой точки λ_k в критерии (3.9)) необходимо численно найти $\inf_{\theta_{i,j} \in \Theta_j}$. Это обстоятельство делает

задачу чрезвычайно трудной с вычислительной точки зрения. В последующих параграфах мы опишем эффективные численные процедуры для нахождения планов, оптимальных с точки зрения критерия (3.9).

3.3 Теорема эквивалентности для T_P -оптимальных планов

В прошлом параграфе мы показали, что изучение байесовских T_P -оптимальных планов сводится к изучению обычных T_P -оптимальных планов в случае дискретных априорных распределений. В этом параграфе рассмотрим некоторые свойства T_P -оптимальных планов и сформулируем теорему эквивалентности для них.

Пусть имеет место

Предположение 4. Для любого $i = 1, \dots, \nu$ функция $\eta_i(x, \theta_i)$ непрерывно дифференцируема по параметрам $\theta_i \in \Theta_i$.

Пусть ξ есть произвольный план на $\mathcal{X}$. Введем обозначение

$$\Theta_{i,j}^*(\xi) = \arg \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \xi(dx) \quad (3.10)$$

Теорема 6. Пусть выполнено предположение 4. План ξ^* является локальным T_P -оптимальным тогда и только тогда, когда существуют распределения μ_{ij}^* на множествах $\Theta_{i,j}^*(\xi^*)$, которые определены в (3.10), такие что неравенство

$$\sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi^*)} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \mu_{ij}^*(d\theta_{i,j}) \leq T_P(\xi^*) \quad (3.11)$$

выполнено для всех $x \in \mathcal{X}$. Более того, если ξ^* — локальный T_P -оптимальный план, то для его опорных точек в (3.11) достигается равенство.

Теорема 6 является обобщением соответствующей теоремы из [13] и доказывается аналогично. Для планов ξ, ζ на $\mathcal{X}$ введем функцию

$$Q(\zeta, \xi) = \int_{\mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_{i,j}^*(\xi)} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \zeta(dx), \quad (3.12)$$

где ζ — это план эксперимента и множество $\Theta_{ij}^*(\xi)$ определяется в (3.10). Используя лемму 1 из главы 1 нетрудно проверить, что

$$\left. \frac{\partial T_P(\xi(\alpha))}{\partial \alpha} \right|_{\alpha=0} = Q(\zeta, \xi) - T_P(\xi)$$

где $\xi(\alpha) = (1 - \alpha)\xi + \alpha\zeta$ обозначает выпуклую комбинацию планов ξ и ζ .

Теорема 7. Пусть выполнено предположение 4. Если план ξ не является T_P -оптимальным, тогда существует план ζ^* , такой что выполнено неравенство

$$Q(\zeta^*, \xi) > T_P(\xi).$$

Доказательство теоремы 7 можно найти в работе [15].

Для того, чтобы получить более удобное представление для этого результата, обозначим за $\hat{\mu}_{i,j}(\xi)$ меру на множестве $\Theta_{i,j}^*(\xi)$ ($i, j = 1, \dots, \nu$), для которой функция

$$\max_{x \in \mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi^*)} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \mu_{i,j}(d\theta_{i,j})$$

принимает минимальное значение. Определим также функцию

$$\Psi(x, \xi) = \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi)} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \hat{\mu}_{i,j}(d\theta_{i,j}). \quad (3.13)$$

Теорема 8. Пусть выполнено предположение 4. Если план ξ не является T_P -оптимальным, тогда существует точка $\bar{x} \in \mathcal{X}$, такая что

$$\Psi(\bar{x}, \xi) > T_P(\xi).$$

Доказательство теоремы 8 также можно найти в [15]. Рассмотрим

Предположение 5. Для любых i и j , таких что $p_{i,j} \neq 0$, и для любого плана ξ , для которого $T_P(\xi) > 0$, $\inf_{\theta_{i,j} \in \Theta_j}$ в (3.5) достигается в единственной точке $\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\xi)$ во внутренней области множества Θ_j .

Если кроме предположения 4 выполнено также предположение 5, то функция

$$Q(\zeta, \xi) = \int_{\mathcal{X}} \sum_{i,j=1}^{\nu} p_{i,j} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \hat{\theta}_{i,j})]^2 \zeta(dx).$$

Более того,

$$\Psi(x, \xi) = \sum_{i,j=1}^{\nu} p_{i,j} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \hat{\theta}_{i,j})]^2. \quad (3.14)$$

3.4 Численные алгоритмы для построения T_P -оптимальных планов

В работе [13] был предложен численный алгоритм для построения локальных T_P -оптимальных планов, основанный на многомерной чебышевской аппроксимации. Этот алгоритм достаточно сложен как с точки зрения описания, так и с точки зрения реализации, поэтому в этом параграфе он не приводится. Также он требует существенных вычислительных ресурсов (смотри в [13] параграф 7 с численными результатами), из-за чего применять его можно только в случае небольшого количества попарных сравнений между конкурирующими моделями в критерии (3.5). В этом параграфе мы опишем более эффективный метод, способный вычислить план даже в случае большего числа попарных сравнений. Как уже отмечалось в параграфе 3.2, этот алгоритм также можно будет использовать для построения байесовских планов при дискретных априорных распределениях.

Если выполнено предположение 4, то по теореме 8 существует точка $\bar{x} \in \mathcal{X}$, такая что

$$\Psi(\bar{x}, \xi) > T_P(\xi),$$

для любого плана ξ , который не является локальным T_P -оптимальным, где функция $\Psi(x, \xi)$ определяется в (3.13), а $T_P(\xi)$ в (3.5). Алгоритм, предложенный в [6], использует похожее свойство для того, чтобы построить последовательность планов, сходящуюся к T -оптимальному. Приведем аналог этого алгоритма для случая T_P -оптимальных планов.

Алгоритм 2 (аналог алгоритма из [6]). Пусть ξ_0 — некоторый начальный план и $(\alpha_s)_{s=0}^{\infty}$ — последовательность положительных вещественных чисел,

удовлетворяющая условиям $\lim_{s \rightarrow \infty} \alpha_s = 0$, $\sum_{s=0}^{\infty} \alpha_s = \infty$, $\sum_{s=0}^{\infty} \alpha_s^2 < \infty$. Тогда для $s = 0, 1, \dots$ поочередно выполняем два шага:

1. Находим $x_{s+1} = \arg \max_{x \in \mathcal{X}} \Psi(x, \xi_s)$,
2. Берем $\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi(x_{s+1})$.

Как уже говорилось ранее, условие остановки для всех алгоритмов в работе: достижение планом ξ_{s+1} заранее заданной нижней границы эффективности, получаемой из теоремы эквивалентности. Относительно сходимости этого алгоритма известен следующий результат:

Теорема 9. Пусть выполнено предположение 4 и пусть $\{\xi_s\}_{s=0,1,\dots}$ — это последовательность планов, получаемая с помощью алгоритма 2. Тогда

$$\lim_{s \rightarrow \infty} T_P(\xi_s) = T_P(\xi^*),$$

где ξ^* — локальный T_P -оптимальный план.

Доказательство этой теоремы аналогично доказательству теоремы 3 из главы 1.

Основная проблема алгоритма 2 состоит в том, что он порождает последовательность планов ξ_s с постоянно возрастающим количеством опорных точек, поэтому и финальный план, получаемый после остановки, имеет очень большой носитель. Проблему для финального плана можно решить с помощью кластеризации или через нахождение экстремумов функции $\Psi(x, \xi_m)$, где m — это номер итерации, на которой произошла остановка, но с увеличением числа точек в носителе ξ_s во время итерационного процесса бороться гораздо сложнее. Для достижения финальным планом нужной эффективности может потребоваться несколько сотен итераций алгоритма, что влечет большую сложность при вычислении оптимальных значений

$$\hat{\theta}_{i,j} = \arg \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 \xi(dx) \quad (3.15)$$

в критерии (3.5). Также в работе [13] было показано, что алгоритм 2 может не сходиться в том случае, если в критерии (3.5) слишком много попарных сравнений.

Далее предлагается альтернативный алгоритм, который также подразумевает последовательное выполнение двух шагов: раздельной максимизации

критерия по опорным точкам (шаг 1) и по весам (шаг 2). Первый шаг не представляет затруднений и подробно мы его описывать не будем. Для второго шага предложены несколько вариантов оптимизационных процедур, которые мы опишем ниже по тексту в параграфах 3.4.1 и 3.4.2.

Алгоритм 3. Пусть ξ_0 — некоторый начальный план. Тогда для $s = 0, 1, \dots$ поочередно выполняем два шага:

1. Пусть $\mathcal{S}_{[s]}$ — носитель плана ξ_s . Найдем множество $\mathcal{E}_{[s]}$ всех локальных максимумов функции $\Psi(x, \xi_s)$ на $\mathcal{X}$ и определим $\mathcal{S}_{[s+1]} = \mathcal{S}_{[s]} \cup \mathcal{E}_{[s]}$.
2. Определим $\xi = \{\mathcal{S}_{[s+1]}, \omega\}$ как план с носителем $\mathcal{S}_{[s+1]}$ и вектором весов ω . Теперь вычислим локальный TP -оптимальный план в классе всех планов с носителем $\mathcal{S}_{[s+1]}$, то есть найдем такой вектор $\omega_{[s+1]}$, который доставляет максимум

$$\begin{aligned} g(\omega) &= TP(\{\mathcal{S}_{[s+1]}, \omega\}, \bar{\theta}_1, \dots, \bar{\theta}_\nu) \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \sum_{x \in \mathcal{S}_{[s+1]}} [\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j})]^2 w_x, \end{aligned}$$

где w_x — это вес в точке $x \in \mathcal{S}_{[s+1]}$. Уберем из множества $\mathcal{S}_{[s+1]}$ те точки, которым в полученном оптимальном векторе весов $\omega_{[s+1]}$ соответствуют нули, и получим новый носитель, который также обозначим $\mathcal{S}_{[s+1]}$. Наконец, определим план ξ_{s+1} как план с носителем $\mathcal{S}_{[s+1]}$ и соответствующими ненулевыми весами из $\omega_{[s+1]}$.

На первом шаге каждой итерации все локальные максимумы функции $\Psi(x, \xi_s)$ добавляются к множеству возможных опорных точек плана ξ_{s+1} . Накапливания точек в носителе не происходит благодаря следующему важному свойству функции $\Psi(x, \xi_{s+1})$:

Лемма 3. Пусть выполнены предположения 4 и 5. В конце каждой итерации алгоритма 3 функция $\Psi(x, \xi_{s+1})$ принимает одно и то же значение во всех опорных точках плана ξ_{s+1} .

Доказательство леммы 3 можно найти в [15]. Из этой леммы сразу следует

Теорема 10. Пусть выполнены предположения 4 и 5. Тогда в конце каждой итерации алгоритма 3 число точек в носителе $\mathcal{S}_{[s+1]}$ плана ξ_{s+1} не превосхо-

дит максимального числа корней уравнения

$$\Psi(x, \xi_{s+1}) = c, \quad c \in [0, \max_{x \in \mathcal{X}} \Psi(x, \xi_{s+1})].$$

В программной реализации на каждом шаге из носителя промежуточного плана исключаются те точки, итоговый вес которых после решения оптимизационной задачи по весам w меньше, чем $m^{0.25}$, где $m = 2.2 \times 10^{-16}$ — самое маленькое число с плавающей запятой в $\mathbb{R}$.

Для алгоритма 3 верен следующий результат о сходимости:

Теорема 11. Пусть выполнено предположение 4 и пусть $\{\xi_s\}_{s=0,1,\dots}$ — это последовательность планов, получаемая с помощью алгоритма 3. Тогда

$$\lim_{s \rightarrow \infty} T_P(\xi_{s+1}) = T_P(\xi^*),$$

где ξ^* — локальный T_P -оптимальный план.

Доказательство. Очевидно, что неравенство

$$T_P(\{\mathcal{S}_{[s]}, \omega_{[s]}\}) \leq T_P(\{\mathcal{S}_{[s+1]}, \omega_{[s+1]}\})$$

выполнено для всех s , так как $\mathcal{S}_{[s]} \subseteq \mathcal{S}_{[s+1]}$, то есть последовательность $T_P(\xi_s)$, $s = 0, 1, \dots$ — неубывающая. Более того, эта последовательность также ограничена сверху величиной $T_P(\xi^*)$, а значит она имеет предел. Обозначим этот предел как T_P^{**} . Существует подпоследовательность планов ξ_{s_j} , $j = 1, 2, \dots$, которая сходится к некоторому плану ξ^{**} . Функция T_P является полу-непрерывной сверху как инфимум непрерывной функции, поэтому $T_P(\xi^{**}) = T_P^{**}$. Теперь предположим, что

$$T_P(\xi^{**}) < T_P(\xi^*).$$

Тогда ξ^{**} не является локальным T_P -оптимальным планом и по теореме 7 существует константа $\delta > 0$, такая что

$$\sup_{\zeta} Q(\zeta, \xi^{**}) - T_P(\xi^{**}) = 2\delta,$$

где функция Q определяется в формуле (3.12). Поэтому при достаточно больших j , таких что $j \geq N$ для некоторого заранее фиксированного N , получаем (используя тот факт, что функция $\sup_{\zeta} Q(\zeta, \xi)$ является полунепрерывной снизу)

$$\sup_{\zeta} Q(\zeta, \xi_{s_j}) - T_P(\xi_{s_j}) > \delta,$$

при всех $j \geq N$. Так как по построению последовательность $T_P(\xi_s)$, $s = 0, 1, \dots$ не убывает, верно неравенство

$$T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j}) \geq T_P(\xi_{s_j+1}) - T_P(\xi_{s_j}). \quad (3.16)$$

Чтобы оценить правую часть этого неравенства, определим при $j \geq N$ и $\alpha \in [0, 1]$ план

$$\tilde{\xi}_{s_{j+1}}(\alpha) = (1 - \alpha)\xi_{s_j} + \alpha\zeta_j,$$

где ζ_j — это мера, на которой функция $Q(\zeta, \xi_{s_j})$ достигает своего максимального значения на множестве всех планов, сосредоточенных в точках локальных максимумов функции $\Psi(x, \xi_{s_j})$, и

$$\alpha_{s_{j+1}} = \arg \max_{0 \leq \alpha \leq 1} T_P(\tilde{\xi}_{s_{j+1}}(\alpha)).$$

По построению план $\tilde{\xi}_{s_{j+1}}$ — это лучший план, сосредоточенный на множестве точек $\text{supp}(\xi_{s_j}) \cup \text{supp}(\zeta_j)$, поэтому

$$T_P(\xi_{s_{(j+1)}}) \geq T_P(\xi_{s_j+1}) \geq T_P(\tilde{\xi}_{s_{j+1}}(\alpha_{s_{j+1}})). \quad (3.17)$$

Введем обозначение

$$h(j, \alpha) = T_P(\tilde{\xi}_{s_j}(\alpha)),$$

и заметим, что

$$\left. \frac{\partial T_P(\tilde{\xi}_{s_{j+1}}(\alpha))}{\partial \alpha} \right|_{\alpha=0} = Q(\zeta_j, \xi_{s_j}) - T_P(\xi_{s_j}) = \sup_{\zeta} Q(\zeta, \xi_{s_j}) - T_P(\xi_{s_j}) > \delta.$$

Теперь, используя разложение Тейлора, получаем

$$\begin{aligned}
h(j+1, \alpha_{s_{j+1}}) - h(j+1, 0) &= \max_{\alpha \in [0,1]} [T_P(\tilde{\xi}_{s_{j+1}}(\alpha)) - T_P(\tilde{\xi}_{s_{j+1}}(0))] \\
&\geq \max_{\alpha \in [0,1]} \left[\alpha \frac{\partial T_P(\tilde{\xi}_{s_{j+1}}(\alpha))}{\partial \alpha} \Big|_{\alpha=0} - \frac{1}{2} \alpha^2 K \right] \\
&> \max_{\alpha \in [0,1]} \left[\alpha \delta - \frac{1}{2} \alpha^2 K \right] = \frac{\delta^2}{2K},
\end{aligned}$$

где K — это верхняя оценка для второй производной. Поэтому из уравнения (3.17) получаем

$$\begin{aligned}
T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j}) &\geq T_P(\xi_{s_{j+1}}) - T_P(\xi_{s_j}) \\
&\geq T_P(\tilde{\xi}_{s_{j+1}}) - T_P(\xi_{s_j}) \\
&= h(j+1, \alpha_{s_{j+1}}) - h(j+1, 0) \geq \frac{\delta^2}{2K}.
\end{aligned}$$

Теперь просуммируем последнее выражение по j от N до $L-1 > N$ при $L > N+1$ и получим

$$\sum_{j=N}^{L-1} [T_P(\xi_{s_{(j+1)}}) - T_P(\xi_{s_j})] = T_P(\xi_{s_{(L)}}) - T_P(\xi_{s_N}) \geq [L-N] \frac{\delta^2}{2K}.$$

Левая часть последнего неравенства стремится к конечному значению $T(\xi^{**}) - T(\xi_{s_N})$, а правая часть стремится к бесконечности при $L \rightarrow \infty$. Мы пришли к противоречию, а значит наше исходное предположение о том, что $T_P(\xi^{**}) < T_P(\xi^*)$ неверно, что завершает доказательство теоремы 11. $\square$

С практической точки зрения эффективное решение оптимизационной задачи на втором шаге алгоритма 3, то есть нахождение максимума функции $g(\omega)$, является самым важным. Далее описаны два метода оптимизации по весам. Будем считать, что выполнены предположения 4 и 5.

3.4.1 Квадратичное программирование

Пусть $\mathcal{S}_{[s+1]} = \{x_1, \dots, x_n\}$ обозначает множество, получаемое на первом шаге s -й итерации алгоритма 3. Определим ξ как план с носителем $\mathcal{S}_{[s+1]}$ и весами $\omega_1, \dots, \omega_n$, которые и предстоит определить на втором шаге итерации

алгоритма посредством максимизации функции

$$g(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \sum_{k=1}^n \omega_k [\eta_i(x_k, \bar{\theta}_i) - \eta_j(x_k, \hat{\theta}_{i,j})]^2,$$

где $\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\omega)$ определяются в (3.15). Для этой цели предлагается линеаризовать функции $\eta_j(x_k, \theta_{i,j})$ в окрестности точек $\hat{\theta}_{i,j}$. Рассмотрим функцию

$$\begin{aligned} \bar{g}(\omega) &= \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j} \in \mathbb{R}^{d_j}} \sum_{k=1}^n \omega_k \left[\eta_i(x_k, \bar{\theta}_i) - \eta_j(x_k, \hat{\theta}_{i,j}) - \alpha_{i,j}^T \frac{\partial \eta_j(x_k, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j} = \hat{\theta}_{i,j}} \right]^2 \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j} \in \mathbb{R}^{d_j}} \left[\alpha_{i,j}^T \mathbf{J}_{i,j}^T \boldsymbol{\Omega} \mathbf{J}_{i,j} \alpha_{i,j} - 2\omega^T \mathbf{R}_{i,j} \alpha_{i,j} + b_{i,j}^T \omega \right], \end{aligned}$$

где d_j — это размерность множества возможных параметров j -й модели Θ_j ,

$$\boldsymbol{\Omega} = \text{diag}(\omega_1, \dots, \omega_n),$$

а матрицы $\mathbf{J}_{i,j} \in \mathbb{R}^{n \times d_j}$, $\mathbf{R}_{i,j} \in \mathbb{R}^{n \times d_j}$ и векторы $b_{i,j} \in \mathbb{R}^n$ задаются как

$$\begin{aligned} \mathbf{J}_{i,j} &= \left(\frac{\partial \eta_j(x_r, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j} = \hat{\theta}_{i,j}} \right)_{r=1, \dots, n}, \\ \mathbf{R}_{i,j} &= \left([\eta_i(x_r, \bar{\theta}_i) - \eta_j(x_r, \hat{\theta}_{i,j})] \frac{\partial \eta_j(x_r, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j} = \hat{\theta}_{i,j}} \right)_{r=1, \dots, n}, \\ b_{i,j} &= \left([\eta_i(x_r, \bar{\theta}_i) - \eta_j(x_r, \hat{\theta}_{i,j})]^2 \right)_{r=1, \dots, n}, \end{aligned}$$

соответственно. Нетрудно видеть, что минимум по $\alpha_{i,j}$ достигается при

$$\alpha_{i,j} = (\mathbf{J}_{i,j}^T \boldsymbol{\Omega} \mathbf{J}_{i,j})^{-1} \mathbf{R}_{i,j}^T \omega,$$

откуда получаем, что

$$\bar{g}(\omega) = -\omega^T \mathbf{Q}(\omega) \omega + b^T \omega,$$

где

$$\mathbf{Q}(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \mathbf{R}_{i,j} (\mathbf{J}_{i,j}^T \boldsymbol{\Omega} \mathbf{J}_{i,j})^{-1} \mathbf{R}_{i,j}^T.$$

Матрица $\mathbf{Q}(\omega)$ зависит от ω , но если игнорировать эту зависимость и взять матрицу

$$\mathbf{\Omega} = \text{diag}(\bar{\omega}_1, \dots, \bar{\omega}_n)$$

фиксированной, то получим следующую задачу квадратичного программирования

$$\begin{aligned} \phi(\omega, \bar{\omega}) &= -\omega^T \mathbf{Q}(\bar{\omega}) \omega + b^T \omega \rightarrow \max_{\omega}, \\ \sum_{k=1}^n \omega_k &= 1; \quad \omega_k \geq 0, \quad k = 1, \dots, n. \end{aligned} \quad (3.18)$$

Решаем эту задачу итерационно до достижения сходимости, подставляя на каждой новой итерации решение, полученное на предыдущей итерации вместо $\bar{\omega}$. Похожая процедура предлагалась в работе [13].

