

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

На правах рукописи

КОРОБЕЙНИКОВ Антон Иванович

**Исследование специальных моделей кривых
дожития в условиях неполных данных**

05.13.18 – Математическое моделирование,
численные методы и комплексы программ

ДИССЕРТАЦИЯ

на соискание ученой степени

кандидата физико-математических наук

Научный руководитель

доктор физ.-мат. наук, профессор

ЕРМАКОВ Сергей Михайлович

Санкт-Петербург – 2010

Содержание

Введение	5
Глава 1. Оценки типа максимального правдоподобия	13
1.1. Модель интервального цензурирования смешанного типа	13
1.2. Построение оценок	17
1.3. Состоятельность оценок	18
1.3.1. Асимптотические свойства функции правдоподобия	19
1.3.2. Сходимость к предельному множеству	21
1.3.3. Идентифицируемость	23
1.3.4. Строгая состоятельность оценок	25
1.3.5. Случай информативного цензурирования	27
1.4. Асимптотическое распределение оценок	28
1.4.1. Условия регулярности типа Крамера	29
1.4.2. Слабые условия асимптотической нормальности	31
Глава 2. Оценки по минимуму расстояния Кульбака-Лейблера	45
2.1. Построение оценок	46
2.2. Оценивание линейных функционалов относительно мер	47
2.2.1. Используемые обозначения и теоремы	48
2.2.2. Вычисление информационных границ в случае интервального цензурирования	53
2.2.3. ОМП функционалов в случае цензурирования	66
2.3. Состоятельность оценок	67
2.4. Асимптотическое распределение оценок	71
2.4.1. Условия типа Крамера	71
2.4.2. Слабые условия асимптотической нормальности	73

2.5.	Вычисление непараметрической оценки функции распределения.	78
2.5.1.	Редукция.	78
2.5.2.	Оптимизация. EM-алгоритм.	81
Глава 3.	Информационные критерии типа Акайке	84
3.1.	Информационные критерии в случае ОМКЛ	86
3.1.1.	Случай интервального цензурирования первого типа	89
3.2.	Информационные критерии в случае ОМП	90
3.2.1.	Случай интервального цензурирования первого типа	93
Глава 4.	Моделирование	94
4.1.	Модели данных типа времени жизни	95
4.1.1.	Распределения, связанные с экспоненциальным	96
4.1.2.	Модель Гомперца-Макегама	97
4.1.3.	Модель ExrCos А.Г. Барта	97
4.2.	Асимптотические свойства ОМП	98
4.2.1.	Распределение Вейбулла	99
4.2.2.	Обобщенное гамма-распределение	99
4.2.3.	Модель Гомперца-Макегама	99
4.2.4.	Модель ExrCos	103
4.3.	Асимптотические свойства ОМКЛ	104
4.3.1.	Сравнение оценок $\tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)}, \tilde{\theta}_n^{(3)}, \tilde{\theta}_n^{(4)}, \tilde{\theta}_n^{(jk)}$	105
4.4.	Сравнение ОМП $\hat{\theta}_n$ и ОМКЛ $\tilde{\theta}_n$	107
Глава 5.	Анализ реальных данных	109
5.1.	Пример из стоматологии	109
5.2.	Пример из кардиологии	111
5.3.	Пример из фармакологии	112

Заключение	116
Литература	118
Приложение А. Разрешимость информационного уравнения в случае интервального цензурирования смешанного вида . .	128
Приложение Б. Доказательство теоремы 2.3	133
Б.1. Используемые результаты и обозначения	133
Б.2. Ход доказательства	136
Б.3. Шаги доказательства	137

Введение

Актуальность темы. Анализ данных типа времени жизни является одной из активно развивающихся областей современной прикладной статистики. Данные такого вида возникают не только в медико-биологических задачах, но и во многих других областях, где требуется изучение «времени до наступления некоторого события», например, при анализе демографических, экономических, финансовых, эпидемиологических показателей, в теории надежности и социологии.

Несмотря на бурное развитие в последние годы непараметрических и семипараметрических методов, специальные (параметрические) модели кривых дожития по-прежнему являются основным инструментом для исследования данных типа времени жизни. Этот факт обусловлен, в первую очередь, высокой информативностью параметрических моделей и возможностью их интерпретации экспериментаторами.

В силу ряда объективных причин для анализа данных типа времени жизни требуются специальные статистические методы. Одной из таких причин, отделяющих эту область от других областей прикладной статистики, является наличие так называемого *цензурирования*: в процессе сбора данных вместо интересующей случайной величины наблюдается другая, менее информативная. Таким образом, при анализе данных типа времени жизни имеет место проблема неполной информации о выборке. Механизмы цензурирования могут быть достаточно сложными и, вследствие этого, требуют отдельного подхода. Стандартные методы анализа данных, как правило, просто не могут быть адекватно применены к случаю цензурирования.

Задача оценивания параметров и выбора специальной (параметрической) модели кривых дожития рассматривалась в работах многих авторов [3, 15, 33, 81]. Как правило, большинство полученных результатов предполагают

наличие так называемого случайного правого цензурирования. Однако, такая модель не очень часто встречается при анализе реальных данных [14], и может рассматриваться только как достаточно простая аппроксимация; вопрос адекватности полученных в таком предположении результатов остается открытым. Вместо этой модели в приложениях более подходящей представляется модель интервального цензурирования [61], специальным случаем которой является вышеупомянутое случайное правое цензурирование.

В связи со сложностью этой модели и ее специальной структурой необходимо развитие специальных методов для оценивания параметров и выбора адекватных параметрических моделей (теория для случайного правого цензурирования не применима здесь по крайней мере без серьезной доработки).

Настоящая работа призвана в известной степени заполнить обозначенный пробел и посвящена построению оценок параметров специальных моделей кривых дожития, исследованию асимптотических свойств полученных оценок, а также развитию методологии сравнения различных параметрических моделей в условиях интервального цензурирования.

Целью работы являются:

1. построение оценок параметров для специальных моделей кривых дожития в условиях интервального цензурирования и изучение асимптотических свойств полученных оценок;
2. разработка методологии выбора адекватной параметрической модели посредством модификации информационных критериев типа Акайке на случай интервального цензурирования;
3. разработка численных методов и систем программ, позволяющих производить оценивание параметров в указанных моделях.

Общая методика работы. В работе применяются методы статистического моделирования, теории вероятностей и математической статистики (оценки максимального правдоподобия, законы больших чисел и центральные предельные теоремы, теория эмпирических процессов), функционального анализа (теория Фредгольмовых операторов), линейной алгебры. Программирование осуществлялось в статистическом пакете R.

Научная новизна. В данной работе впервые получены достаточные условия строгой состоятельности оценок типа максимального правдоподобия в условиях интервального цензурирования и исследованы их асимптотические свойства. Помимо этого, были предложены робастные оценки параметров в случае, когда предполагаемая параметрическая модель не точна. Показано, что эти оценки являются оптимальными с точки зрения расстояния Кульбака-Лейблера между предполагаемой параметрической моделью и истинным распределением данных. При помощи построенных оценок информационные критерии типа Акайке выбора адекватной параметрической модели впервые были распространены на случай интервального цензурирования.

Теоретическая и практическая ценность. В работе математически обоснована применимость (строгая состоятельность, асимптотическое распределение) двух классов оценок к анализу данных в случае интервального цензурирования. Созданы программы, в которых эффективно реализованы разработанные методы оценивания параметров. Методология информационных критериев может быть успешно использована экспериментаторами для подбора адекватной параметрической модели данных.

Апробация работы. Основные результаты диссертации докладывались и обсуждались на семинаре кафедры статистического моделирования математико-механического факультета СПбГУ, а также на конференциях:

- II Всероссийская научно-практическая конференция с международным участием «Высокотехнологичные методы диагностики и лечения заболеваний сердца, крови и эндокринных органов», Федеральный центр сердца им. В.А. Алмазова, г. Санкт-Петербург, 20 – 22 Мая 2008 г.
- 18th Population Approach Group in the Europe (PAGE) Meeting, Saint Petersburg, 23 – 26 June, 2009.
- 6th Saint Petersburg Workshop on Simulation, Saint Petersburg, June 28 – July 4, 2009.