Отметим, что получение оптимальных значений (3.15) требует в алгоритме 3 гораздо меньше вычислительных ресурсов, чем в алгоритме 2, так как количество опорных точек в промежуточном плане, который находится в ходе итерации, существенно меньше за счет удаления точек с нулевыми весами.

Предложенный способ вычисления весов можно также применить в алгоритме 2 вместо второго шага итерации. Вычислительные эксперименты показывают, что алгоритм 3, где на первом шаге итерации мы добавляем в множество опорных точек все локальные максимумы функции Ψ , а на втором шаге используем описанную выше процедуру, примерно в два раза быстрее, чем модификация алгоритма 2 с тем же вторым шагом, где на первом шаге мы добавляем к опорным точкам только глобальный максимум Ψ .

При практической реализации предложенной процедуры для шага 2 рекомендуется проводить только первые несколько итераций и прерывать процедуру, как только будет достигнута существенная разница между значениями критерия при начальном для шага 2 плане и плане, полученном в результате итерационного процесса. При этом общая скорость сходимости алгоритма 3 существенно возрастает, однако равенство значений функции Ψ для опорных точек вычисленного на шаге 2 плана (как указано в лемме 3) может и не достигаться.

3.4.2 Градиентный метод

Теперь рассмотрим альтернативную процедуру, представляющую из себя специализированный градиентный метод, которым можно воспользоваться на втором шаге алгоритма 3. Необходимо найти максимум по ω для функции

$$g(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \sum_{k=1}^n \omega_k [\eta_i(x_k, \bar{\theta}_i) - \eta_j(x_k, \hat{\theta}_{i,j})]^2, \quad (3.19)$$

где $\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\omega)$ определяются в (3.15). Введем функции

$$v_k(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} [\eta_i(x_k, \bar{\theta}_i) - \eta_j(x_k, \hat{\theta}_{i,j}(\omega))]^2, \quad k = 1, \dots, n,$$

и посчитаем последовательность векторов $(\omega_{(\gamma)})_{\gamma=0,1,\dots}$ по следующему алгоритму. Вначале возьмем какой-нибудь вектор $\omega_{(0)} = \bar{\omega}$ (например, с равными весами). Если $\omega_{(\gamma)} = (\omega_{(\gamma),1}, \dots, \omega_{(\gamma),n})$ задан, будем производить вычисления для $\gamma = 0, 1, \dots$ по следующей схеме. Найдем индексы $\bar{k}$ и $\underline{k}$, такие что

$$\bar{k} = \arg \max_{1 \leq k \leq n} v_k(\omega_{(\gamma)}), \quad \underline{k} = \arg \min_{1 \leq k \leq n} v_k(\omega_{(\gamma)}).$$

Обозначим

$$\alpha^* = \arg \max_{0 \leq \alpha \leq \omega_{(\gamma),\underline{k}}} g(\bar{\omega}_{(\gamma)}(\alpha)), \quad (3.20)$$

где вектор $\bar{\omega}_{(\gamma)}(\alpha) = (\bar{\omega}_{(\gamma),1}(\alpha), \dots, \bar{\omega}_{(\gamma),n}(\alpha))$ определяется как

$$\bar{\omega}_{(\gamma),i}(\alpha) = \begin{cases} \omega_{(\gamma),i} + \alpha, & \text{если } i = \bar{k}; \\ \omega_{(\gamma),i} - \alpha, & \text{если } i = \underline{k}; \\ \omega_{(\gamma),i}, & \text{иначе.} \end{cases}$$

Вектор $\omega_{(\gamma+1)}$ на следующей итерации определяется как $\omega_{(\gamma+1)} = \bar{\omega}_{(\gamma)}(\alpha^*)$. Следующая теорема показывает, что сгенерированная таким образом последовательность сходится к оптимальному значению ω^* , доставляющему максимум функции g из (3.19).

Теорема 12. *Последовательность $(\omega_{(\gamma)})_{\gamma \in \mathbb{N}}$ сходится к $\omega^* \in \arg \max g(\omega)$.*

Доказательство теоремы можно найти в [15].

Стоит отметить, что одномерная оптимизационная задача (3.20) является достаточно дорогой с вычислительной точки зрения. В программной реализации этого метода мы используем линеаризацию подобно той, что была описана в параграфе 3.4.1.

3.5 Примеры

Теперь проиллюстрируем предлагаемые алгоритмы с помощью двух примеров. Алгоритм 3 с двумя способами оптимизации по весам из параграфов 3.4.1 и 3.4.2 был реализован в качестве пакета [17] для языка *R*. Примеры из этого параграфа также можно найти в пакете. В программе пользователь задает начальный план ξ_0 , метод решения оптимизационной задачи на шаге 2, веса $p_{i,j}$ и априорную информацию о параметрах $\bar{\theta}_i$. Таблицу

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,\nu-1} & p_{1,\nu} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{\nu,1} & p_{\nu,2} & \cdots & p_{\nu,\nu-1} & p_{\nu,\nu} \end{bmatrix}$$

мы будем называть таблицей сравнения для задачи нахождения локального T_P -оптимального плана. Эта таблица должна быть задана экспериментатором. Так как задача нахождения байесовского T_P -оптимального плана в случае дискретных априорных распределений для параметров может быть сведена к задаче нахождения локального T_P -оптимального плана с большим числом конкурирующих моделей и попарных сравнений, мы опишем соответствующую таблицу сравнения для байесовского критерия. Для наглядности в первом примере мы рассмотрим случай двух конкурирующих моделей $\nu = 2$. Байесовский критерий T_P -оптимальности в этом случае задается формулой (3.9), где априорное распределение на Θ_1 имеет носитель $\lambda_1, \dots, \lambda_\ell$ и веса $\tau_1, \dots, \tau_\ell$. Этот критерий может быть переписан в виде

$$T_P(\xi) = \sum_{i,j=1}^{\ell+1} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} \left[\eta_i(x, \bar{\theta}_i) - \eta_j(x, \theta_{i,j}) \right]^2 \xi(dx), \quad (3.21)$$

где таблица сравнения задается как

$$P = (p_{i,j})_{i,j=1,\dots,\ell+1} = \begin{bmatrix} 0 & 0 & \dots & 0 & \tau_1 \\ 0 & 0 & \dots & 0 & \tau_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \tau_\ell \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{\ell+1 \times \ell+1}, \quad (3.22)$$

$\eta_i(x, \bar{\theta}_i) = \eta_1(x, \lambda_i)$, $i = 1, \dots, \ell$ и $\eta_{\ell+1}(x, \theta_{i,j}) = \eta_2(x, \theta_{i,\ell+1})$. Таким образом, байесовский критерий для дискретного априорного распределения предусматривает ℓ попарных сравнений вместо одного сравнения в случае локального критерия. Построение таблиц сравнения для случая дискриминации более чем двух моделей делается по аналогии.

Для всех рассматриваемых в этом параграфе регрессионных моделей выполнено предположение 4. Проверка предположения 5 оказывается проблематичной. Тем не менее в последующих примерах мы будем действовать так, как если бы оно было также выполнено. На практике мы не столкнулись с трудностями в связи с этим допущением.

3.5.1 Байесовские T_P -оптимальные планы для дискриминации экспоненциальных моделей

Рассмотрим задачу дискриминации между двумя конкурирующими экспоненциальными регрессионными моделями

$$\begin{aligned} \eta_1(x, \theta_1) &= \theta_{1,1} - \theta_{1,2} \exp(-\theta_{1,3} x^{\theta_{1,4}}), \\ \eta_2(x, \theta_2) &= \theta_{2,1} - \theta_{2,2} \exp(-\theta_{2,3} x), \end{aligned} \quad (3.23)$$

где множество планирования — это интервал $[0, 10]$. Экспоненциальные модели вида (3.23) широко используются в различных приложениях. Например, модель η_2 часто возникает в сельскохозяйственных исследованиях, где у нее есть специальное название (Mitscherlich growth law), при описании зависимости урожайности от количества использованного удобрения. В исследованиях в области рыболовства эта же модель носит название закона роста Берталанффи и используется при описании зависимости длины рыбы от ее возраста (смотри [33]). Оптимальные планы для экспоненциальных моделей обсуждались, например, в

работе [34]. Нетрудно показать, что T_P -оптимальный план для дискриминации моделей (3.23) не зависит от линейных параметров модели η_1 , поэтому можно выбрать их произвольными, например, $\bar{\theta}_{1,1} = 2$ и $\bar{\theta}_{1,2} = 1$. Для параметров $\bar{\theta}_{1,3}$ и $\bar{\theta}_{1,4}$ были взяты независимые априорные распределения, сосредоточенные в точках

$$\mu_j + \frac{\sigma(i-3)}{2} \quad i = 1, \dots, 5; \quad j = 3, 4, \quad (3.24)$$

где $\mu_3 = 0.8$, $\mu_4 = 1.5$. Веса в этих точках пропорциональны (в обоих случаях) величинам

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(i-3)^2}{8}\right); \quad i = 1, \dots, 5. \quad (3.25)$$

Отметим, что описанные условия соответствуют 25 попарным сравнениям в байесовском критерии (3.9). Байесовские T_P -оптимальные планы для дискриминации моделей (3.23), имеющие эффективность

$$\text{Eff}_{T_P}(\xi) = \frac{T_P(\xi)}{\sup_{\eta} T_P(\eta)} \geq 0.999,$$

были найдены тремя способами: 1) алгоритмом 3 со вторым шагом из § 3.4.1, 2) алгоритмом 3 со вторым шагом из § 3.4.2, и 3) алгоритмом 2. В таблице 4 представлены планы, полученные первым способом, при различных значениях параметра σ^2 , а в таблице 5 — сравнительное время работы в секундах до получения плана с нужной эффективностью для всех трех алгоритмов. В качестве начального плана во всех случаях был взят план, сосредоточенный в 11 точках $0, 1, \dots, 10$ с одинаковыми весами.

Вычисления проводились на персональном компьютере с центральным процессором intel core i7-4790K. Из таблицы 5 видно, что предложенные в настоящей главе алгоритмы значительно превосходят стандартный алгоритм по времени работы.

Отметим, что в случае небольшой дисперсии σ^2 , оптимальные планы имеют четыре опорные точки, а при $\sigma^2 \geq 0.285$ байесовские T_P -оптимальные планы сосредоточены в пяти точках. Графики функции Ψ из теоремы эквивалентности 6 показаны на рисунке 3.1 при разных значениях дисперсии σ^2 .

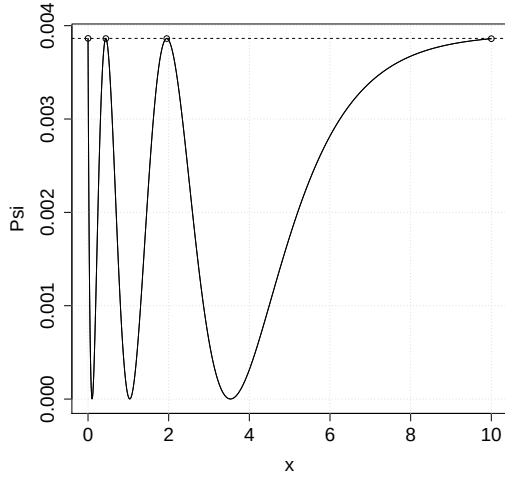
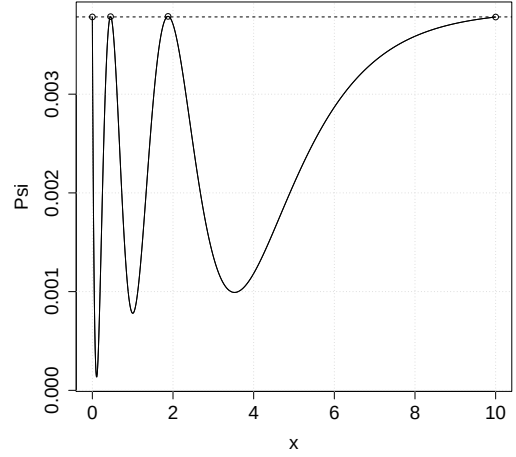
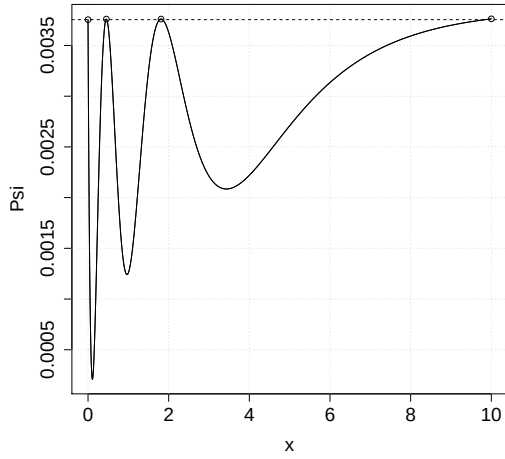
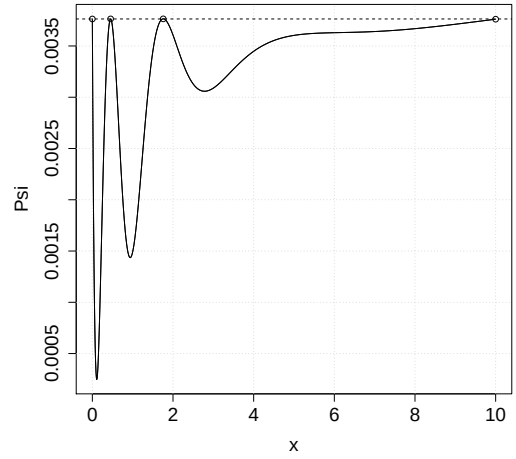
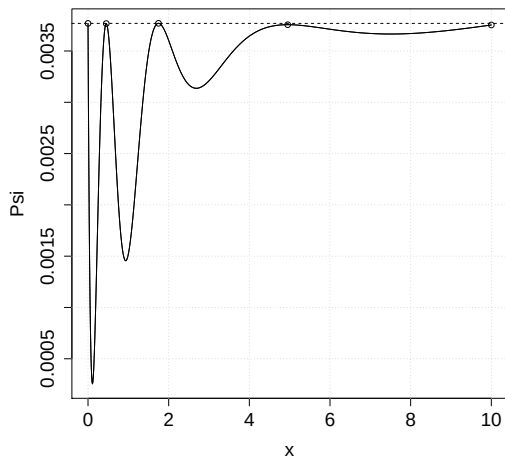
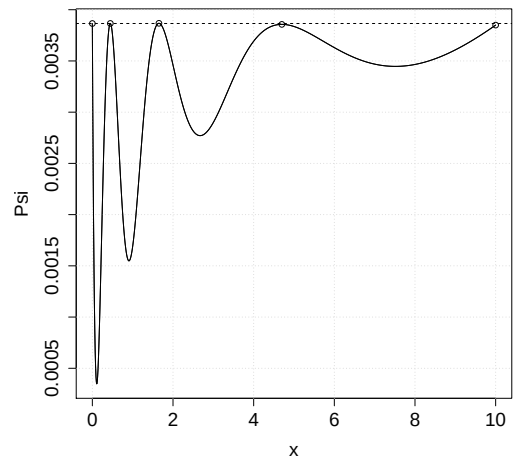
Таблица 4 — Байесовские T_P -оптимальные планы для дискриминации моделей (3.23)

σ^2	Оптимальный план $\xi_{\sigma^2}^*$				
0.0	0.000	0.441	1.952	10.000	
	0.209	0.385	0.291	0.115	
0.1	0.000	0.452	1.877	10.000	
	0.209	0.391	0.290	0.110	
0.2	0.000	0.455	1.811	10.000	
	0.208	0.394	0.291	0.107	
0.285	0.000	0.453	1.758	10.000	
	0.207	0.396	0.292	0.105	
0.3	0.000	0.452	1.747	4.951	10.000
	0.207	0.396	0.292	0.003	0.102
0.4	0.000	0.446	1.651	4.699	10.000
	0.200	0.384	0.290	0.060	0.066

Таблица 5 — Время работы различных алгоритмов в секундах

Пример	σ^2	Алг. 3(1)	Алг. 3(2)	АФ
(3.23)	0.0	0.03	0.11	12.4
	0.4	1.4	11.6	218.3
(3.26)	0.0	0.09	0.75	5.7
	37^2	7.8	37.1	762.3

Примечание — Алг. 3(1) и Алг. 3(2) — алгоритм 3 с использованием методов оптимизации по весам из параграфов 3.4.1 и 3.4.2 соответственно; АФ — алгоритм 2

а) $\sigma^2 = 0$ б) $\sigma^2 = 0.1$ в) $\sigma^2 = 0.2$ г) $\sigma^2 = 0.245$ д) $\sigma^2 = 0.3$ е) $\sigma^2 = 0.4$

Непрерывной линией обозначен график функции $\Psi(x, \xi_{\sigma^2}^*)$, задаваемой формулой (3.14), пунктирной — значение функционала $T_P(\xi_{\sigma^2}^*)$, определенного в (3.5), а выколотыми точками — опорные точки T_P -оптимального плана $\xi_{\sigma^2}^*$.

Рисунок 3.1 — Графики функций $\Psi(x, \xi_{\sigma^2}^*)$ для планов из таблицы 4

3.5.2 Байесовские T_P -оптимальные планы для дискриминации моделей зависимости эффективности препарата от дозы

Нелинейные регрессионные модели также часто используются при проведении исследований зависимости эффективности препарата от дозы. Например, для этих целей в работе [35] рассматривались следующие модели

$$\begin{aligned}\eta_1(x, \theta_1) &= \theta_{1,1} + \theta_{1,2}x; \\ \eta_2(x, \theta_2) &= \theta_{2,1} + \theta_{2,2}x(\theta_{2,3} - x); \\ \eta_3(x, \theta_3) &= \theta_{3,1} + \theta_{3,2}x/(\theta_{3,3} + x); \\ \eta_4(x, \theta_4) &= \theta_{4,1} + \theta_{4,2}/(1 + \exp(\theta_{4,3} - x)/\theta_{4,4});\end{aligned}\tag{3.26}$$

где множество планирования есть интервал $\mathcal{X} = [0, 500]$. В этой работе также приводились априорные значения для параметров этих моделей

$$\begin{aligned}\bar{\theta}_1 &= (60, 0.56), \\ \bar{\theta}_2 &= (60, 7/2250, 600), \\ \bar{\theta}_3 &= (60, 294, 25), \\ \bar{\theta}_4 &= (49.62, 290.51, 150, 45.51).\end{aligned}$$

Локальный T_P -оптимальный план для моделей (3.26) был получен в работе [13] в случае $p_{i,j} = 1/6$, ($1 \leq j < i \leq 4$), то есть критерий T_P -оптимальности (3.5) предусматривает в этом случае 6 попарных сравнений моделей.

Демонстрацию новых методов, предложенных в параграфе 3.4, мы начнем с вычисления того же локального T_P -оптимального плана для дискриминации моделей (3.26), что был представлен в работе [13]. Алгоритму 3 со вторым шагом из параграфа 3.4.1 требуется четыре итерации и 0.09 секунд для получения плана с эффективностью ≥ 0.999 . В качестве стартового использовался стартовый план из [13]. График функции $\Psi(x, \tilde{\xi}_1)$ после первой итерации алгоритма представлен на рисунке 3.2. Опорные точки плана $\tilde{\xi}_1$ показаны на этом рисунке выколотыми точками. Отметим, что во всех опорных точках $\tilde{\xi}_1$ функция $\Psi(x, \tilde{\xi}_1)$ принимает одно и то же значение, как и утверждается в лемме 3. Чтобы найти план с аналогичной эффективностью, алгоритму из статьи [13] понадобилось 9 итераций, каждая из которых по утверждению авторов выполняется за 1-2 секунды.

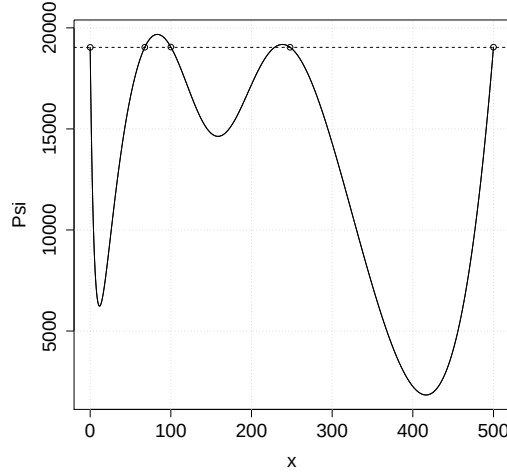


Рисунок 3.2 — График функции $\Psi(x, \tilde{\xi}_1)$ после первой итерации алгоритма 3

Теперь для тех же моделей рассмотрим байесовский T_P -оптимальный план. Для простоты мы задаем априорное распределение только для четырехмерного вектора параметров последней модели $\bar{\theta}_4$, в то время как $\bar{\theta}_1$, $\bar{\theta}_2$ и $\bar{\theta}_3$ остаются фиксированными векторами. Это априорное распределение задано в 81 точке из $\mathbb{R}^4$ следующим образом:

$$\{\mu_{e_1, e_2, e_3, e_4} \mid e_1, e_2, e_3, e_4 \in \{-1, 0, 1\}\}, \quad (3.27)$$

где

$$\begin{aligned} \mu_{e_1, e_2, e_3, e_4} &= (\mu_1 + e_1\sigma, \mu_2 + e_2\sigma, \mu_3 + e_3\sigma, \mu_4 + e_4\sigma), \\ \mu &= (\mu_1, \mu_2, \mu_3, \mu_4) = (49.62, 290.51, 150, 45.51). \end{aligned}$$

План, как и в прошлом примере, строится для разных значений σ^2 . Веса плана в соответствующих точках пропорциональны величинам

$$\frac{1}{(2\pi\sigma^2)^2} \exp\left(-\frac{\|\mu_{e_1, e_2, e_3, e_4} - \mu\|_2^2}{2\sigma^2}\right), \quad e_1, e_2, e_3, e_4 \in \{-1, 0, 1\}, \quad (3.28)$$

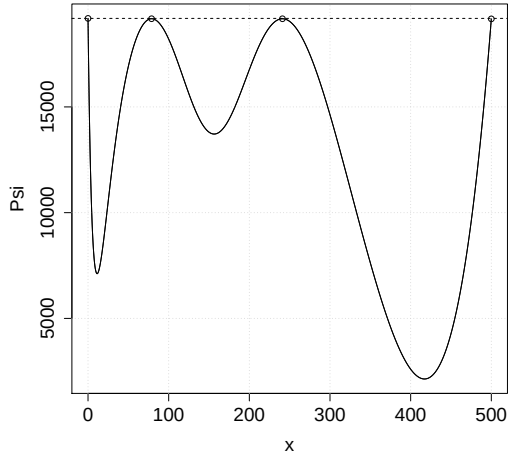
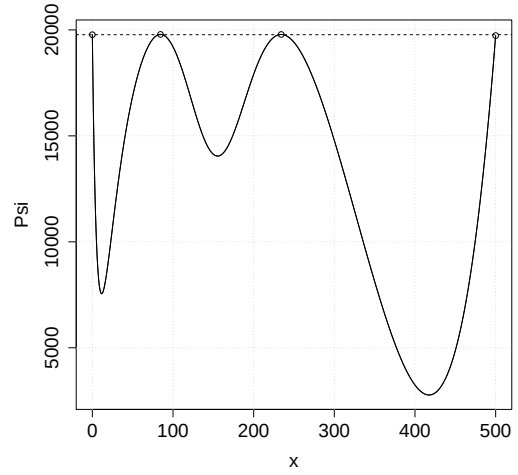
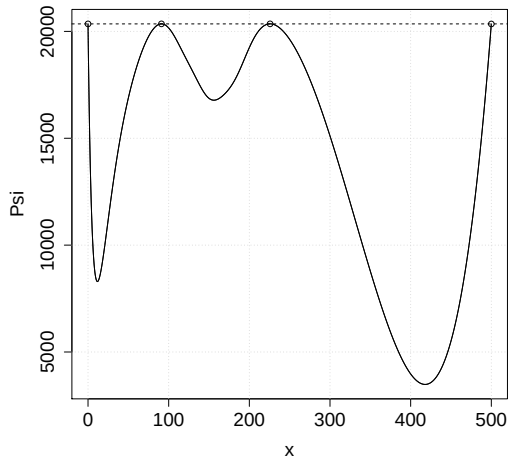
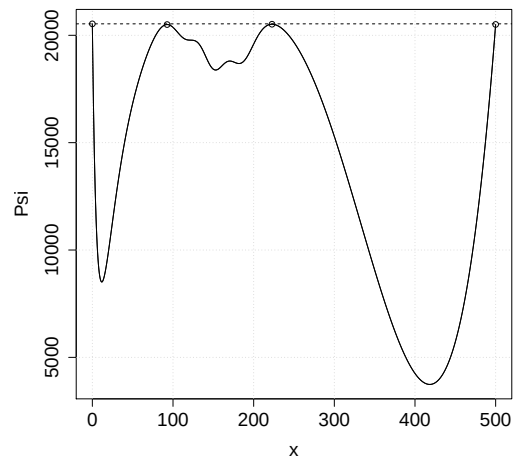
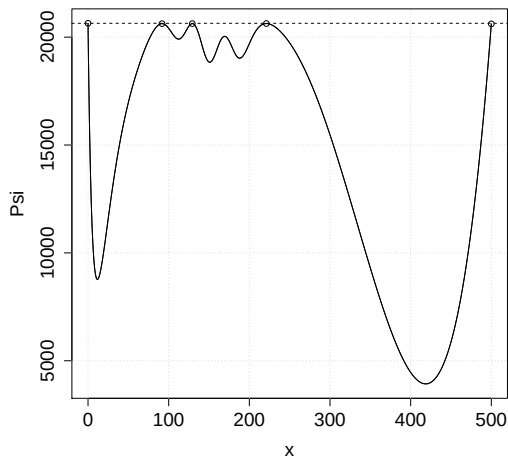
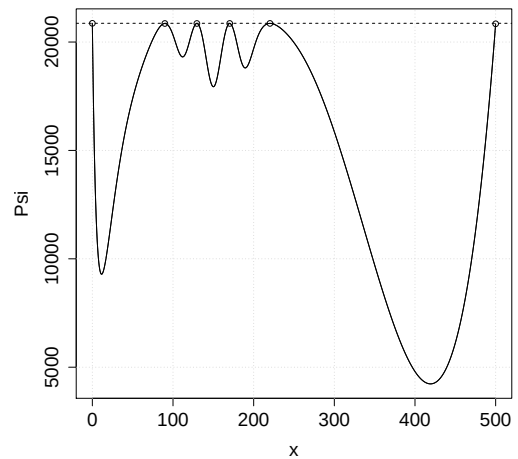
где $\|\cdot\|_2$ есть евклидова норма. Конечный байесовский критерий (3.8) состоит из 246 попарных сравнений конкурирующих моделей.

Для разных значений параметра σ^2 байесовский T_P -оптимальный план снова был найден тремя способами (как в предыдущем примере). Планы, найденные с помощью алгоритма 3 со вторым шагом из § 3.4.1, представлены в

Таблица 6 — Байесовские T_P -оптимальные планы для дискриминации моделей (3.26)

σ^2	Оптимальный план $\xi_{\sigma^2}^*$					
0	0.000	78.783	241.036	500.0		
	0.255	0.213	0.357	0.175		
20^2	0.000	84.467	234.134	500.0		
	0.257	0.225	0.351	0.167		
30^2	0.000	91.029	225.713	500.0		
	0.259	0.237	0.345	0.159		
33^2	0.000	92.692	222.735	500.0		
	0.260	0.240	0.344	0.156		
35^2	0.000	91.743	129.322	221.118	500.0	
	0.260	0.214	0.036	0.336	0.154	
37^2	0.000	89.881	129.590	170.306	220.191	500.0
	0.260	0.170	0.091	0.019	0.310	0.150

таблице 6. Сравнительное время работы для всех трех алгоритмов можно найти в таблице 5. Мы снова видим, что предложенные в главе методы работают существенно быстрее. При небольших значениях параметра σ^2 байесовский T_P -оптимальный план сосредоточен в 4 точках считая граничные точки множества планирования. При увеличении параметра σ^2 количество опорных точек растет. Так, если $\sigma^2 = 35^2$ или 37^2 , байесовский T_P -оптимальный план имеет 5 или 6 точек, считая граничные точки множества планирования. Графики функции Ψ из теоремы эквивалентности 6 приведены на рисунке 3.3.

а) $\sigma^2 = 0$ б) $\sigma^2 = 20^2$ в) $\sigma^2 = 30^2$ г) $\sigma^2 = 33^2$ д) $\sigma^2 = 35^2$ е) $\sigma^2 = 37^2$

Непрерывной линией обозначен график функции $\Psi(x, \xi_{\sigma^2}^*)$, задаваемой формулой (3.14), пунктирной — значение функционала $T_P(\xi_{\sigma^2}^*)$, определенного в (3.5), а выколотыми точками — опорные точки T_P -оптимального плана $\xi_{\sigma^2}^*$.

Рисунок 3.3 — Графики функций $\Psi(x, \xi_{\sigma^2}^*)$ для планов из таблицы 6

Глава 4. Построение байесовских KL_P -оптимальных планов

Критерий T -оптимальности (как и обобщающий его критерий T_P -оптимальности) предполагает гомоскедастичность и нормальную распределенность случайных ошибок наблюдения. На случай дискриминации двух моделей при гетероскедастичных нормальных ошибках T -критерий был обобщен в работе [9]. В [10] был предложен критерий KL -оптимальности, основанный на расстояниях Кульбака–Лейблера, не требующий нормальности и пригодный для дискриминации двух произвольных конкурирующих моделей для плотностей ошибок, причем критерии из работ [6] и [9] являются его частными случаями.

В этой главе предложен метод построения байесовских KL_P -оптимальных планов для дискриминации произвольного количества регрессионных моделей. Байесовский критерий KL_P -оптимальности вводится по аналогии с байесовским критерием T_P -оптимальности. Численные алгоритмы из главы 3 обобщаются на случай нового критерия. Изложение в данной главе опирается на материал из работы [16]. Алгоритмы, описанные в параграфе 4.4, также реализованы в пакете [17] для языка R.

4.1 Обобщение критерия T -оптимальности на случай произвольно распределенных ошибок наблюдения

Рассмотрим стандартное уравнение регрессии:

$$Y = \eta(x, \theta) + \varepsilon, \quad (4.1)$$

где $x \in \mathcal{X}$, а случайные ошибки ε независимы и имеют нулевое среднее и заданную дисперсию. В работе [10] было предложено следующее обобщение критерия T -оптимальности для случая произвольно распределенных ошибок наблюдения. Будем считать, что при фиксированных x и θ величина Y имеет абсолютно непрерывное распределение с плотностью $f(y, x, \theta)$, среднее и дисперсия которой равны

$$\begin{aligned} \eta(x, \theta) &= \mathbb{E}Y = \int_{\mathcal{X}} y f(y, x, \theta) dy, \\ v^2(x, \theta) &= \text{var}Y = \int_{\mathcal{X}} (y - \mathbb{E}Y)^2 f(y, x, \theta) dy. \end{aligned}$$

Рассмотрим теперь две конкурирующие модели для плотности $f(y, x, \theta)$:

$$\begin{aligned} f_1(y, x, \theta_1), & \quad y \in \mathcal{Y}, x \in \mathcal{X}, \theta_1 = (\theta_{1,1}, \dots, \theta_{1,d_1})^T \in \Theta_1; \\ f_2(y, x, \theta_2), & \quad y \in \mathcal{Y}, x \in \mathcal{X}, \theta_2 = (\theta_{2,1}, \dots, \theta_{2,d_2})^T \in \Theta_2, \end{aligned}$$

где $\mathcal{Y}$ — общая область определения для обеих плотностей. Расстояние Кульбака–Лейблера между конкурирующими плотностями $f_1(y, x, \theta_1)$ и $f_2(y, x, \theta_2)$ обозначим как

$$I_{1,2}(x, \theta_1, \theta_2) = \int_{\mathcal{Y}} f_2(y, x, \theta_2) \log \frac{f_2(y, x, \theta_2)}{f_1(y, x, \theta_1)} dy. \quad (4.2)$$

Предположим теперь, что параметры первой модели фиксированы и обозначим их как $\bar{\theta}_1$. Для дискриминации между двумя конкурирующими плотностями $f_1(y, x, \theta_1)$ и $f_2(y, x, \theta_2)$ в [10] было предложено искать такой план эксперимента, который максимизирует величину

$$\text{KL}_{1,2}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, \bar{\theta}_1, \theta_2) \xi(dx). \quad (4.3)$$

План, максимизирующий критерий (4.3), называется KL -оптимальным.

Отметим, что критерий (4.3) не является симметричным. По сути, поменяв f_1 и f_2 в (4.2) местами или заменив $\inf_{\theta_2 \in \Theta_2}$ на $\inf_{\theta_1 \in \Theta_1}$ в (4.3), мы получим другие критерии (всего четыре разных варианта).

Статистическая интерпретация критерия KL -оптимальности такова. Рассмотрим задачу проверки гипотезы

$$H_0 : f(x, \theta) = f_1(x, \bar{\theta}_1),$$

против альтернативы

$$H_1 : f(x, \theta) = f_2(x, \theta_2), \quad \theta \in \Theta_2.$$

Отношение правдоподобия для наблюдения y в точке x имеет следующий вид:

$$L(\theta_2, \xi_x) = \frac{f_1(y, x, \bar{\theta}_1)}{f_2(y, x, \theta_2)}.$$

Тест отношения правдоподобия основан на статистике

$$R(\theta_2, \xi_x) = -2 \log L(\theta_2, \xi_x) = 2 \log \frac{f_2(y, x, \theta_2)}{f_1(y, x, \bar{\theta}_1)}.$$

При больших значениях статистики R гипотеза H_0 отвергается. Рассмотрим математическое ожидание статистики R при том, что верна гипотеза H_1 , в случае плана, сосредоточенного в единственной точке x :

$$\mathbb{E}_{H_1} R(\theta_2, \xi_x) = 2 \int f_2(y, x, \theta_2) \log \frac{f_2(y, x, \theta_2)}{f_1(y, x, \bar{\theta}_1)} = 2I_{1,2}(x, \bar{\theta}_1, \theta_2) dy.$$

Мощность теста отношения правдоподобия тем больше, чем больше значение $\mathbb{E}_{H_1} R(\theta_2, \xi_x)$. Теперь рассмотрим точный план $\xi_N = (x_1, \dots, x_n; q_1, \dots, q_n)$, такой что $q_i = r_i/N$, $\sum_{i=1}^n r_i = N$. Используя независимость наблюдений получаем

$$\begin{aligned} \mathbb{E}_{H_1} R(\theta_2, \xi_N) &= 2 \int f_2(y, x_i, \theta_2) \log \prod_{i=1}^n \left\{ \frac{f_2(y, x_i, \theta_2)}{f_1(y, x_i, \bar{\theta}_1)} \right\}^{r_i} dy \\ &= 2N \sum_{i=1}^n q_i \int f_2(y, x_i, \theta_2) \log \frac{f_2(y, x_i, \theta_2)}{f_1(y, x_i, \bar{\theta}_1)} dy \\ &= 2N \int_{\mathcal{X}} I_{1,2}(x, \bar{\theta}_1, \theta_2) \xi_N(dx). \end{aligned}$$

Переходя от приближенного плана ξ_N к точному плану ξ и рассматривая значение $\mathbb{E}_{H_1} R(\theta_2, \xi)$ при наименее благоприятном θ_2 , когда $\mathbb{E}_{H_1} R(\theta_2, \xi)$ минимально, получаем критерий (4.3).

Теперь рассмотрим обобщение критерия KL -оптимальности на случай дискриминации произвольного количества конкурирующих плотностей

$$f_1(y, x, \theta_1), \dots, f_\nu(y, x, \theta_\nu).$$

Фиксированные параметры рассматриваемых моделей обозначим как $\bar{\theta}_1, \dots, \bar{\theta}_\nu$. Расстояние Кульбака–Лейблера между i -й и j -й плотностью обозначим как

$$I_{i,j}(x, \theta_i, \theta_j) = \int f_j(y, x, \theta_j) \log \frac{f_j(y, x, \theta_j)}{f_i(y, x, \theta_i)} dy,$$

а соответствующий критерий

$$\text{KL}_{i,j}(\xi, \bar{\theta}_i) = \inf_{\theta_j \in \Theta_j} \int_{\mathcal{X}} I_{i,j}(x, \bar{\theta}_i, \theta_j) \xi(dx).$$

Здесь параметры i -й плотности фиксируются, а по параметрам j -й плотности происходит оптимизация. Следуя обозначениям главы 3, пусть $p_{i,j}$ — это неотрицательные веса, отвечающие за вклад попарного сравнения плотностей $f_i(y, x, \theta_i)$ и $f_j(y, x, \theta_j)$ в финальном критерии. По аналогии с тем, как мы делали это в предыдущей главе для критерия T -оптимальности, введем симметричную версию критерия KL -оптимальности для $\nu \geq 2$ конкурирующих моделей $f_1(y, x, \theta_1), \dots, f_\nu(y, x, \theta_\nu)$:

$$\text{KL}_P(\xi) = \sum_{i,j=1}^{\nu} p_{i,j} \text{KL}_{i,j}(\xi, \bar{\theta}_i) = \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} I_{i,j}(x, \bar{\theta}_i, \theta_{i,j}) \xi(dx). \quad (4.4)$$

План, максимизирующий критерий (4.4), будем называть KL_P -оптимальным планом для дискриминации моделей $f_1(y, x, \theta_1), \dots, f_\nu(y, x, \theta_\nu)$. Будем далее всегда предполагать, что множества $\mathcal{X}$, Θ_i компактны и все плотности непрерывны по x . Отметим, что критерий (4.4) зависит от неизвестных на практике параметров $\bar{\theta}_1, \dots, \bar{\theta}_\nu$, которые устанавливаются экспериментатором для конкурирующих моделей $f_1(y, x, \theta_1), \dots, f_\nu(y, x, \theta_\nu)$. Следовательно, критерий является локальным в смысле работы [27]. Если фиксированные значения параметров $\bar{\theta}_1, \dots, \bar{\theta}_\nu$ далеки от истинных значений, найденный оптимальный план может иметь низкую эффективность.

4.2 Байесовские KL_P -оптимальные планы и их сведение к локально оптимальным

Аналогично тому, как мы делали это в предыдущей главе для T_P -оптимальных планов, сформулируем байесовскую версию критерия KL_P -оптимальности, являющуюся более устойчивой по отношению к неправильному выбору фиксированных параметров. Байесовским KL_P -оптимальным планом будем на-

зывать дискретную вероятностную меру, доставляющую максимум величине

$$\begin{aligned} \text{KL}_P^B(\xi) &= \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_i} \text{KL}_{i,j}(\xi, \theta_i) \mathcal{P}_i(d\theta_i), \\ &= \sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_i} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} I_{i,j}(x, \theta_i, \theta_{i,j}) \xi(dx) \mathcal{P}_i(d\theta_i) \end{aligned} \quad (4.5)$$

Здесь $p_{i,j} \geq 0$ и $p_{i,i} = 0$ при любом i , а мера $\mathcal{P}_i$ для каждого $i = 1, \dots, \nu$ задает априорное распределение для параметров θ_i плотности f_i , такое что все интегралы в формуле (4.5) определены.

Отметим, что для случая двух конкурирующих моделей критерий (4.5) рассматривался в работе [36]. Как было показано в [15], вычисление байесовских планов, оптимальных с точки зрения критерия (4.5), можно свести к вычислению локально оптимальных планов для большего количества конкурирующих моделей. В большинстве приложений интегралы в (4.5) вычисляются численными методами, предполагающими аппроксимацию произвольных априорных распределений мерами с конечным носителем. Если априорные распределения $\mathcal{P}_i$ дискретны и сосредоточены в точках $\lambda_{i1}, \dots, \lambda_{i\ell_i}$ с весами $\tau_{i1}, \dots, \tau_{i\ell_i}$, то критерий (4.5) можно переписать в виде

$$\text{KL}_P^B(\xi) = \sum_{i,j=1}^{\nu} \sum_{k=1}^{\ell_i} p_{i,j} \tau_{ik} \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} I_{i,j}(x, \lambda_{ik}, \theta_{i,j}) \xi(dx). \quad (4.6)$$

Критерий (4.6) — это локальный KL_P -оптимальный критерий вида (4.4), где конкурирующие модели задаются как $\{f_i(y, x, \lambda_{ik}) \mid k = 1, \dots, \ell_i; i = 1, \dots, \nu\}$. Единственная разница между критерием, полученным через дискретизацию байесовского подхода, и критерием (4.4) состоит в том, что из-за дискретизации априорных распределений $\mathcal{P}_1, \dots, \mathcal{P}_\nu$ итоговый критерий (4.6) предусматривает большее количество попарных сравнений конкурирующих моделей $f_i(y, x, \lambda_{ik})$. Поэтому нахождение байесовских KL -оптимальных планов представляет из себя вычислительно трудную задачу.