Публикации. По теме диссертации опубликованы работы [37, 78, 82, 83]. Статья [83] опубликована в журнале, входящем в перечень ВАК по специальности 05.13.18. Статья [78] написана в соавторстве, в ней автору принадлежит доказательство теоремы об асимптотических свойствах оценок параметров специальной модели кривой дожития.

Структура и объем диссертации. Диссертация состоит из введения, 5 глав, заключения, списка литературы и 2 глав приложения. Библиография содержит 87 наименований. Общий объем работы 144 страницы.

Первая глава посвящена построению оценок типа максимального правдоподобия в условиях интервального цензурирования смешанного типа и исследованию их асимптотических свойств, таких как состоятельность и асимптотическая нормальность.

Вопросу состоятельности оценок максимального правдоподобия в случае полных данных уделяется достаточно большое внимание в статистической литературе. Классический подход к этому вопросу был впервые рассмотрен в работе [69] и в дальнейшем обобщен в ряде других работ, среди которых стоит отметить работы [25, 30, 42, 48, 49, 79].

Случай неполных данных (цензурирования) изучен не так хорошо. Так, формальная процедура оценивания при помощи функции правдоподобия ти-

па (1.6) рассматривается во многих монографиях, посвященных анализу данных типа времени жизни (например, [61]). Во всех этих работах стандартные условия типа Крамера используются для того, чтобы показать существование состоятельной последовательности корней уравнения правдоподобия. Глава VI в монографии [3] является типичным примером такого подхода, примененного к данным в условиях случайного правого цензурирования.

Этот подход обладает двумя серьезными недостатками. Во-первых, такие условия представляются существенно более сильными, чем требуется: как правило, они включают в себя требование существования равномерно ограниченной третьей частной производной по параметру логарифма функции правдоподобия. В зависимости от выбранной модели для данных проверка этого условия может оказаться достаточно нетривиальной; или, сверх того, выбранная модель данных может и не обладать подобным свойством.

Вторая проблема заключается в том, что показывается только существование состоятельной последовательности корней уравнения правдоподобия, единственность необходимо проверять отдельно. Без условия единственности, как правило, невозможно выбрать правильную (состоятельную) последовательность корней уравнения правдоподобия, обладая информацией только о выборке и модели данных.

Не- или семипараметрический подход к оцениванию в случае интервального цензурирования рассматривался, например, в работах [27–29]. Однако, параметрические модели обычно более предпочтительны в приложениях, так как параметрам модели часто можно придать некоторый смысл, например «среднее время дожития» и т.п. Так же, в сравнении с непараметрическими, параметрические модели обычно демонстрируют лучшую скорость сходимости оценок параметров и меньшую дисперсию.

Альтернативные условия типа Вальда (Wald) [69] состоятельности оценок максимального правдоподобия лишены отмеченных выше недостатков.

Однако, они рассматривались только в работе [59] применительно к данным со случайным правым цензурированием. В диссертации впервые представлены достаточные условия состоятельности типа Вальда в случае интервального цензурирования смешанного типа.

Далее в главе выводятся достаточные условия асимптотической нормальности оценок типа максимального правдоподобия в случае интервального цензурирования. Основным инструментом здесь служит теория эмпирических процессов [67]. Выбор именно такого способа доказательства продиктован универсальностью, эти результаты с небольшими изменениями будут использованы далее.

Во **второй главе** рассматривается класс оценок, минимизирующий расстояние Кульбака-Лейблера (КЛ-оптимальный) между модельным и истинным распределением. Заметим, что в случае оценивания по полным данным (то есть при отсутствии цензурирования) такие оценки совпадают с оценками максимального правдоподобия [1, 8, 70]. При наличии цензурирования это, вообще говоря, не так. Свойство минимизации расстояния между модельным и истинным распределением становится особенно важным в приложениях, так как не всегда присутствует полная уверенность, что предполагаемая модель точна.