4.3 Теорема эквивалентности для KL_P -оптимальных планов

Пусть выполнено следующее

Предположение 6. Функции $f_1(y, x, \theta_1), \dots, f_\nu(y, x, \theta_\nu)$ непрерывно дифференцируемы по $\theta_i \in \Theta_i$.

Пусть ξ есть произвольный план на $\mathcal{X}$. Введем обозначение

$$\Theta_{i,j}^*(\xi) = \arg \inf_{\theta_{i,j} \in \Theta_j} \int_{\mathcal{X}} I_{i,j}(x, \bar{\theta}_i, \theta_{i,j}) \xi(dx). \quad (4.7)$$

Сформулируем теорему эквивалентности, которая дает необходимые и достаточные условия оптимальности для KL_P -оптимальности плана.

Теорема 13. Пусть выполнено предположение 6. План ξ^* является локальным KL_P -оптимальным тогда и только тогда, когда существуют распределения ρ_{ij}^* на множествах $\Theta_{i,j}^*(\xi^*)$, которые определены в (4.7), такие что неравенство

$$\sum_{i,j=1}^{\nu} p_{i,j} \int_{\Theta_{i,j}^*(\xi^*)} I_{i,j}(x, \bar{\theta}_i, \theta_{i,j}) \rho_{ij}^*(d\theta_{i,j}) \leq KL_P(\xi^*) \quad (4.8)$$

выполнено для всех $x \in \mathcal{X}$. Более того, если ξ^* — локальный KL_P -оптимальный план, то для его опорных точек в (4.8) достигается равенство.

Эта теорема доказывается аналогично теореме 6 из предыдущей главы. Пусть также выполнено

Предположение 7. Для любого плана ξ , такого что $KL_P(\xi) > 0$ и соответствующий вес $p_{i,j} \neq 0$, инфимумы в (4.4) достигаются в единственной точке $\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\xi)$ во внутренней области множества Θ_j .

Тогда все меры ρ_{ij}^* в теореме 13 сосредоточены в одной точке и левая часть неравенства (4.8) упрощается до

$$\Psi_{KL}(x, \xi) = \sum_{i,j=1}^{\nu} p_{i,j} I_{i,j}(x, \bar{\theta}_i, \hat{\theta}_{i,j}). \quad (4.9)$$

Следовательно, если ξ не является локальным KL_P -оптимальным планом, то существует точка $\bar{x} \in \mathcal{X}$, такая что $\Psi_{KL}(\bar{x}, \xi) > KL_P(\xi)$.

Далее будем считать, что предположения 6 и 7 выполнены.

4.4 Численные алгоритмы для построения KL_P -оптимальных планов

В этом параграфе мы опишем обобщение алгоритма 3 из главы 3 на случай байесовского KL_P -критерия для дискретных априорных распределений. В качестве базового алгоритма для нахождения KL_P -оптимальных планов мы снова будем использовать версию алгоритма из [6]:

Алгоритм 4. Пусть ξ_0 — это некоторый начальный план и пусть $(\alpha_s)_{s=0}^\infty$ — это последовательность положительных вещественных чисел, таких что $\lim_{s \rightarrow \infty} \alpha_s = 0$, $\sum_{s=0}^\infty \alpha_s = \infty$, $\sum_{s=0}^\infty \alpha_s^2 < \infty$. Для $s = 0, 1, \dots$ поочередно выполняем два шага:

1. Находим $x_{s+1} = \arg \max_{x \in \mathcal{X}} \Psi_{KL}(x, \xi_s)$.
2. Берем $\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi(x_{s+1})$.

Порождаемая алгоритмом 4 последовательность планов $(\xi_s)_{s \in \mathbb{N}}$ сходится к KL_P -оптимальному плану ξ^* в том смысле, что $\lim_{s \rightarrow \infty} \text{KL}_P(\xi_s) = \text{KL}_P(\xi^*)$. Если в KL_P -критерии задействовано достаточно большое количество попарных сравнений, то алгоритм 4 может не сойтись к плану с желаемой нижней границей эффективности (смотри пример в параграфе 4.5). Сложности возникают из-за того, что алгоритм 4 порождает последовательность планов с постоянно увеличивающимся количеством опорных точек.

Алгоритм 3 обобщается на случай KL_P -критерия следующим образом:

Алгоритм 5. Пусть ξ_0 — это некоторый начальный план. Определим последовательность планов $(\xi_s)_{s=0,1,\dots}$ следующим образом:

1. Пусть $\mathcal{S}_{[s]}$ — носитель плана ξ_s . Найдем множество $\mathcal{E}_{[s]}$ всех локальных максимумов функции $\Psi_{KL}(x, \xi_s)$ на $\mathcal{X}$ и положим $\mathcal{S}_{[s+1]} = \mathcal{S}_{[s]} \cup \mathcal{E}_{[s]}$.
2. Определим $\xi = \{\mathcal{S}_{[s+1]}, \omega\}$ как план с носителем $\mathcal{S}_{[s+1]}$ и вектором весов ω . Найдем локальный KL_P -оптимальный план в классе всех планов с носителем $\mathcal{S}_{[s+1]}$, то есть найдем вектор $\omega_{[s+1]}$, доставляющий максимум функции

$$g(\omega) = \text{KL}_P(\{\mathcal{S}_{[s+1]}, \omega\}) = \sum_{i,j=1}^{\nu} p_{i,j} \inf_{\theta_{i,j} \in \Theta_j} \sum_{x \in \mathcal{S}_{[s+1]}} I_{i,j}(x, \bar{\theta}_i, \theta_{i,j}) \omega_x,$$

где w_x — это вес в точке $x \in \mathcal{S}_{s+1}$. Уберем из $\mathcal{S}_{s+1}$ те точки, которым в полученном оптимальном векторе весов $\omega_{[s+1]}$ соответствуют нули, и получим новый носитель, который также обозначим $\mathcal{S}_{[s+1]}$. Наконец, план ξ_{s+1} определим как план носителем $\mathcal{S}_{[s+1]}$ и соответствующими ненулевыми весами из $\omega_{[s+1]}$.

Для алгоритма 5 верен следующий результат о сходимости:

Теорема 14. Пусть выполняется предположение 6 и пусть $\{\xi_s\}_{s=0,1,\dots}$ — это последовательность планов, получаемая с помощью алгоритма 5. Тогда

$$\lim_{s \rightarrow \infty} KL_P(\xi_{s+1}) = KL_P(\xi^*),$$

где ξ^* — локальный KL_P -оптимальный план.

Эта теорема доказывается аналогично теореме 11 из предыдущей главы.

В параграфах 4.4.1 и 4.4.2 мы также рассмотрим два численных метода оптимизации по весам, аналогичных тем, что описаны в параграфах 3.4.1 и 3.4.2 предыдущей главы.

4.4.1 Квадратичное программирование

Пусть $\mathcal{S}_{[s+1]} = \{x_1, \dots, x_n\}$ обозначает множество, получаемое на первом шаге s -й итерации алгоритма 5. Определим ξ как план с носителем $\mathcal{S}_{[s+1]}$ и весами $\omega_1, \dots, \omega_n$, которые предстоит определить на втором шаге алгоритма посредством максимизации функции

$$g(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} \sum_{k=1}^n \omega_k I_{i,j}(x_k, \bar{\theta}_i, \hat{\theta}_{i,j}),$$

где

$$\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\omega) = \arg \inf_{\theta_{i,j} \in \Theta_j} \sum_{k=1}^n \omega_k I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j}). \quad (4.10)$$

Мы аппроксимируем функцию $g(\omega)$ в окрестности фиксированных точек $\widehat{\theta}_{i,j}$ с помощью первых трех членов ряда Тейлора

$$\begin{aligned}
g(\omega) &\approx \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j}} \sum_{k=1}^n \omega_k \left\{ I_{i,j}(x_k, \bar{\theta}_i, \widehat{\theta}_{i,j}) + \alpha_{i,j}^T \left[\frac{\partial I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r)} \Big|_{\theta_{i,j}=\widehat{\theta}_{i,j}} \right]_{r=1}^{d_j} \right. \\
&\quad \left. + \frac{1}{2} \alpha_{i,j}^T \left[\frac{\partial^2 I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r) \partial \theta_{i,j}(s)} \Big|_{\theta_{i,j}=\widehat{\theta}_{i,j}} \right]_{r,s=1}^{d_j, d_j} \alpha_{i,j} \right\} \\
&= \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j}} \left\{ b_{i,j}^T \omega + \omega^T \mathbf{R}_{i,j} \alpha_{i,j} + \frac{1}{2} \alpha_{i,j}^T \mathbf{H}_{i,j} \alpha_{i,j} \right\},
\end{aligned} \tag{4.11}$$

где

$$\begin{aligned}
\mathbf{b}_{i,j} &= \mathbf{b}_{i,j}(\widehat{\theta}_{i,j}) = \left[I_{i,j}(x_k, \bar{\theta}_i, \widehat{\theta}_{i,j}) \right]_{k=1}^n \in \mathbb{R}^n, \\
\mathbf{R}_{i,j} &= \mathbf{R}_{i,j}(\widehat{\theta}_{i,j}) = \left[\frac{\partial I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r)} \Big|_{\theta_{i,j}=\widehat{\theta}_{i,j}} \right]_{k,r=1}^{n, d_j} \in \mathbb{R}^{n \times d_j}, \\
\mathbf{H}_{i,j} &= \mathbf{H}_{i,j}(\omega, \widehat{\theta}_{i,j}) = \sum_{k=1}^n \omega_k \left[\frac{\partial^2 I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r) \partial \theta_{i,j}(s)} \Big|_{\theta_{i,j}=\widehat{\theta}_{i,j}} \right]_{r,s=1}^{d_j, d_j} \\
&= \sum_{k=1}^n \omega_k \mathbf{H}_{i,j,k} \in \mathbb{R}^{d_j \times d_j}.
\end{aligned}$$

Минимум по $\alpha_{i,j} \in \mathbb{R}^{d_j}$ в формуле (4.11) достигается при

$$\widehat{\alpha}_{i,j} = -[\mathbf{H}_{i,j}]^{-1} \mathbf{R}_{i,j}^T \omega,$$

поэтому получаем

$$g(\omega) \approx \mathbf{b}^T \omega - \omega^T \mathbf{Q}(\omega) \omega,$$

где

$$\mathbf{b} = \sum_{i,j=1}^{\nu} p_{i,j} \mathbf{b}_{i,j}, \quad \mathbf{Q}(\omega) = \frac{1}{2} \sum_{i,j=1}^{\nu} p_{i,j} \mathbf{R}_{i,j} [\mathbf{H}_{i,j}(\omega)]^{-1} \mathbf{R}_{i,j}^T.$$

Введем обозначение

$$\Delta = \{\omega \in \mathbb{R}^n \mid \omega_i \geq 0 \ (i = 1, \dots, n) \sum_{i=1}^n \omega_i = 1\} \subset \mathbb{R}^n.$$

Если мы игнорируем зависимость матрицы $\mathbf{Q}(\omega)$ от ω и считаем ее фиксированной для некоторого заданного $\bar{\omega}$, то задача нахождения максимума по весам для аппроксимации $g(\omega)$ сводится к задаче квадратичного программирования

$$\phi(\omega, \bar{\omega}) = -\omega^T \mathbf{Q}(\bar{\omega}) \omega + \mathbf{b}^T \omega \rightarrow \max_{\omega \in \Delta}. \quad (4.12)$$

Последнюю задачу предлагается решать итерационно, подставляя каждый раз вместо $\bar{\omega}$ значение весов, полученное на предыдущей итерации.

В работе [10] рассматривалась задача дискриминации регрессионных моделей в рамках уравнения (4.1) для логнормальных плотностей ошибок с параметрами $\mu(x, \theta)$ и $\sigma^2(x, \theta)$. Среднее и дисперсия логнормальных плотностей задаются через эти параметры следующим образом:

$$\begin{aligned} \mathbb{E}[Y] &= \eta(x, \theta) = \exp \left\{ \frac{\sigma^2(x, \theta)}{2} + \mu(x, \theta) \right\}, \\ \text{Var}(Y) &= v^2(x, \theta) = \eta^2(x, \theta) \{ \exp \{ \sigma^2(x, \theta) \} - 1 \}. \end{aligned}$$

Сами плотности откликов Y равны

$$f(y, x, \theta) = \frac{1}{x \sqrt{2\pi} \sigma(x, \theta)} \exp \left\{ - \frac{\{\log(y) - \mu(x, \theta)\}^2}{2\sigma^2(x, \theta)} \right\}.$$

В [10] было показано, что расстояние Кульбака–Лейблера между двумя логнормальными плотностями с параметрами $\mu_\ell(x, \theta_\ell)$ и $\sigma_\ell^2(x, \theta_\ell)$ ($\ell = i, j$) равно

$$I_{i,j}(x, \theta_i, \theta_{i,j}) = \frac{1}{2} \left\{ s_{i,j}(x, \theta_{i,j}) + \frac{[\mu_i(x, \theta_i) - \mu_j(x, \theta_{i,j})]^2}{\sigma_i^2(x, \theta_i)} - 1 \right\}, \quad (4.13)$$

где

$$s_{i,j}(x, \theta_{i,j}) = \log \left[\frac{\sigma_i^2(x, \theta_i)}{\sigma_j^2(x, \theta_{i,j})} \right] + \frac{\sigma_j^2(x, \theta_{i,j})}{\sigma_i^2(x, \theta_i)},$$

и

$$\begin{aligned} \sigma_i^2(x, \theta_i) &= \log [1 + v_i^2(x, \theta_i)/\eta_i^2(x, \theta_i)] , \\ \mu_i(x, \theta_i) &= \log [\eta_i(x, \theta_i)] - \sigma_i^2(x, \theta_i)/2. \end{aligned}$$

Посчитаем первые и вторые производные по варьируемым параметрам для функции (4.13):

$$\begin{aligned} \frac{\partial I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r)} &= \frac{1}{2} \frac{\partial s_{i,j}(x, \theta_{i,j})}{\partial \theta_{i,j}(r)} - \frac{\mu_i(x, \theta_i) - \mu_j(x, \theta_{i,j})}{\sigma_i^2(x, \theta_i)} \frac{\partial \mu_j(x, \theta_{i,j})}{\partial \theta_{i,j}(r)}, \\ \frac{\partial^2 I_{i,j}(x_k, \bar{\theta}_i, \theta_{i,j})}{\partial \theta_{i,j}(r) \partial \theta_{i,j}(s)} &= \frac{1}{2} \frac{\partial^2 s_{i,j}(x, \theta_{i,j})}{\partial \theta_{i,j}(r) \partial \theta_{i,j}(s)} \\ &+ \frac{1}{\sigma_i^2(x, \theta_i)} \frac{\partial \mu_j(x, \theta_{i,j})}{\partial \theta_{i,j}(r)} \frac{\partial \mu_j(x, \theta_{i,j})}{\partial \theta_{i,j}(s)} - \frac{\mu_i(x, \theta_i) - \mu_j(x, \theta_{i,j})}{\sigma_i^2(x, \theta_i)} \frac{\partial^2 \mu_j(x, \theta_{i,j})}{\partial \theta_{i,j}(r) \partial \theta_{i,j}(s)}. \end{aligned} \quad (4.14)$$

Для ускорения вычислений в (4.14) предлагается игнорировать слагаемые, содержащие вторые производные. Тогда функция $g(\omega)$ может быть аппроксимирована функцией

$$\bar{g}(\omega) = \frac{1}{2} \sum_{i,j=1}^{\nu} p_{i,j} \min_{\alpha_{i,j}} \left[\alpha_{i,j}^T \bar{\mathbf{J}}_{i,j}^T \boldsymbol{\Omega}_i \bar{\mathbf{J}}_{i,j} \alpha_{i,j} + \omega^T \mathbf{R}_{i,j} \alpha_{i,j} + \mathbf{b}_{i,j}^T \omega \right],$$

где $\mathbf{\Omega}_i = \text{diag}(\omega_1, \dots, \omega_n)$,

$$\begin{aligned} \bar{\mathbf{J}}_{i,j} &= \left(\frac{\frac{\partial \mu_j(x_k, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j}=\hat{\theta}_{i,j}}}{\sigma_i(x_1, \bar{\theta}_i)} \right)_{k=1, \dots, n} ; \\ \mathbf{R}_{i,j} &= \left(\frac{\partial s_{i,j}(x_k, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j}=\hat{\theta}_{i,j}} - 2 \frac{[\mu_i(x_k, \bar{\theta}_i) - \mu_j(x_k, \hat{\theta}_{i,j})] \frac{\partial \mu_i(x_k, \theta_{i,j})}{\partial \theta_{i,j}} \Big|_{\theta_{i,j}=\hat{\theta}_{i,j}}}{\sigma_i^2(x_k, \bar{\theta}_i)} \right)_{k=1, \dots, n} ; \\ \mathbf{b}_{i,j} &= \left(I_{i,j}(x_k, \bar{\theta}_i, \hat{\theta}_{i,j}) \right)_{k=1, \dots, n} . \end{aligned}$$

4.4.2 Градиентный метод

В этом параграфе мы опишем специализированный градиентный метод для нахождения максимума по весам на втором шаге алгоритма 5. Введем функции

$$v_k(\omega) = \sum_{i,j=1}^{\nu} p_{i,j} I_{i,j}(x_k, \bar{\theta}_i, \hat{\theta}_{i,j}(\omega)), \quad k = 1, \dots, n,$$

где $\hat{\theta}_{i,j} = \hat{\theta}_{i,j}(\omega)$ определено в (4.10). Далее будем вычислять итерационно последовательность векторов $(\omega_{(\gamma)})_{\gamma=0,1,\dots}$ начиная с вектора $\omega_{(0)} = \bar{\omega}$ (который, например, имеет равные веса). Для $\omega_{(\gamma)} = (\omega_{(\gamma),1}, \dots, \omega_{(\gamma),n})$ найдем такие индексы $\bar{k}$ и $\underline{k}$, которые соответствуют $\max_{1 \leq k \leq n} v_k(\omega_{(\gamma)})$ и $\min_{1 \leq k \leq n} v_k(\omega_{(\gamma)})$. Определим

$$\alpha^* = \arg \max_{0 \leq \alpha \leq \omega_{(\gamma),\underline{k}}} g(\bar{\omega}_{(\gamma)}(\alpha)), \quad (4.15)$$

где вектор $\bar{\omega}_{(\gamma)}(\alpha) = (\bar{\omega}_{(\gamma),1}(\alpha), \dots, \bar{\omega}_{(\gamma),n}(\alpha))$ задается как

$$\bar{\omega}_{(\gamma),i}(\alpha) = \begin{cases} \omega_{(\gamma),i} + \alpha, & \text{если } i = \bar{k}; \\ \omega_{(\gamma),i} - \alpha, & \text{если } i = \underline{k}; \\ \omega_{(\gamma),i}, & \text{иначе.} \end{cases}$$

Вектор $\omega_{(\gamma+1)}$ на следующей итерации определяется как $\omega_{(\gamma+1)} = \bar{\omega}_{(\gamma)}(\alpha^*)$. Получаемая таким образом последовательность сходится к ω^* , которое доставляет максимум функции $g(\omega)$ (это можно показать так же, как в [15]).

4.5 Примеры

В этом параграфе мы проиллюстрируем работу предложенных алгоритмов для нахождения байесовских KL -оптимальных планов в случае дискриминации нескольких моделей с логнормальными ошибками. Алгоритм 5 с двумя способами оптимизации по весам из параграфов 4.4.1 и 4.4.2 был реализован в пакете [17] для R , в котором также можно найти примеры из параграфов 4.5.1, 4.5.2 и 4.5.3. Сначала приведем некоторые дополнительные замечания, касающиеся реализации алгоритмов.

1. На этапе 1 алгоритма 5 все локальные максимумы $\Psi_{KL}(x, \xi_s)$ добавляются в носитель на каждой итерации. Для того, чтобы избежать чрезмерного накопления точек в носителе промежуточного плана, мы удаляем на каждой итерации те точки, вес которых опускается ниже $m^{0.25}$, где $m = 2.2 \times 10^{-16}$ рабочая точность для чисел с плавающей точкой в языке R.
2. Замечание о реализации метода оптимизации по весам на втором шаге алгоритма 5, основанного на квадратичном программировании. Во время шага 2 мы проводим только несколько итераций предлагаемой процедуры, не дожидаясь сходимости к оптимальному вектору весов при фиксированном носителе. Это позволяет существенно сократить время выполнения программы. При этом в рассмотренных примерах на общую сходимость алгоритма раннее прерывание не повлияло.
3. Замечание о реализации метода оптимизации по весам на втором шаге алгоритма 5, основанного на градиентном методе. Мы использовали линеаризацию, подобную описанной в параграфе 4.4.1, для ускорения вычислений.

4.5.1 KL -оптимальные планы для дискриминации ЕМАХ-модели и модели Михаэлиса–Ментен

Начнем с примера, аналогичного рассмотренному в [10]. Найдем локальный KL -оптимальный план для дискриминации двух логнормальных моделей

со средними

$$\begin{aligned}\eta_1(x, \theta_1) &= \frac{\theta_{1,1}x}{\theta_{1,2} + x} + \theta_{1,3}x, \\ \eta_2(x, \theta_2) &= \frac{\theta_{2,1}x}{\theta_{2,2} + x}\end{aligned}\tag{4.16}$$

на интервале $\mathcal{X} = [0.1, 5]$. Пусть первая модель η_1 имеет фиксированные параметры $\bar{\theta}_1 = (1, 1, 1)$. Рассмотрим ситуацию, когда в критерии (4.4) $\nu = 2$ и $p_{1,2} = 1$, $p_{2,1} = 0$. В таблице 7 представлены оптимальные планы, найденные при разных функциях для дисперсии:

$$\begin{aligned}(1) \quad & v_1^2(x, \theta_1) = v_2^2(x, \theta_2) = 1 \\ (2) \quad & \sigma_1^2(x, \theta_1) = \sigma_2^2(x, \theta_2) = 1 \\ (3) \quad & v_i^2(x, \theta_i) = \exp(\eta_i(x, \theta_i))\end{aligned}\tag{4.17}$$

Нижняя оценка эффективности для всех представленных планов равна 0.999. Каждый план был вычислен тремя методами: 1) алгоритмом 4, рассматриваемым [10], 2) алгоритмом 5 со вторым шагом из параграфа 4.4.1, использующим квадратичное программирование, 3) алгоритмом 5 со вторым шагом из параграфа 4.4.2, использующим градиентный метод. Для случая (2), который предполагает равенство функций дисперсии в (4.17), соответствующая функция в (4.13) упрощается до $(\mu_1(x, \bar{\theta}_1) - \mu_2(x, \theta_{1,2}))^2$, поэтому мы можем использовать процедуру для вычисления оптимального плана в случае нормальных ошибок наблюдения, описанную в главе 3, где $\mu_i(x, \theta_i) = \log \eta_i(x, \theta_i)$, $i = 1, 2$. Эта процедура работает значительно быстрее, что видно из таблицы 7, в которой указано время работы каждого алгоритма в секундах. Все тесты выполнялись на компьютере с процессором intel core i7-4790K. Предложенные методы работают существенно быстрее, нежели чем стандартный алгоритм из [10]. Так градиентный метод показывает прирост от 5 до 30 раз, а подход, основанный на квадратичном программировании — от 25 до 120 раз по сравнению со стандартным алгоритмом.

Таблица 7 — Локальные KL_P -оптимальные планы для дискриминации моделей (4.16) при логнормальных откликах с дисперсиями из (4.17)

(4.17)	Оптимальный план			AF	grad	quad
(1)	0.130	2.501	5.000	7.15	0.56	0.06
	0.489	0.378	0.133			
(2)	0.100	1.569	5.000	2.74	0.52	0.01
	0.294	0.500	0.206			
(3)	0.100	1.218	5.000	10.87	0.33	0.08
	0.326	0.510	0.164			

Примечание — В колонках 3–5 показано время работы различных алгоритмов: AF — алгоритм 4 из [10]; grad — алгоритм 5 со вторым шагом из § 4.4.2; quad — алгоритм 5 со вторым шагом из § 4.4.1

4.5.2 Байесовские KL_P -оптимальные планы для дискриминации экспоненциальных моделей

Теперь найдем байесовский KL_P -оптимальный план для дискриминации двух конкурирующих экспоненциальных моделей:

$$\begin{aligned}\eta_1(x, \theta_1) &= \theta_{1,1} - \theta_{1,2} \exp(-\theta_{1,3} x^{\theta_{1,4}}); \\ \eta_2(x, \theta_2) &= \theta_{2,1} - \theta_{2,2} \exp(-\theta_{2,3} x).\end{aligned}\tag{4.18}$$

на интервале $[0, 10]$. Снова будем считать, что параметры первой модели η_1 фиксированы. В случае нормально распределенных ошибок наблюдения оптимальный план для этих моделей был найден в работе [15] (смотри предыдущую главу). Посмотрим на то, как изменятся планы в предположении логнормальности ошибок наблюдения. Как и в предыдущем примере, рассмотрим три типа логнормальных ошибок, определяемых соотношениями (4.17). Отметим, что оптимальный план для дискриминации моделей (4.18) не зависит от параметров $\bar{\theta}_{2,1}$ и $\bar{\theta}_{2,2}$, поэтому их можно взять произвольными. Рассмотрим независимые априорные распределения в точках

$$\mu_j + \frac{\sqrt{0.3}(i-3)}{2}, \quad i = 1, \dots, 5; \quad j = 3, 4, \tag{4.19}$$

для параметров $\bar{\theta}_{1,3}$ и $\bar{\theta}_{1,4}$, где $\mu_3 = 0.8$, $\mu_4 = 1.5$. Соответствующие веса в обоих случаях пропорциональны

$$\frac{1}{\sqrt{2\pi \cdot 0.3}} \exp\left(-\frac{(i-3)^2}{8}\right); \quad i = 1, \dots, 5. \quad (4.20)$$

В таблице 8 представлены байесовские KL_P -оптимальные планы для дискриминации моделей (4.18) в случае логнормальных ошибок наблюдения для трех вариантов функции дисперсии, которые описаны в (4.17). На рисунке 4.1 представлены графики функций из левой части неравенства (4.8) в теореме эквивалентности 13. Из результатов в таблице 8 видно, что алгоритм 5 со вторым шагом, использующим квадратичное программирование, существенно превосходит другие алгоритмы по времени работы. Интересно также сравнить байесовские KL_P -оптимальные планы для логнормальных ошибок с байесовским T_P -оптимальным планом для тех же конкурирующих моделей из главы 3. Соответствующий T_P -оптимальный план сосредоточен в пяти точках 0.000, 0.452, 1.747, 4.951 и 10.000 с весами 0.207, 0.396, 0.292, 0.003 и 0.102. Эффективности вида

$$\text{Eff}_{KL_P}^{(j)}(\xi_{(i)}^*) = \frac{KL_P^{(j)}(\xi_{(i)}^*)}{\sup_{\eta} KL_P^{(j)}(\eta)}$$

приведены в таблице 9. Так, например, эффективность плана $\xi_{(0)}^*$, вычисленного в предположении гомоскедастичных нормальных ошибок наблюдения, в модели с логнормальными ошибками наблюдения типа (2) из (4.17) равна 0.953. Из таблицы видно, что байесовские планы, найденные в предположении о нормальности ошибок наблюдения, достаточно устойчивы и имеют высокие эффективности в моделях с логнормальными ошибками.

4.5.3 Байесовские KL_P -оптимальные планы для дискриминации моделей зависимости эффективности препарата от дозы

В качестве еще одного примера построим байесовский KL_P -оптимальный план для дискриминации четырех моделей зависимости эффективности препа-

Таблица 8 — Байесовские KL_P -оптимальные планы для дискриминации моделей (4.18) при логнормальных откликах с дисперсиями из (4.17)

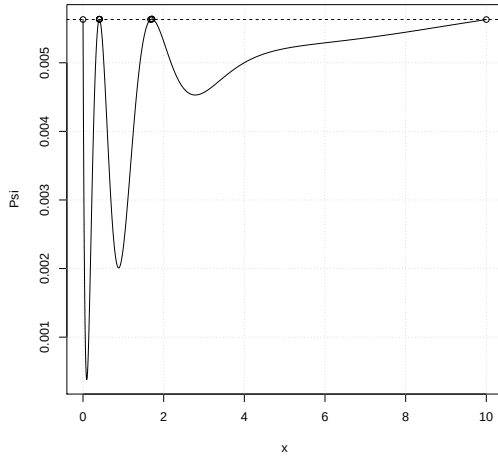
(4.17)	Оптимальный план $\xi_{(i)}^*$					AF	grad	quad
(1)	0 0.186	0.406 0.418	1.706 0.289	10 0.107		298.37	44.36	3.7
(2)	0 0.189	0.374 0.397	1.650 0.311	10 0.103		390.44	7.39	2.39
(3)	0 0.186	0.356 0.394	1.604 0.313	10 0.107		570.45	39.19	4.42

Примечание — В колонках 3–5 показано время работы различных алгоритмов в секундах: AF — алгоритм 4 из [10]; grad — алгоритм 5 со вторым шагом из § 4.4.2; quad — алгоритм 5 со вторым шагом из § 4.4.1

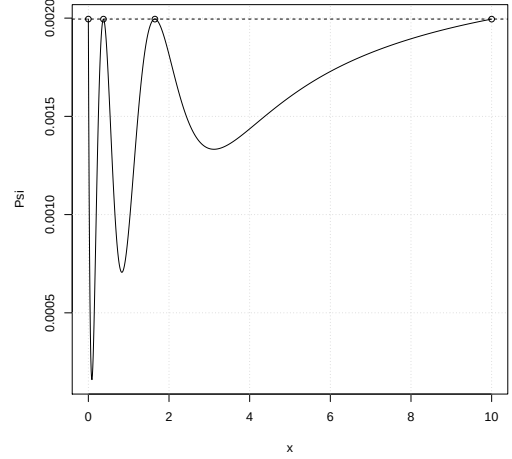
Таблица 9 — Относительные эффективности планов из таблицы 8 при различных предположениях о распределении ошибок

	(0)	(1)	(2)	(3)
(0)	1	0.978	0.953	0.908
(1)	0.981	1	0.988	0.966
(2)	0.951	0.987	1	0.992
(3)	0.923	0.970	0.996	1

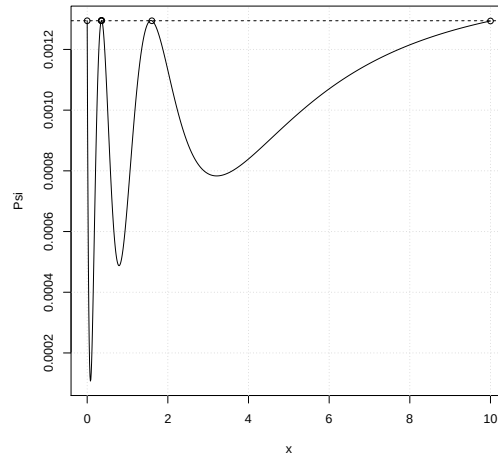
Примечание — (0): гомоскедастичное нормальное распределение; (1) - (3): логнормальные распределения с различными функциями дисперсии из (4.17)



а) Случай (4.17)(1)



б) Случай (4.17)(2)



в) Случай (4.17)(3)

Непрерывной линией обозначен график функции $\Psi_{KL}(x, \xi_{(i)}^*)$, задаваемой формулой (4.9), пунктирной — значение функционала $KL_P(\xi_{(i)}^*)$, определенного в (4.4), а выколотыми точками — опорные точки

KL_P -оптимального плана $\xi_{(i)}^*$.

Рисунок 4.1 — Графики функций $\Psi_{KL}(x, \xi_{(i)}^*)$ для планов из таблицы 8

рата от дозы, рассмотренных в работе [35]:

$$\begin{aligned}
 \eta_1(x, \theta_1) &= \theta_{1,1} + \theta_{1,2}x; \\
 \eta_2(x, \theta_2) &= \theta_{2,1} + \theta_{2,2}x(\theta_{2,3} - x); \\
 \eta_3(x, \theta_3) &= \theta_{3,1} + \theta_{3,2}x/(\theta_{3,3} + x); \\
 \eta_4(x, \theta_4) &= \theta_{4,1} + \theta_{4,2}/(1 + \exp(\theta_{4,3} - x)/\theta_{4,4});
 \end{aligned} \tag{4.21}$$

на интервале $\mathcal{X} = [0, 500]$. В [35] также приводится априорная информация о параметрах:

$$\begin{aligned}
 \bar{\theta}_1 &= (60, 0.56), \\
 \bar{\theta}_2 &= (60, 7/2250, 600), \\
 \bar{\theta}_3 &= (60, 294, 25), \\
 \bar{\theta}_4 &= (49.62, 290.51, 150, 45.51).
 \end{aligned}$$

В главе 3 были найдены байесовские TP -оптимальные планы для дискриминации этих моделей в предположении о нормальности ошибок наблюдения. Априорное распределение задавалось только для параметров θ_4 . В соответствующем критерии использовались веса $p_{i,j} = 1/6$, ($1 \leq j < i \leq 4$). В этом параграфе мы рассмотрим похожий пример для логнормальных ошибок наблюдения. Априорное равномерное распределение для θ_4 зададим в 81 точке из $\mathbb{R}^4$:

$$(49.62 + c_1, 290.51 + c_2, 150 + c_3, 45.51 + c_4), \tag{4.22}$$

где $c_1, c_2, c_3, c_4 \in \{-20, 0, 45\}$. Итоговый байесовский критерий KL_P -оптимальности (4.6) предусматривает 246 попарных сравнений конкурирующих моделей. Планы, оптимальные относительно этого критерия, представлены в таблице 10. Рассматривается три типа задач:

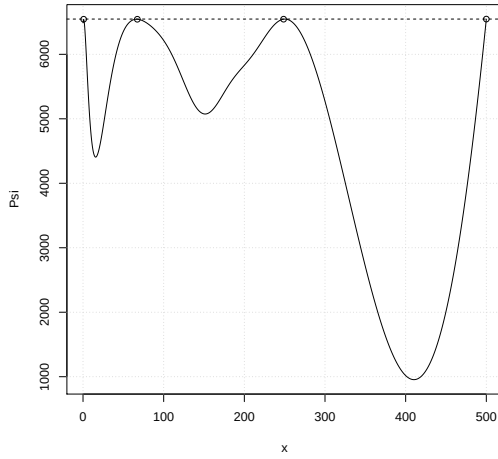
$$\begin{aligned}
 (1) \quad & v_1^2(x, \theta_1) = 1, \quad i = 1, 2, 3, 4; \\
 (2) \quad & \sigma_i^2(x, \theta_i) = 1, \quad i = 1, 2, 3, 4; \\
 (3) \quad & v_i^2(x, \theta_i) = \exp(\eta_i(x, \theta_i)/100), \quad i = 1, 2, 3, 4.
 \end{aligned} \tag{4.23}$$

Все представленные планы имеют нижнюю границу эффективности 0.999. Графики функций из левой части теоремы эквивалентности 13 для всех трех слу-

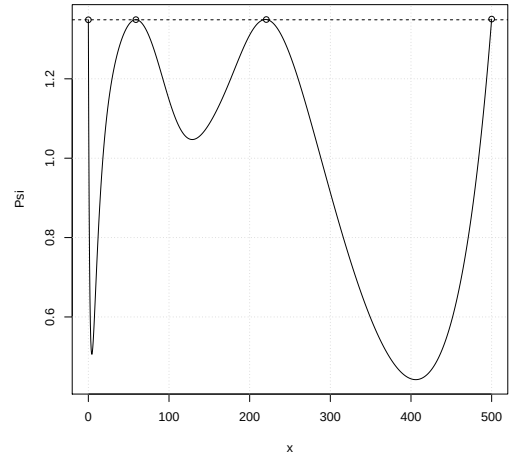
Таблица 10 — Байесовские KL_P -оптимальные планы для дискриминации моделей (4.21) при логнормальных откликах с дисперсиями из (4.23)

(4.23)	Оптимальный план $\xi_{(i)}^*$							AF	grad	quad
(1)	0.759 67.32 248.6 500 0.419 0.156 0.233 0.192							1674.14	679.52	48.91
(2)	0 58.9 220.6 500 0.200 0.354 0.247 0.199							-	255.03	33.42
(3)	0 33.12 78.0 161.6 215.7 500 0.279 0.092 0.225 0.003 0.224 0.177							2382.64	631.53	82.33

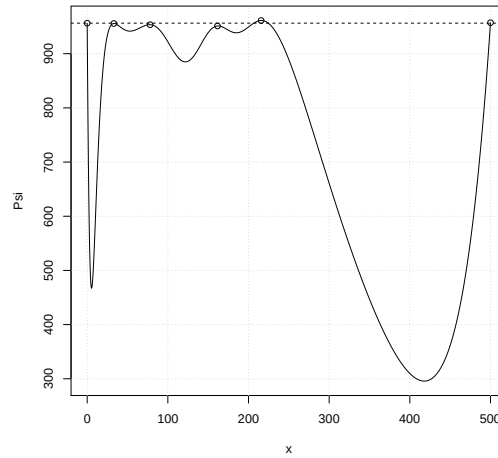
чаев приведены на рисунке 4.2. Стандартный алгоритм не сошелся к плану с заданной эффективностью в случае (4.23)(2). Из таблицы 10 видно существенное преимущество новых алгоритмов в скорости работы по сравнению со стандартным методом.



а) Случай (4.23)(1)



б) Случай (4.23)(2)



в) Случай (4.23)(3)

Непрерывной линией обозначен график функции $\Psi_{KL}(x, \xi_{(i)}^*)$, задаваемой формулой (4.9), пунктирной — значение функционала $KL_P(\xi_{(i)}^*)$, определенного в (4.4), а выколотыми точками — опорные точки

KL_P -оптимального плана $\xi_{(i)}^*$.

Рисунок 4.2 — Графики функций $\Psi_{KL}(x, \xi_{(i)}^*)$ для планов из таблицы 10

Глава 5. Построение полу-параметрических оптимальных планов

В литературе по планированию дискриминационных экспериментов основное внимание, как правило, уделяется случаю, когда конкурирующие модели, каждая из которых определяется связкой из функции регрессии и закона распределения для случайных ошибок наблюдения, заранее заданы с точностью до неизвестных параметров. В пионерских работах [6; 12] рассматривается случай нормально распределенных ошибок наблюдения с постоянной дисперсией, [9] обобщают этот подход на случай нормальных ошибок с произвольной дисперсией, а более новые результаты [10] позволяют проводить дискриминацию в случае произвольно распределенных ошибок с заранее заданными законами распределения с помощью критерия, основанного на расстоянии Кульбака-Лейблера. В работе [11] был предложен полу-параметрический критерий оптимальности для планирования дискриминационных экспериментов, обобщающий критерий из [10]. Суть этого критерия состоит в том, что функции регрессии для всех конкурирующих моделей и закон распределения ошибок для одной из них задаются а priori, как и раньше, а закон распределения для оставшейся модели получается как решение некоторой оптимизационной задачи.

В этой главе мы опишем новый способ, позволяющий достаточно просто находить оптимальные относительно полу-параметрического критерия планы. Мы также сформулируем соответствующие теоремы эквивалентности, позволяющие проверять полученные планы на оптимальность и выведем несколько интересных соотношений между полу-параметрическими и параметрическими критериями оптимальности.

5.1 Полу-параметрические оптимальные планы

Пусть Y — абсолютно непрерывная зависимая переменная, а x — независимая переменная, заданная на компактном множестве планирования $\mathcal{X}$. Плотность Y при фиксированном x относительно меры Лебега обозначим как $f(y, x)$. Мы хотим построить план эксперимента, который бы позволил эффективно дискриминировать две конкурирующие модели для плотности $f(y, x)$. Пусть плотности $f_j(y, x, \theta_j)$, $j = 1, 2$ заданы с точностью до параметров $\theta_j \in \Theta_j$, где Θ_j

компактно. Расстояние Кульбака-Лейблера между ними зададим как

$$I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) = \int f_1(y, x, \theta_1) \log \frac{f_1(y, x, \theta_1)}{f_2(y, x, \theta_2)} dy. \quad (5.1)$$

Пусть первая модель имеет фиксированные параметры $\bar{\theta}_1$. Критерий KL -оптимальности в этом случае записывается как

$$KL_{1,2}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx). \quad (5.2)$$

В работе [11] было предложено в критерии (5.2) одну из моделей задать параметрически, а другую модель получать как решение некоторой задачи вариационного исчисления с ограничениями. Рассмотрим подробнее постановку задачи из [11]. Опорное множество плотности $f_j(y, x, \theta_j)$ обозначим как

$$\mathcal{S}_{f_j, \theta_j, x} = \{y \mid f_j(y, x, \theta_j) > 0\}, \quad j = 1, 2, \quad (5.3)$$

а математическое ожидание для $f_j(y, x, \theta_j)$ при фиксированных x и θ_j как

$$\eta_j(x, \theta_j) = \int y f_j(y, x, \theta_j) dy, \quad j = 1, 2.$$

Пусть плотность $f_1(y, x, \bar{\theta}_1)$ задана параметрически. Обозначим как

$$\mathcal{F}_{2,x,\theta_2} = \left\{ f_2(y, x, \theta_2) : \int f_2(y, x, \theta_2) dy = 1, \right. \\ \left. \int y f_2(y, x, \theta_2) dy = \eta_2(x, \theta_2), \mathcal{S}_{f_2, \theta_2, x} = \mathcal{S}_{f_1, \bar{\theta}_1, x} \right\} \quad (5.4)$$

класс всех плотностей $f_2(y, x, \theta_2)$ при фиксированных x и θ_2 , имеющих математическое ожидание $\eta_2(x, \theta_2)$. План ξ^* , максимизирующий критерий

$$K_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \inf_{f_2 \in \mathcal{F}_{2,x,\theta_2}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx) \quad (5.5)$$

будем называть $SKL_{(a)}$ -оптимальным. Этот критерий является локальным в смысле работы [27], так как он зависит от параметра $\bar{\theta}_1$.