КЛ-оптимальные оценки вида (2.3) впервые были предложены в работе [47] для случая правого цензурирования. В этой же работе были представлены эвристические аргументы для обоснования асимптотических свойств оценок. Исчерпывающее исследование свойств было проведено в статье [24] для случая правого цензурирования в том числе и в отсутствие предположения о правильности параметрической модели. Для установления асимптотических свойств оценок параметров в этой работе были использованы условия типа Крамера.

В главе выводятся достаточные условия состоятельности КЛ-оптимальных оценок в случае интервального цензурирование смешанного типа. При этом обобщаются результаты работ [18–20] о сходимости и эффективности гладких функционалов относительно мер со случая интервального цензурирования второго типа на случай интервального цензурирования смешанного типа.

Далее результаты первой главы используются для установления достаточных условий асимптотической нормальности КЛ-оптимальных оценок.

Вопросу сравнения и выбора «наилучшей» параметрической модели посвящена **третья глава**. Одним из стандартных способов сравнения моделей является информационный критерий Акайке [1] и его модификации [35, 36]. Фактически он представляет из себя несмещенную оценку расстояния Кульбака-Лейблера между истинным распределением и предполагаемым модельным. Основной задачей здесь является оценивание смещения расстояния, так как оно, вообще говоря, зависит от способа построения оценок параметров, размерности пространства параметров и т.п.

В диссертации впервые построены информационные критерии типа Акайке для выбора наилучшей параметрической модели при оценивании по выборке с интервальным цензурированием смешанного типа. В качестве оценок параметров при этом рассматривались как оценки типа максимального правдоподобия, так и КЛ-оптимальные оценки.

В **четвертой главе** результаты, полученные в предыдущих главах, проверяются посредством моделирования. В качестве используемых параметрических моделей были выбраны модели, используемые в литературе для анализа данных типа времени жизни: распределения, связанные с экспоненциальным (Вейбулла, обобщенное гамма [58]), модель Гомперца-Макегама [44], модель ExpCos А.Г. Барта [76].

На модельных выборках проверяются такие свойства оценок параметров как состоятельности и скорость сходимости к предельному распределению. Кроме этого, два предложенных в диссертации класса оценок сравниваются с точки зрения дисперсии и среднеквадратичного отклонения в двух случаях: оценивание по модели (то есть, когда предполагаемая параметрическая модель точна) и в случае оценивания вне модели, когда предполагаемая модель не точна.

Пятая глава посвящена применению результатов, полученных в диссертации, к реальным данным из стоматологии, кардиологии и фармакологии.

Оценки типа максимального правдоподобия

1.1. Модель интервального цензурирования смешанного типа

Пусть X_1, \dots, X_n — последовательность независимых одинаково распределенных случайных величин с функцией распределения F_{θ_0} , где $\theta_0 \in \Theta$ и Θ — некоторое метрическое пространство. Пока мы не накладываем никаких дополнительных ограничений на пространство параметров Θ , так, оно вполне может быть бесконечномерным. Мы предполагаем, что случайные величины X имеют смысл времени до наступления некоторого события, времени *наработки на отказ*, и поэтому $\text{supp } F_{\theta_0} \subset [0, \infty)$.