Также мы можем зафиксировать плотность $f_2(y, x, \theta_2)$ и обозначить как

$$\mathcal{F}_{1,x,\bar{\theta}_1} = \left\{ f_1(y, x, \bar{\theta}_1) : \int f_1(y, x, \bar{\theta}_1) dy = 1, \right. \\ \left. \int y f_1(y, x, \bar{\theta}_1) dy = \eta_1(x, \bar{\theta}_1), \mathcal{S}_{f_1, \bar{\theta}_1, x} = \mathcal{S}_{f_2, \theta_2, x} \right\} \quad (5.6)$$

класс всех плотностей $f_1(y, x, \bar{\theta}_1)$ при фиксированных x и $\bar{\theta}_1$, которые имеют математическое ожидание $\eta_1(x, \bar{\theta}_1)$. План ξ^* , максимизирующий критерий

$$K_{(b)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_X \inf_{f_1 \in \mathcal{F}_{1,x,\bar{\theta}_1}} I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2) \xi(dx), \quad (5.7)$$

будем называть $SKL_{(b)}$ -оптимальным.

Для простоты мы далее будем предполагать, что функции $f_1(y, x, \theta_1)$, $f_2(y, x, \theta_2)$, $\eta_1(x, \theta_1)$, $\eta_2(x, \theta_2)$ дифференцируемы по y , x , θ_1 и θ_2 соответственно, несмотря на то, что эти предположения можно ослабить при необходимости.

В статье [11] для критериев (5.5) и (5.7) были получены явные представления. Для критерия (5.5)

$$K_{(a)}(\xi, \bar{\theta}_1) = \inf_{\theta_2 \in \Theta_2} \int_X \left\{ \mu + 1 + \int \log \{ -\mu - \lambda(y - \eta_2(x, \theta_2)) \} f_1(y, x, \bar{\theta}_1) dy \right\} \xi(dx), \quad (5.8)$$

где величины λ и μ , зависящие от x , $\bar{\theta}_1$ и θ_2 , находятся из системы уравнений

$$- \int \frac{f_1(y, x, \bar{\theta}_1)}{\mu + \lambda(y - \eta_2(x, \theta_2))} dy = 1, \quad \int \frac{(y - \eta_2(x, \theta_2)) f_1(y, x, \bar{\theta}_1)}{\mu + \lambda(y - \eta_2(x, \theta_2))} dy = 0 \quad (5.9)$$

и удовлетворяют ограничениям

$$\mu + \lambda(y - \eta_2(x, \theta_2)) < 0, \quad \forall y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}.$$

Аналогичное представление было найдено для (5.7).

Ниже доказано, что внутренняя оптимизационная задача в критериях (5.5) и (5.7) может быть сведена к решению одного нелинейного уравнения. Также получены более простые представления для критериев (5.5) и (5.7), что позволило существенно ускорить численное нахождение полу-параметрических планов.

Теорема 15.

(а) Пусть при всех $x \in \mathcal{X}$ опорное множество для плотности $f_1(y, x, \bar{\theta}_1)$ является интервалом, то есть $\mathcal{S}_{f_1, \bar{\theta}_1, x} = [y_{x, \min}, y_{x, \max}]$, причем $y_{x, \min} < \eta_2(x, \theta_2) < y_{x, \max}$ при всех $\theta_2 \in \Theta_2$. Пусть также при всех $x \in \mathcal{X}$ и при всех $\theta_2 \in \Theta_2$, уравнение

$$\int \frac{f_1(y, x, \bar{\theta}_1)}{1 + \lambda(y - \eta_2(x, \theta_2))} dy = 1 \quad (5.10)$$

имеет единственный ненулевой корень $\bar{\lambda}(x, \bar{\theta}_1, \theta_2)$ который удовлетворяет условию

$$-\frac{1}{y_{x, \max} - \eta_2(x, \theta_2)} < \bar{\lambda}(x, \bar{\theta}_1, \theta_2) < -\frac{1}{y_{x, \min} - \eta_2(x, \theta_2)}. \quad (5.11)$$

Тогда критерий (5.5) принимает вид

$$\begin{aligned} K_{(a)}(\xi, \bar{\theta}_1) &= \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int f_1(y, x, \bar{\theta}_1) \log \frac{f_1(y, x, \bar{\theta}_1)}{f_2^*(y, x, \theta_2)} dy \xi(dx) \\ &= \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int \log \{1 + \bar{\lambda}(x, \bar{\theta}_1, \theta_2)(y - \eta_2(x, \theta_2))\} f_1(y, x, \bar{\theta}_1) dy \xi(dx), \end{aligned} \quad (5.12)$$

где “оптимальная” плотность f_2^* в (5.5) задается как

$$f_2^*(y, x, \theta_2) = \frac{f_1(y, x, \bar{\theta}_1)}{1 + \bar{\lambda}(x, \bar{\theta}_1, \theta_2)(y - \eta_2(x, \theta_2))}. \quad (5.13)$$

(b) Пусть интегралы

$$\int f_2(y, x, \theta_2) \exp(-\lambda y) dy \quad \text{и} \quad \int y f_2(y, x, \theta_2) \exp(-\lambda y) dy$$

существуют при всех $x \in \mathcal{X}$ и при всех λ . Тогда критерий (5.7) принимает вид

$$\begin{aligned} K_{(b)}(\xi) &= \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int f_1^*(y, x, \bar{\theta}_1) \log \frac{f_1^*(y, x, \bar{\theta}_1)}{f_2(y, x, \theta_2)} dy \xi(dx) \\ &= \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \int [\log \{\mu'(\bar{\lambda}_x)\} - \bar{\lambda}_x y] f_2(y, x, \theta_2) \exp(-\bar{\lambda}_x y) \mu'(\bar{\lambda}_x) dy \xi(dx), \end{aligned} \quad (5.14)$$

где “оптимальная” плотность f_1^* в (5.7) задается как

$$f_1^*(y, x, \bar{\theta}_1) = \frac{f_2(y, x, \theta_2) \exp(-\bar{\lambda}(x, \bar{\theta}_1, \theta_2)y)}{\int f_2(y, x, \theta_2) \exp(-\bar{\lambda}(x, \bar{\theta}_1, \theta_2)y) dy}, \quad (5.15)$$

$\bar{\lambda}_x = \bar{\lambda}(x, \bar{\theta}_1, \theta_2)$ — это ненулевой корень уравнения

$$\frac{\int y f_2(y, x, \theta_2) \exp(-\lambda y) dy}{\int f_2(y, x, \theta_2) \exp(-\lambda y) dy} = \eta_1(x, \bar{\theta}_1), \quad (5.16)$$

а $\mu'(\lambda) = \mu'(\lambda, x, \theta_2) = 1 / \int f_2(y, x, \theta_2) \exp(-\lambda y) dy$.

Доказательство. (а) Для простоты введем обозначения

$$f_1(y, x, \bar{\theta}_1) = f_1(y), \quad f_2(y, x, \theta_2) = f_2(y), \quad \eta_1(x, \bar{\theta}_1) = \eta_1, \quad \eta_2(x, \theta_2) = \eta_2. \quad (5.17)$$

Из формы критерия (5.5) следует, что нужно минимизировать

$$I_{1,2}(x, f_1, f_2, \theta_1, \theta_2) = \int \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} f_1(y) dy \quad (5.18)$$

при ограничениях

$$\int f_2(y) dy = 1, \quad \text{и} \quad \int y f_2(y) dy = \eta_2. \quad (5.19)$$

Если

$$H = \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} f_1(y) + \mu f_2(y) + \lambda y f_2(y),$$

то мы получаем уравнение Эйлера-Лагранжа

$$\frac{\partial H}{\partial f_2} = -\frac{f_1(y)}{f_2(y)} + \mu + \lambda y = 0,$$

откуда выводится выражение для “оптимальной” плотности $f_2(y)$:

$$f_2(y) = \frac{f_1(y)}{\mu + \lambda y}. \quad (5.20)$$

Естественно предполагать, что $\lambda \neq 0$, так как иначе $\mu = 1$ и $f_2(y) = f_1(y)$. Из уравнений (5.19) и (5.20), получаем

$$\begin{aligned} \eta_2 &= \int y f_2(y) dy = \int \frac{y}{\mu + \lambda y} f_1(y) dy \\ &= \frac{1}{\lambda} \int \frac{\mu + \lambda y}{\mu + \lambda y} f_1(y) dy - \frac{\mu}{\lambda} \int \frac{1}{\mu + \lambda y} f_1(y) dy = \frac{1}{\lambda} - \frac{\mu}{\lambda}. \end{aligned}$$

Отсюда следует, что $\mu = 1 - \eta_2 \lambda$, и

$$f_2(y) = \frac{f_1(y)}{1 + \lambda(y - \eta_2)}.$$

Подставляя новое выражение для $f_2(y)$ в оставшееся ограничительное условие, получаем

$$\int \frac{f_1(y)}{1 + \lambda(y - \eta_2)} dy = 1. \quad (5.21)$$

По предположению теоремы уравнение (5.21) имеет единственное ненулевое решение $\bar{\lambda}$ в интервале (5.11) и поэтому условие $1 + \bar{\lambda}(y - \eta_2) > 0$ выполнено для всех $y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}$. Отсюда следует, что f_2 — это плотность.

(b) Из формы критерия (5.7) следует, что нужно минимизировать (5.18) при ограничениях

$$\int f_1(y) dy = 1, \quad \text{и} \quad \int y f_1(y) dy = \eta_1. \quad (5.22)$$

Если

$$H = \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} f_1(y) + \mu f_1(y) + \lambda y f_1(y),$$

то мы получаем уравнение Эйлера-Лагранжа

$$\frac{\partial H}{\partial f_1} = \log \left\{ \frac{f_1(y)}{f_2(y)} \right\} + 1 + \mu + \lambda y = 0$$

откуда выводится выражение для “оптимальной” плотности $f_1(y)$:

$$f_1(y) = f_2(y) \exp(-1 - \mu - \lambda y).$$

Обозначим

$$\mu' = \exp(-1 - \mu)$$

и подставим полученное выражение для $f_1(y)$ в оба ограничительных уравнения

$$\begin{aligned} \int_Y f_1(y) dy &= \mu' \int_Y f_2(y) \exp(-\lambda y) dy = 1, \\ \int_Y y f_1(y) dy &= \mu' \int_Y y f_2(y) \exp(-\lambda y) dy = \eta_1. \end{aligned}$$

Объединяем эти уравнения и получаем финальное уравнение для λ

$$\frac{\int_Y y f_2(y) \exp(-\lambda y) dy}{\int_Y f_2(y) \exp(-\lambda y) dy} = \eta_1, \quad (5.23)$$

которое уже необходимо решать численно. □

Таким образом, нахождение значений критериев (5.5) и (5.7) при фиксированном ξ сводится к численному нахождению корней уравнений (5.10) и (5.16) относительно λ .

Когда мы решаем уравнение (5.10), то естественно предполагать, что $\lambda < 0$, если $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, так как в этом случае функция $1/[1 + \lambda(y - \eta_2(x, \theta_2))]$ возрастает, если $y \in \mathcal{S}_{f_1, \bar{\theta}_1, x}$, что позволяет нам сместить среднее функции $f_1(y, x, \bar{\theta}_1)/[1 + \lambda(y - \eta_2(x, \theta_2))]$ вправо. Аналогично, если $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$,

то естественно искать $\lambda > 0$. Следующая лемма формализует приведенные рассуждения:

Лемма 4. *Пусть*

$$v_2^2(x, \theta_2) = \int (y - \eta_2(x, \theta_2))^2 f_2(y, x, \theta_2) dy$$

существует и является строго положительной. Если $\bar{\lambda}$ есть решение уравнения (5.10), которое удовлетворяет (5.11), $\bar{\lambda}$ имеет тот же знак, что и разность $\eta_1(x, \theta_1) - \eta_2(x, \theta_2)$.

Доказательство. При доказательстве леммы будем также использовать обозначения (5.17). Пусть $\bar{\lambda}$ есть решение уравнения (5.10) и f_2^* есть “оптимальная” плотность, задаваемая уравнением (5.13). Тогда

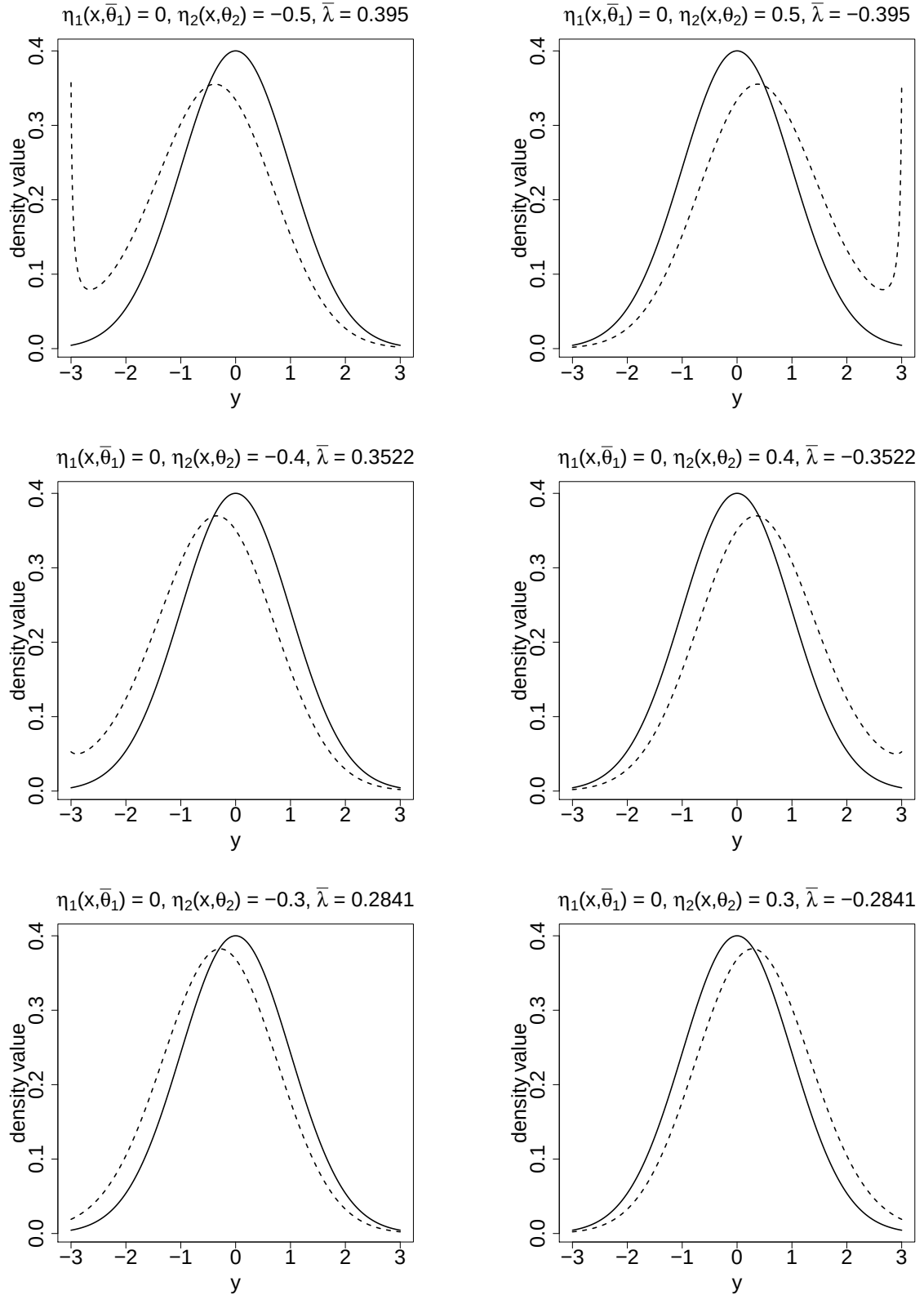
$$\begin{aligned} \eta_1 - \eta_2 &= \int_Y y f_1(y) dy - \int_Y y f_2^*(y) dy = \int_Y y f_1(y) \left\{ 1 - \frac{1}{1 + \bar{\lambda}(y - \eta_2)} \right\} dy \\ &= \int_Y y f_1(y) \frac{\bar{\lambda}(y - \eta_2)}{1 + \bar{\lambda}(y - \eta_2)} dy = \bar{\lambda} \int_Y y(y - \eta_2) f_2(y) dy \\ &= \bar{\lambda} \int_Y \{ (y^2 - y\eta_2 - y\eta_2 + \eta_2^2) + y\eta_2 - \eta_2^2 \} f_2(y) dy \\ &= \bar{\lambda} \int_Y (y - \eta_2^2) f_2(y) dy + \underbrace{\int_Y (y\eta_2 - \eta_2^2) f_2(y) dy}_{=0} = \bar{\lambda} v_2^2, \end{aligned}$$

где v_2^2 есть дисперсия плотности $f_2(y)$ и $v_2^2 > 0$. □

Пример 1. Пусть $f_1(y, x, \bar{\theta}_1)$ есть усеченная плотность нормального распределения $\mathcal{N}(\eta(x, \bar{\theta}_1), 1)$ заданная на интервале $[-3 + \eta_1(x, \bar{\theta}_1), 3 + \eta_1(x, \bar{\theta}_1)]$. Эта плотность является функцией от $\eta_1(x, \bar{\theta}_1)$ и из соотношения (5.13) следует, что в данном случае “оптимальная” плотность $f_2^*(y, x, \theta_2)$ является функцией от $\eta_1(x, \bar{\theta}_1)$ и $\eta_2(x, \theta_2)$. На рисунке 5.1 показаны графики функции f_2^* для $\eta_1(x, \bar{\theta}_1) \equiv 0$ и различных значений $\eta_2(x, \theta_2)$ на интервале $[-3, 3]$. Значение корня уравнения (5.10) $\bar{\lambda}$ показано сверху над каждым рисунком.

Основная разница между подходом, предложенным в [11], и подходом, описанным в теореме 15, заключается в более простом способе вычисления

$$\inf_{f_2 \in \mathcal{F}_{2,x,\theta_2}} I_{1,2}(x, f_1, f_2, \theta_1, \theta_2). \quad (5.24)$$



Сплошная линия — график плотности f_1 усеченного стандартного нормального распределения, заданного на интервале $[-3, 3]$; пунктирная линия — график плотности f_2^* , которая задается формулой (5.13), для различных значений $\eta_2(x, \theta_2) = \mp 0.5, \mp 0.4, \mp 0.3$.

Рисунок 5.1 — Иллюстрация к примеру 1 параграфа 5.1

Это упрощение оказывает существенное влияние на скорость численного решения задачи, а также на его устойчивость. Отметим, что результат из [11] требует численного решения системы из двух нелинейных уравнений (например, с использованием метода Ньютона) каждый раз при вычислении (5.9) для всех точек x , использующихся при получении оптимального плана, максимизирующего критерий (5.7), и для всех значений параметров $\theta_2 \in \Theta_2$, задействованных при минимизации в упрощенной версии критерия (5.8) из [11]. С вычислительной точки зрения это довольно серьезная задача, сходимость которой зависит от выбора начального приближения для итерационной процедуры. С другой стороны, теорема 15 сводит задачу вычисления (5.24) к решению одного нелинейного уравнения на заданном интервале. Эта задача легко решается, например, с помощью метода бисекции или метода золотого сечения.

В ходе численных экспериментов (смотри параграф 5.4) было обнаружено, что для многих значений независимой переменной x значение функции (5.9) не могло быть получено, так как метод Ньютона не сходил к такому решению системы, которое удовлетворяло бы условию $\mu + \lambda(y - \eta_2(x, \theta_2)) < 0$, причем такая ситуация наблюдалась даже в тех случаях, когда начальное приближение в методе Ньютона бралось очень близко к ожидаемому решению (полученному с помощью подхода из теоремы 15). В тех же случаях, когда метод из [11] давал положительный результат, метод из теоремы 15 работал чуть более чем в два раза быстрее (смотри последний пример в § 5.4).

5.2 Теоремы эквивалентности для полу-параметрических оптимальных планов

В этом параграфе мы сформулируем теоремы эквивалентности для критериев (5.5) и (5.7):

Теорема 16. Пусть выполнены предположения теоремы 15 и $\inf_{\theta_2 \in \Theta_2}$ в критериях (5.5) и (5.7) достигается в единственной точке $\theta_2^* \in \Theta_2$ для оптимального плана ξ^* .

(а) Тогда план ξ^* является $SKL_{(a)}$ -оптимальным тогда и только тогда, когда для всех $x \in \mathcal{X}$ выполнено неравенство:

$$I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2^*) - \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2^*) \xi^*(dx) \leq 0. \quad (5.25)$$

При этом равенство в последнем выражении достигается только в опорных точках оптимального плана ξ^* . Здесь $I_{1,2}(x, f_1, f_2, \bar{\theta}_1, \theta_2)$ определяется в (5.1) и

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1, f_2^*, \bar{\theta}_1, \theta_2) \xi^*(dx),$$

$$f_2^*(y, x, \theta_2) = \frac{f_1(y, x, \bar{\theta}_1)}{1 + \lambda(y - \eta_2(x, \theta_2))},$$

а λ находится как корень уравнения (5.10).

(b) Тогда план ξ^* является $SKL_{(b)}$ -оптимальным тогда и только тогда, когда для всех $x \in \mathcal{X}$ выполнено неравенство:

$$I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) - \int_{\mathcal{X}} I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) \xi^*(dx) \leq 0. \quad (5.26)$$

При этом равенство в последнем выражении достигается только в опорных точках оптимального плана ξ^* . Здесь

$$\theta_2^* = \arg \inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} I_{1,2}(x, f_1^*, f_2, \bar{\theta}_1, \theta_2^*) \xi^*(dx),$$

$$f_1^*(y, x, \bar{\theta}_1) = \frac{f_2(y, x, \theta_2) \exp(-\lambda y)}{\int f_2(y, x, \theta_2) \exp(-\lambda y) dy}$$

и λ находится как корень уравнения (5.16).

Эта теорема является прямым следствием теоремы эквивалентности для KL -оптимальных планов из работы [10]. Часть (а) утверждает, что критерий $K_{(a)}(\xi, \bar{\theta}_1)$ является частным случаем критерия KL -оптимальности для дискриминации между плотностью $f_1(y, x, \bar{\theta}_1)$ и плотностью $f_2^*(y, x, \theta_2)$, которая задается формулой (5.13). Часть (b) утверждает, что критерий $K_{(b)}(\xi, \bar{\theta}_1)$ является частным случаем критерия KL -оптимальности для дискриминации между плотностью $f_1^*(y, x, \bar{\theta}_1)$, которая задается формулой (5.15), и плотностью $f_2(y, x, \theta_2)$.

5.3 Связь полу-параметрических критериев с критерием T -оптимальности

В этом параграфе мы докажем две теоремы, связывающие полу-параметрические критерии с критерием T -оптимальности. Следующий результат опи-

сывает достаточное условие для того, чтобы T -оптимальный план был также $SKL_{(a)}$ -оптимальным планом.

Теорема 17. Пусть выполнены предположения пункта (а) теоремы 15 и плотность $f_1(y, x, \bar{\theta}_1)$ удовлетворяет условию:

$$f_1(y, x, \bar{\theta}_1) = g(y - \eta_1(x, \bar{\theta}_1)), \quad (5.27)$$

где функция g есть симметричная плотность, заданная на конечном интервале $[-a, a]$, то есть плотность $f_1(y, x, \bar{\theta}_1)$ имеет носитель

$$[-a + \eta_1(x, \bar{\theta}_1), a + \eta_1(x, \bar{\theta}_1)].$$

Тогда T -оптимальный план также является $SKL_{(a)}$ -оптимальным.

Доказательство. Пусть $\Delta_\eta = \eta_2 - \eta_1$. Используя представление (5.27), имеем следующее выражение для расстояния Кульбака-Лейблера из (5.1):

$$\begin{aligned} F(\eta_1, \eta_2, \lambda) &= \int_{-a+\eta_1}^{a+\eta_1} \log \{1 + \lambda(y - \eta_2)\} f_1(y) dy \\ &= \int_{-a}^a \log \{1 + \lambda(t - \Delta_\eta)\} g(t) dt =: F(\Delta_\eta, \lambda). \end{aligned}$$

Аналогично для левой части уравнения (5.10) мы можем получить следующее представление

$$G(\eta_1, \eta_2, \lambda) = \int_{-a+\eta_1}^{a+\eta_1} \frac{f_1(y)}{1 + \lambda(y - \eta_2)} dy = \int_{-a}^a \frac{g(t)}{1 + \lambda(t - \Delta_\eta)} dt.$$

По предположению данной теоремы существует ненулевой корень уравнения

$$G(\eta_1, \eta_2, \lambda) = 1 \text{ при всех } x \in \mathcal{X} \text{ и при всех } \theta_2 \in \Theta_2.$$

Обозначим этот ненулевой корень $\bar{\lambda}(\Delta_\eta)$. Также введем обозначение

$$F(\Delta_\eta) = F(\Delta_\eta, \bar{\lambda}(\Delta_\eta)).$$

Заметим, что функция

$$h(t, \Delta_\eta) = \frac{g(t)}{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)}$$

является плотностью со средним Δ_η , так как функция $g(t)$ симметрична на интервале и имеет нулевое среднее. Посчитаем теперь производную функции $F(\Delta_\eta)$ по Δ_η :

$$\begin{aligned} \frac{\partial F(\Delta_\eta)}{\partial \Delta_\eta} &= \frac{\partial}{\partial \Delta_\eta} \int_{-a}^a \log \{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)\} g(t) dt \\ &= \int_{-a}^a \left[\frac{\partial \bar{\lambda}(\Delta_\eta)}{\partial \Delta_\eta} (t - \Delta_\eta) - \bar{\lambda}(\Delta_\eta) \right] \frac{g(t)}{1 + \bar{\lambda}(\Delta_\eta)(t - \Delta_\eta)} dt \\ &= \frac{\partial \bar{\lambda}(\Delta_\eta)}{\partial \Delta_\eta} \int_{-a}^a (t - \Delta_\eta) h(t, \Delta_\eta) dt - \bar{\lambda}(\Delta_\eta) \int_{-a}^a h(t, \Delta_\eta) dt = -\bar{\lambda}(\Delta_\eta). \end{aligned}$$

В последней формуле первый интеграл равен нулю, так как функция $h(t, \Delta_\eta)$ имеет среднее, равное Δ_η , а второй интеграл равен единице, так как $h(t, \Delta_\eta)$ — это плотность. Из леммы 4 получаем, что

$$\begin{aligned} \text{если } \eta_1 > \eta_2, \text{ то } \Delta_\eta < 0 &\Rightarrow \bar{\lambda}(\Delta_\eta) > 0 \Rightarrow F(\Delta_\eta) \text{ убывает,} \\ \text{если } \eta_1 < \eta_2, \text{ то } \Delta_\eta > 0 &\Rightarrow \bar{\lambda}(\Delta_\eta) < 0 \Rightarrow F(\Delta_\eta) \text{ возрастает.} \end{aligned}$$

Также отметим, что из симметричности функции g следует симметричность функции F , то есть $F(\Delta_\eta) = F(-\Delta_\eta)$. Пусть

$$\xi_T^* = \begin{bmatrix} x_1^* & \dots & x_k^* \\ \omega_1^* & \dots & \omega_k^* \end{bmatrix}$$

есть T -оптимальный план. Определим

$$\theta_2^* = \arg \inf_{\theta_2} \int [\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2 \xi_T^*(dx).$$

Из теоремы эквивалентности для T -оптимальных планов из работы [7] следует, что для всех $x \in \mathcal{X}$ выполняется соотношение

$$|\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)| \leq \epsilon = |\eta_1(x_1^*, \bar{\theta}_1) - \eta_2(x_1^*, \theta_2^*)| = \dots = |\eta_1(x_n^*, \bar{\theta}_1) - \eta_2(x_n^*, \theta_2^*)|.$$

Учитывая последнее соотношение и представление (5.13) для “оптимальной” плотности f_2^* , при всех $x \in \mathcal{X}$ получаем

$$\begin{aligned} I_{1,2}(x, f_1, f_2^*, \theta_1, \theta_2^*) &= F(\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)) \\ &= F(|\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2^*)|) \leq F(\epsilon) = \sum_{i=1}^k \omega_i^* F(\epsilon) \\ &= \sum_{i=1}^k \omega_i^* F(\eta_1(x_i^*, \bar{\theta}_1) - \eta_2(x_i^*, \theta_2^*)) = \int I_{1,2}(x, f_1, f_2^*, \theta_1, \theta_2^*) d\xi^*(x). \end{aligned}$$

Второе равенство следует из симметричности функции $F(\Delta_\eta)$, а неравенство является следствием монотонности $F(\Delta_\eta)$. По пункту (а) теоремы эквивалентности 16 из последнего неравенства следует, что план ξ^* является $SKL_{(a)}$ -оптимальным. $\square$

Похожий результат можно получить для $SKL_{(b)}$ -оптимальных планов. Пусть $f_2(y, x, \theta_2)$ — это плотность нормального распределения $\mathcal{N}(\eta_2(x, \theta_2), v_2^2(x, \theta_2))$ и $f_1(y, x, \bar{\theta}_1)$ — это плотность нормального распределения $\mathcal{N}(\eta_1(x, \bar{\theta}_1), v_2^2(x, \theta_2))$. План, доставляющий максимум величине

$$\inf_{\theta_2 \in \Theta_2} \int_{\mathcal{X}} \frac{[\eta_1(x, \bar{\theta}_1) - \eta_2(x, \theta_2)]^2}{v_2^2(x, \theta_2)} \xi(dx), \quad (5.28)$$

является KL -оптимальным для дискриминации между двумя плотностями нормального распределения с одинаковой функцией для дисперсии (смотри работу [9]). Следующая теорема показывает, что этот план также является $SKL_{(b)}$ -оптимальным.

Теорема 18. Пусть плотность $f_2(y, x, \theta_2)$ — это плотность нормального распределения со средним $\eta_2(x, \theta_2)$ и дисперсией $v_2^2(x, \theta_2)$. Тогда наилучшая аппроксимация $f_1^*(y, x, \bar{\theta}_1)$ будет плотностью нормального распределения со средним $\eta_1(x, \bar{\theta}_1)$ и дисперсией $v_2^2(x, \theta_2)$ и оптимальный план, доставляющий максимум выражению (5.28), также будет $SKL_{(b)}$ -оптимальным. Последнее утверждение верно также и в обратную сторону.

Доказательство. Для простоты обозначим

$$\eta_1 = \eta_1(x, \bar{\theta}_1), \quad \eta_2 = \eta_2(x, \theta_2), \quad v_2 = v_2(x, \theta_2).$$

Из части (b) теоремы 15 следует, что

$$\begin{aligned} f_1^*(y, x, \bar{\theta}_1) &\propto f_2(y) \exp(-\lambda y) = \frac{1}{\sqrt{2\pi}v_2} \exp\left[-\frac{\{y - \eta_2\}^2}{2v_2^2} - \lambda y\right] \\ &= \frac{1}{\sqrt{2\pi}v_2} \exp\left[-\frac{\{y - (\eta_2 - v_2^2\lambda)\}^2}{2v_2^2}\right] \exp\left[-\eta_2\lambda + \frac{v_2^2\lambda^2}{2}\right]. \end{aligned}$$

Из того, что $f_1^*(y, x, \bar{\theta}_1)$ — это плотность со средним η_1 , получаем

$$\lambda = \frac{\eta_2 - \eta_1}{v_2^2}, \quad \mu' = \exp\left[\eta_2\lambda - \frac{v_2^2\lambda^2}{2}\right],$$

откуда следует, что $f_1^*(y, x, \bar{\theta}_1)$ является плотностью нормального распределения со средним η_1 и дисперсией v_2^2 . Теперь нетрудно увидеть, что расстояние Кульбака-Лейблера между плотностями $f_1^*(y, x, \bar{\theta}_1)$ и $f_2(y, x, \theta_2)$ равно $[\eta_1 - \eta_2]^2/v_2^2$, откуда получаем критерий (5.28). $\square$

5.4 Примеры

Для численного нахождения $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов можно воспользоваться любым методом для KL -оптимальных планов, подставляя “оптимальные” функции $f_2^*(y, x, \theta_2)$ и $f_1^*(y, x, \bar{\theta}_1)$, задаваемые соотношениями из теоремы 15, в критерии $K_{(a)}(\xi, \bar{\theta}_1)$ и $K_{(b)}(\xi, \bar{\theta}_1)$, так как эти критерии сводятся к критерию KL -оптимальности. Эффективные алгоритмы для численного нахождения KL -оптимальных планов описаны в главе 4 и в работе [16]. Мы будем использовать версию алгоритма Аткинсона-Федорова из [6], адаптированную для вычисления полу-параметрических планов, которая не является самой эффективной, но достаточно проста в реализации.

Пусть δ — некоторая заранее заданная маленькая положительная константа. Из леммы 4 и неравенства (5.11) предлагается искать численное решение уравнения (5.10) следующим образом:

1. Если $\eta_1(x, \bar{\theta}_1) = \eta_2(x, \theta_2)$, то $\lambda = 0$;
2. Если $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, то решение ищем на интервале

$$\Lambda^- = [-1/(y_{x, \max} - \eta_2(x, \theta_2)), -\delta];$$

3. Если $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$, то решение ищем на интервале

$$\Lambda^+ = [\delta, -1/(y_{x, \min} - \eta_2(x, \theta_2))].$$

Аналогично может быть найдено решение уравнения (5.16). Будем искать $\lambda > 0$, если $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, так как такое λ смещает среднее заданной плотности $f_2(y, x, \theta_2)$ влево. Из тех же соображений будем искать $\lambda < 0$, если $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$. Если δ — это достаточно маленькая положительная константа и β — это достаточно большая положительная константа, мы можем предполагать, что решение уравнения (5.16) находится в интервале $[-\beta, +\beta]$. Тогда численное решение уравнения (5.16) предлагается искать следующим образом:

1. Если $\eta_1(x, \bar{\theta}_1) = \eta_2(x, \theta_2)$, то $\lambda = 0$;
2. Если $\eta_1(x, \bar{\theta}_1) < \eta_2(x, \theta_2)$, то решение ищем на интервале $\Lambda^+ = [+ \delta, + \beta]$;
3. Если $\eta_1(x, \bar{\theta}_1) > \eta_2(x, \theta_2)$, то решение ищем на интервале $\Lambda^- = [-\beta, -\delta]$.

Теперь рассмотрим три численных примера, в которых T -оптимальные планы и полу-параметрические оптимальные планы не совпадают.

5.4.1 Полу-параметрические оптимальные планы для дискриминации ЕМАХ-модели и модели Михаэлиса–Ментен

Рассмотрим пару моделей из работы [10]:

$$\begin{aligned} \eta_1(x, \theta_1) &= \theta_{1,1}x + \frac{\theta_{1,2}x}{x + \theta_{1,3}}, \\ \eta_2(x, \theta_2) &= \frac{\theta_{2,1}x}{x + \theta_{2,2}}, \end{aligned} \tag{5.29}$$

при $x \in [0.1, 5]$. Пусть параметры первой модели $\bar{\theta}_1 = (1, 1, 1)$ фиксированы. Для этого случая найдены четыре разных типа оптимальных планов:

1. T -оптимальный план,
2. KL -оптимальный план для дискриминации двух логнормальных плотностей со средними $\eta_1(x, \bar{\theta}_1)$ и $\eta_2(x, \theta_2)$ и фиксированными дисперсиями $v_1^2(x, \bar{\theta}_1) = v_2^2(x, \theta_2) = 0.1$ (далее просто KL -оптимальный план),
3. $SKL_{(a)}$ -оптимальный план для усеченной плотности логнормального распределения $f_1(y, x, \bar{\theta}_1)$ с параметрами $\mu_1(x, \bar{\theta}_1)$ и $\sigma_1^2(x, \bar{\theta}_1)$,

Таблица 11 — T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальные планы для дискриминации моделей (5.29)

Тип плана	ξ^*			θ_2^*
T -оптимальный	0.508 0.580	2.992 0.298	5.000 0.122	(22.564, 14.637)
KL -оптимальный	0.218 0.629	2.859 0.260	5.000 0.111	(21.112, 13.436)
$SKL_{(a)}$ -оптимальный	0.454 0.531	2.961 0.344	5.000 0.125	(22.045, 14.197)
$SKL_{(b)}$ -оптимальный	0.509 0.611	2.994 0.273	5.000 0.116	(22.824, 14.857)

4. $SKL_{(b)}$ -оптимальный план для усеченной плотности логнормального распределения $f_2(y, x, \theta_2)$ с параметрами $\mu_2(x, \theta_2)$ и $\sigma_2^2(x, \theta_2)$,

где

$$\mu_i(x, \theta) = \log [\eta_i(x, \theta)] - \frac{1}{2} \log [1 + v_i^2(x, \theta)/\eta_i^2(x, \theta)], \quad i = 1, 2,$$

$$\sigma_i^2(x, \theta) = \log [1 + v_i^2(x, \theta)/\eta_i^2(x, \theta)], \quad i = 1, 2.$$

Эти усеченные плотности заданы на интервалах

$$[Q_1(0.0001, x, \bar{\theta}_1), Q_1(0.9999, x, \bar{\theta}_1)], [Q_2(0.0001, x, \theta_2), Q_2(0.9999, x, \theta_2)]$$

соответственно, где $Q_i(p, x, \theta)$ — это функция квантилей для обычного логнормального распределения со средним $\eta_i(x, \theta)$ и дисперсией $v_i^2(x, \theta) = 0.1$. Заметим, что из-за усечения по краям $\eta_1(x, \bar{\theta}_1)$ и $\eta_2(x, \theta_2)$ не совпадают с реальными средними плотностей $f_1(y, x, \bar{\theta}_1)$ и $f_2(y, x, \theta_2)$, но достаточно близки к ним. В таблице 11 представлены найденные численно планы для дискриминации, оптимальные относительно четырех критериев, описанных выше, а также “оптимальные” значения параметров θ_2^* для второй модели, соответствующие значениям, на которых в критериях достигается минимум по θ_2 . Отметим, что $SKL_{(b)}$ -оптимальный план сосредоточен почти в тех же точках, что и T -оптимальный план, но веса у него другие. На рисунке 5.2 показаны функции влияния для всех четырех планов.

В таблице 12 представлены эффективности T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов относительно T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -критериев опти-

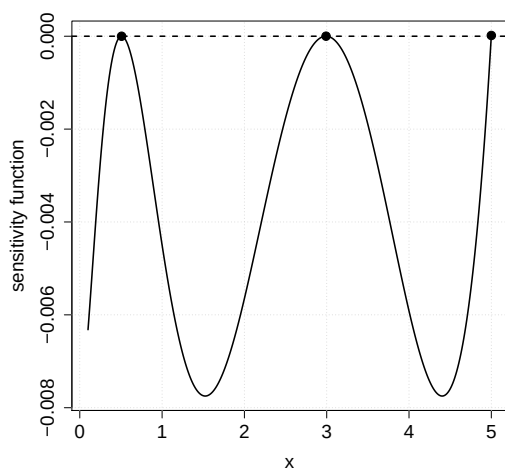
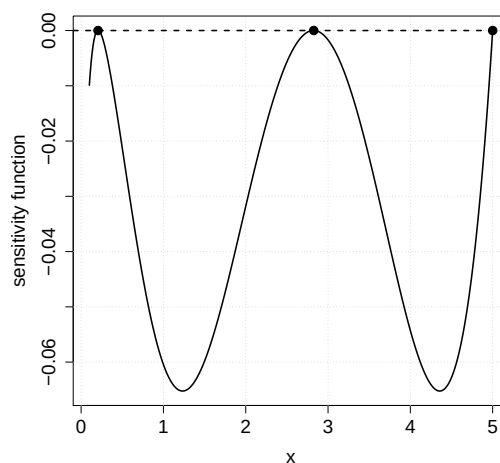
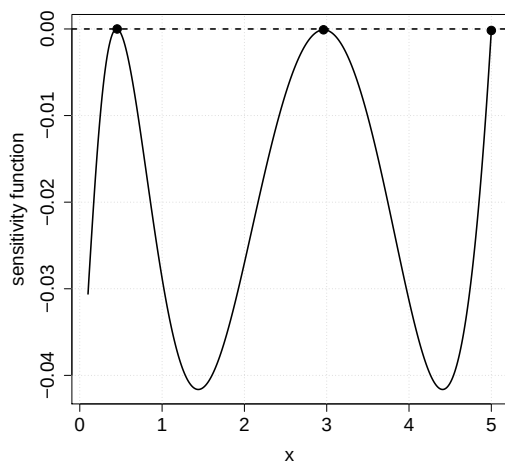
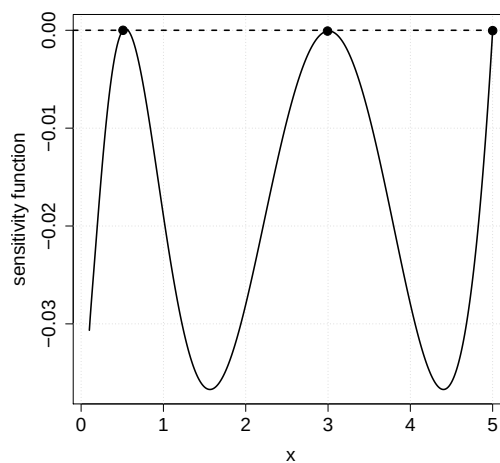
а) T -оптимальный.б) KL -оптимальный.в) $SKL_{(a)}$ -оптимальный.г) $SKL_{(b)}$ -оптимальный.

Рисунок 5.2 — Графики функций влияния для T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов дискриминации из таблицы 11

Таблица 12 — Эффективности T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов для дискриминации моделей (5.29) относительно различных критериев оптимальности

$K(\xi) \setminus \xi$	T	KL	$SKL_{(a)}$	$SKL_{(b)}$
T	1	0.321	0.741	0.830
KL	0.739	1	0.796	0.650
$SKL_{(a)}$	0.552	0.544	1	0.454
$SKL_{(b)}$	0.876	0.254	0.633	1

мальности. Так, например, значение 0.321 в первой строке таблицы обозначает эффективность KL -оптимального плана относительно критерия T -оптимальности.

5.4.2 Полу-параметрические оптимальные планы для дискриминации экспоненциальной модели и модели Михаэлиса–Ментен

Рассмотрим похожий пример с другой функцией $\eta_1(x, \theta_1)$ из работы [37]. Необходимо дискриминировать модели

$$\begin{aligned}\eta_1(x, \theta_1) &= \theta_{1,1} \{1 - \exp(-\theta_{1,2}x)\}, \\ \eta_2(x, \theta_2) &= \frac{\theta_{2,1}x}{\theta_{2,2} + x},\end{aligned}\tag{5.30}$$

при $x \in [0.1, 5]$. Фиксируем параметры $\bar{\theta}_1 = (1, 1)$ первой модели (5.30). Для этого случая найдем те же типы оптимальных планов: T -оптимальный, KL -оптимальный (для ошибок, имеющих логнормальное распределение), $SKL_{(a)}$ -оптимальный и $SKL_{(b)}$ -оптимальный для ошибок, имеющих усеченные логнормальные распределения. Дисперсии ошибок для случая дискриминации с помощью KL -критерия: $v_1^2(x, \bar{\theta}_1) = v_2^2(x, \theta_2) = 0.02$, для $SKL_{(a)}$ -критерия: $v_1^2(x, \bar{\theta}_1) = 0.02$, для $SKL_{(b)}$ -критерия: $v_2^2(x, \theta_2) = 0.02$. В таблице 13 представлены оптимальные планы всех четырех видов, а также соответствующие оптимальные значения для параметров θ_2^* второй модели. В таблице 14 находятся значения относительных эффективностей. В целом результаты напоминают предыдущий пример. На рисунке 5.3 показаны графики функций влияния для T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов.

5.4.3 Сравнение подходов

В общем случае сравнить предложенный в теореме 15 подход с изначально описанным в работе [11] не представляется возможным в связи с вычислительными сложностями, о которых говорится в конце параграфа 5.1. Алгоритм из [11] часто не находит SKL -оптимальный план, в частности, для моделей из предыдущих двух примеров. Для того, чтобы провести сравнительный анализ скорости работы обоих методов, рассмотрим более простой пример, для которого оптимальный план получилось найти методом из [11]. Для пары моде-

Таблица 13 — T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальные планы для дискриминации между моделями (5.30)

Тип плана	ξ^*			θ_2^*
T -оптимальный	0.308	2.044	5.000	(1.223, 0.948)
	0.316	0.428	0.256	
KL -оптимальный	0.136	1.902	5.000	(1.244, 1.020)
	0.297	0.457	0.252	
$SKL_{(a)}$ -оптимальный	0.395	2.090	5.000	(1.216, 0.920)
	0.396	0.355	0.249	
$SKL_{(b)}$ -оптимальный	0.308	2.044	5.000	(1.225, 0.956)
	0.289	0.458	0.253	

Таблица 14 — Эффективности T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов для дискриминации моделей (5.30) относительно различных критериев оптимальности

$K(\xi) \setminus \xi$	T	KL	$SKL_{(a)}$	$SKL_{(b)}$
T	1	0.266	0.663	0.858
KL	0.786	1	0.565	0.879
$SKL_{(a)}$	0.407	0.346	1	0.388
$SKL_{(b)}$	0.882	0.396	0.608	1

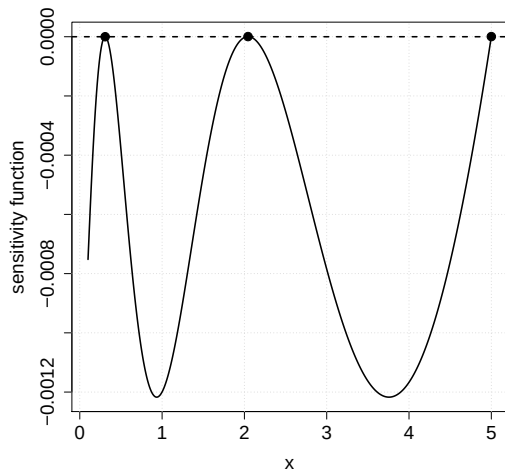
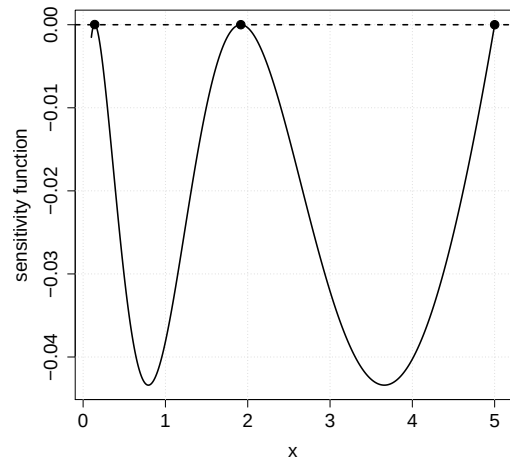
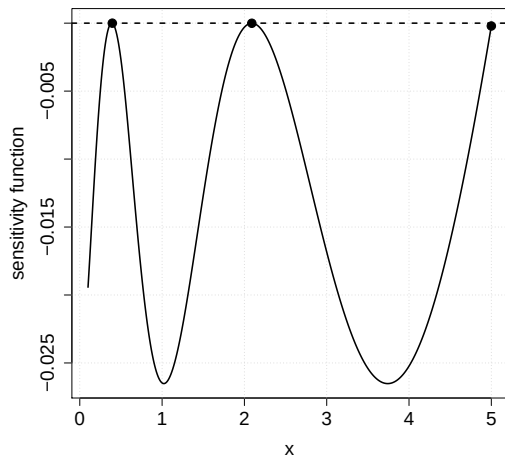
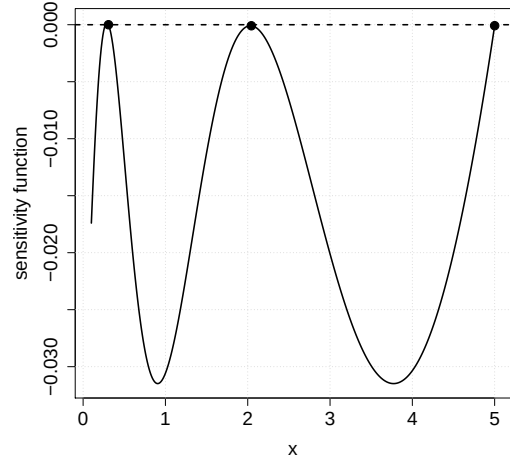
а) T -оптимальный.б) KL -оптимальный.в) $SKL_{(a)}$ -оптимальный.г) $SKL_{(b)}$ -оптимальный.

Рисунок 5.3 — Графики функций влияния для T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов дискриминации из таблицы 13

лей (5.30) построим $SKL_{(a)}$ -оптимальный план. В качестве плотности $f_1(y, x, \bar{\theta}_1)$ возьмем плотность случайной величины $\eta_1(x, \theta_1) + (\varepsilon - m)$, где ε имеет усеченное логнормальное распределение с параметрами $\mu_1 = 0$ и $\sigma_1^2 = 1$ на интервале от $Q(0.001)$ до $Q(0.9)$ (здесь $Q(p)$ функция квантилей для стандартного логнормального распределения). Константа m выбирается для каждого x таким образом, чтобы выполнялось равенство $\mathbb{E}[\varepsilon - m] = 0$. В итоге $SKL_{(a)}$ -оптимальный план задается как

$$\xi^* = \begin{bmatrix} 0.308 & 2.044 & 5.000 \\ 0.323 & 0.415 & 0.262 \end{bmatrix}. \quad (5.31)$$

Отметим, что этот план сосредоточен в тех же точках, что и T -оптимальный план, но имеет другие веса. Для того, чтобы получить этот план, алгоритму, основанному на теореме 15, понадобилось 540 секунд. У алгоритма из работы [11] на решение той же задачи потребовалось 1230 секунд (в обоих случаях использовался адаптированный алгоритм Аткинсона–Федорова, отличался только метод вычисления значения критерия). То есть даже в этом простом случае разница в скорости вычислений является существенной.

Заключение

В диссертационной работе рассмотрен критерий T -оптимальности и связанные с ним критерии для дискриминации конкурирующих регрессионных моделей. Получены следующие результаты.

В аналитическом виде найдены T -оптимальные планы для дискриминации полиномиальных моделей и аналогичных моделей, содержащих дополнительное дробно-рациональное слагаемое.

Предложен метод построения байесовских T_P -оптимальных планов путем их сведения к локально оптимальным планам. Разработан двухэтапный алгоритм для нахождения локальных T_P -оптимальных планов, состоящий в чередовании обновления носителя плана и оптимизации по его весам. Доказана сходимость этого алгоритма. Для оптимизации по весам предложено две эффективные численные процедуры. Проведено сравнение наиболее часто используемого в литературе алгоритма с новым, выявившее значительное преимущество последнего. Создан пакет для языка R, включающий реализацию нового алгоритма и примеры из работы.

Сформулирован байесовский критерий KL_P -оптимальности и теорема эквивалентности для него. Результаты, касающиеся байесовских T_P -оптимальных планов, обобщены на случай байесовских KL_P -оптимальных планов, реализация обобщенного алгоритма и новые численные примеры также включены в пакет для R.

Предложен упрощенный метод численного нахождения полу-параметрических оптимальных планов. Доказаны две теоремы, связывающие полу-параметрические критерии с критерием T -оптимальности.

Возможные направления для дальнейшей работы: поиск аналитических представлений T -оптимальных планов для дискриминации двух дробно-рациональных моделей, например, ЕМАХ-модели и модели Михаэлиса–Ментен, а также исследование комбинированных критериев для дискриминации и оценки неизвестных параметров конкурирующих моделей.

Список литературы

1. *Stigler S.* Optimal experimental design for polynomial regression. // *Journal of the American Statistical Association*. — 1971. — Vol. 66. — Pp. 311–318.
2. *Studden W.J.* D_s -optimal designs for polynomial regression using continued fractions // *Annals of Statistics*. — 1980. — Vol. 8, no. 5. — Pp. 1132–1141.
3. *Dette H., Haller G.* Optimal designs for the identification of the order of a Fourier regression. // *Annals of Statistics*. — 1998. — Vol. 26. — Pp. 1496–1521.
4. *Song D., Wong W.K.* On the construction of G_{rm} -optimal designs. // *Statistica Sinica*. — 1999. — Vol. 9. — Pp. 263–272.
5. *May C., Tommasi C.* Model selection and parameter estimation in non-linear nested models: A sequential generalized DKL-optimum design // *Statistica Sinica*. — 2014. — Vol. 24, no. 1. — Pp. 63–82.
6. *Atkinson A.C., Fedorov V.V.* The designs of experiments for discriminating between two rival models // *Biometrika*. — 1975. — Vol. 62. — Pp. 57–70.
7. *Dette H., Titoff S.* Optimal discrimination designs. // *Annals of Statistics*. — 2009. — Vol. 37, no. 4. — Pp. 2056–2082.
8. *Dette H., Melas V.B., Shpilev P.* T-optimal designs for discrimination between two polynomial models // *Annals of Statistics*. — 2012. — Vol. 40, no. 1. — Pp. 188–205.
9. *Ucinski D., Bogacka B.* T-Optimum Designs for Multiresponse Dynamic Heteroscedastic Models. // *Proceedings of the 7th International Workshop on Model-Oriented Design and Analysis* / Ed. by A. Di Bucchianico, Läuter H., Wynn H.P. — Springer, 2004. — June. — Pp. 191–199.
10. *López-Fidalgo J., Tommasi C., Trandafir P.C.* An optimal experimental design criterion for discriminating between non-normal models // *Journal of the Royal Statistical Society, Series B*. — 2007. — Vol. 69. — Pp. 231–242.
11. *Otsu T.* Optimal experimental design criterion for discriminating semi-parametric models // *Journal of Statistical Planning and Inference*. — 2008. — Vol. 138. — Pp. 4141–4150.

12. *Atkinson A.C., Fedorov V.V.* Optimal design: Experiments for discriminating between several models // *Biometrika*. — 1975. — Vol. 62. — Pp. 289–303.
13. *Braess D., Dette H.* Optimal discriminating designs for several competing regression models // *Annals of Statistics*. — 2013. — Vol. 41, no. 2. — Pp. 897–922.
14. *Гученко Р.А., Мелас В.Б.* T -оптимальные планы для дискриминации дробно-рациональных и полиномиальных моделей // *Вестник СПбГУ. Математика. Механика. Астрономия*. — 2017 — Т.4(62), №2 — С.208-219.
15. *Dette H., Melas V.B., Guchenko R.* Bayesian T -optimal discriminating designs. // *Annals of Statistics*. — 2015. — Vol. 43, no. 5. — Pp. 1959–1985.
16. *Dette H., Guchenko R., Melas V.B.* Efficient computation of Bayesian optimal discriminating designs. // *Journal of Computational and Graphical Statistics*. — 2017. — Vol. 26, no. 2. — Pp. 424–433.
17. *Guchenko R.* rodd: Optimal Discriminating Designs. — R package version 0.2-1. — 2016. — <http://CRAN.R-project.org/package=rodd>.
18. *Dette H., Pepelyshev A., Zhigljavsky A.* Optimal designs in regression with correlated errors // *Ann. Statist.* — 2016. — Vol. 44. — Pp. 113–152.
19. *Ермаков С.М., Жиглявский А.А.* Математическая теория оптимального эксперимента. — Москва: Наука, 1987.
20. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. — 2 edition. — New York, NY, USA: Springer, 2009.
21. *Bishop C.M.* Pattern Recognition and Machine Learning. — Springer, 2006.
22. *Мелас В.Б., Шнилев П.В.* Планирование и анализ для регрессионных моделей. — Санкт-Петербург: Издательство Санкт-Петербургского государственного университета, 2014. — 94с.
23. *Atkinson A., Donev A., Tobias R.* Optimum Experimental Designs, with SAS (Oxford Statistical Science Series). — 2nd edition. — Oxford University Press, USA, 2007.

24. *Kiefer J.* General equivalence theory for optimum designs (approximate theory). // *Annals of Statistics*. — 1974. — Vol. 2, no. 5. — Pp. 849–879.
25. *Pukelsheim F.* Optimal Design of Experiments. — Philadelphia: SIAM, 2006.
26. *Ермаков С.М., Бродский В.З., Жигляевский А.А. и др.* Математическая теория планирования эксперимента. — Москва: Наука, 1983.
27. *Chernoff H.* Locally optimal designs for estimating parameters // *Annals of Mathematical Statistics*. — 1953. — Vol. 24. — Pp. 586–602.
28. *Pshenichny B.N.* Necessary Conditions of an Extremum. — New York: Marcel Dekker, 1971.
29. *Дзядык В.К.* Введение в теорию равномерного приближения функций полиномами. — Москва: Наука, 1977. — 512с.
30. *Ахиезер Н.И.* Лекции по теории аппроксимации. — Москва: Наука, 1965. — 406с.
31. *Бернштейн С.Н.* Экстремальные свойства полиномов и наилучшее приближение непрерывных функций одной вещественной переменной. — Ленинград, Москва: Гостехиздат, 1937. — 203с.
32. *Dette H., Melas V.B., Shpilev P.* Robust T -optimal discriminating designs // *Annals of Statistics*. — 2013. — Vol. 41, no. 4. — Pp. 1693–1715.
33. *Ratkowsky D.A.* Handbook of Nonlinear Regression Models. — New York: Dekker, 1990.
34. *Han C., Chaloner K.* D - and c -optimal designs for exponential regression models used in pharmacokinetics and viral dynamics. // *Journal of Statistical Planning and Inference*. — 2003. — Vol. 115. — Pp. 585–601.
35. *Pinheiro J., Bretz F., Branson M.* Analysis of dose-response studies: Modeling approaches. // Dose Finding in Drug Development / Ed. by N. Ting. — New York: Springer-Verlag, 2006. — Pp. 146–171.
36. *Tommasi C., López-Fidalgo J.* Bayesian optimum designs for discriminating between models with any distribution // *Computational Statistics and Data Analysis*. — 2010. — Vol. 54, no. 1. — Pp. 143–150.

37. *Wiens D. P.* Robust discrimination designs. // *Journal of the Royal Statistical Society, Ser. B.* — 2009. — Vol. 71. — Pp. 805–829.

Список рисунков

2.1	Иллюстрации к теоремам эквивалентности для оптимальных планов ξ_T^* и ξ_{SB}^*	43
3.1	Графики функций $\Psi(x, \xi_{\sigma_2}^*)$ для планов из таблицы 4	64
3.2	График функции $\Psi(x, \tilde{\xi}_1)$ после первой итерации алгоритма 3 . .	66
3.3	Графики функций $\Psi(x, \xi_{\sigma_2}^*)$ для планов из таблицы 6	68
4.1	Графики функций $\Psi_{KL}(x, \xi_{(i)}^*)$ для планов из таблицы 8	86
4.2	Графики функций $\Psi_{KL}(x, \xi_{(i)}^*)$ для планов из таблицы 10	89
5.1	Иллюстрация к примеру 1 параграфа 5.1	98
5.2	Графики функций влияния для T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов дискриминации из таблицы 11 . . .	107
5.3	Графики функций влияния для T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов дискриминации из таблицы 13 . . .	110

Список таблиц

1	Опорные точки T -оптимального плана для дискриминации моделей (2.1)	34
2	Опорные точки T -оптимального плана для дискриминации моделей (2.9)	37
3	Эффективность байесовского плана ξ_{SB}^* относительно локально оптимальных критериев с фиксированными параметрами $\theta_{1,\gamma}^{(i)}$. .	42
4	Байесовские T_P -оптимальные планы для дискриминации моделей (3.23)	63
5	Время работы различных алгоритмов в секундах	63
6	Байесовские T_P -оптимальные планы для дискриминации моделей (3.26)	67
7	Локальные KL_P -оптимальные планы для дискриминации моделей (4.16) при логнормальных откликах с дисперсиями из (4.17)	83
8	Байесовские KL_P -оптимальные планы для дискриминации моделей (4.18) при логнормальных откликах с дисперсиями из (4.17)	85
9	Относительные эффективности планов из таблицы 8 при различных предположениях о распределении ошибок	85
10	Байесовские KL_P -оптимальные планы для дискриминации моделей (4.21) при логнормальных откликах с дисперсиями из (4.23)	88
11	T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальные планы для дискриминации моделей (5.29)	106
12	Эффективности T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов для дискриминации моделей (5.29) относительно различных критериев оптимальности	107
13	T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальные планы для дискриминации между моделями (5.30)	109

14	Эффективности T -, KL -, $SKL_{(a)}$ - и $SKL_{(b)}$ -оптимальных планов для дискриминации моделей (5.30) относительно различных критериев оптимальности	109
----	--	-----