При анализе данных типа времени жизни наблюдать непосредственно время отказа X , как правило, невозможно по тем или иным объективным причинам. Часто вместо этого можно наблюдать пару случайных величин (L, R) , такую, что (см. рисунок 1.1 для примера):

$$-\infty \leq L < X < R \leq +\infty.$$

Например, врача может интересовать время до наступления рецидива некоторого заболевания. В таком случае все, что возможно наблюдать — это момент, в который произошла *диагностика*. Поскольку момент диагностики, вообще говоря, может быть достаточно сильно отдален от момента проявления рецидива, то этот факт необходимо учитывать при построении статистической модели и получении достоверных выводов. Более того, диагностика может лишь констатировать наличие или отсутствие интересующего явления,

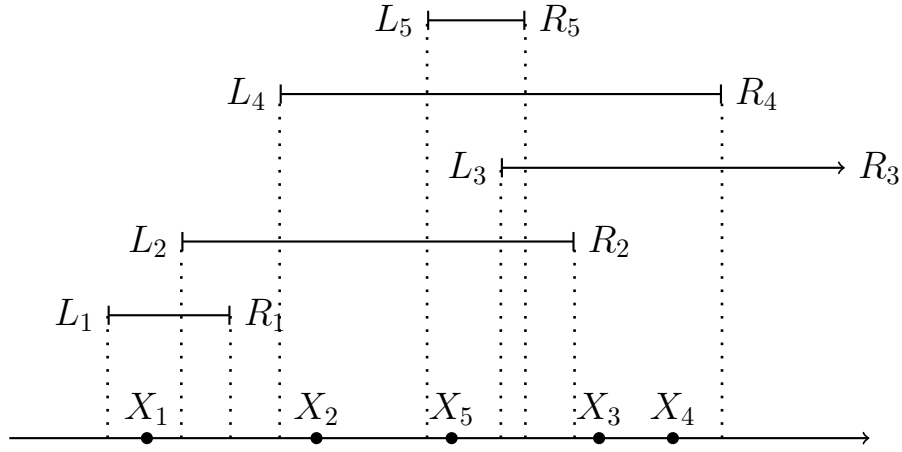


Рис. 1.1. Пример наблюдений с интервальным цензурированием

а наблюдение над индивидом по каким-либо причинам может закончиться до наступления рецидива.

Так, если мы обозначим через X момент рецидива, T — момент последней диагностики, то величина (L, R) может быть задана как

$$(L, R) = \begin{cases} (-\infty, T), & X \leq T, \\ (T, +\infty), & X > T. \end{cases}$$

Для независимых X и T эта модель известна как *интервальное цензурирование первого типа* [23] и пара (L, R) обычно представлена в виде *информации о текущем состоянии* $(T, \Delta = 1_{[X \leq T]})$.

Другой моделью интервального цензурирования является т.н. *интервальное цензурирование второго типа* [23]. Предположим наличие двух моментов диагностики состояния U, V , таких, что $U < V$ и пара (U, V) не зависит от X . В таком случае можно констатировать, произошло ли интересующее событие до момента U , между U и V , или после V . Для этой модели имеем

$$(L, R) = \begin{cases} (-\infty, U), & X \leq U, \\ (U, V), & U < X \leq V, \\ (V, +\infty), & X > V. \end{cases}$$

Естественным обобщением этих моделей является *интервальное цензурирование k -ого типа* [26]. В этом случае предполагается наличие k отдельных моментов диагностики $T_1 < \dots < T_k$, не зависящих от X , и доступна информация, какому из интервалов $(-\infty, T_1], \dots, (T_k, +\infty)$ принадлежит реализация X . Естественно, модель интервального цензурирования k -ого типа при $k > 2$ может быть формально сведена к модели цензурирования второго типа с U и V в виде функций X и T_1, \dots, T_k . Однако, в этом случае нарушится ключевое условие независимости пары (U, V) от X и многие общие результаты, полученные для модели цензурирования второго типа, не будут иметь места.

На практике, однако, нечасто можно наблюдать возможность применения модели интервального цензурирования k -ого вида при $k \geq 2$ в чистом виде: маловероятно, что число наблюдений состояния индивида будет всегда одинаковым. Отдельные индивиды могут выбывать из процесса наблюдения до окончания эксперимента, по тем или иным причинам пропускать диагностику и т.д.

В связи с этим представляется правдоподобной модель *интервального цензурирования смешанного типа* [55], когда количество моментов наблюдений состояния может меняться от индивида к индивиду и, при фиксированном индивиде (а, значит, при фиксированном числе наблюдений), представляет собой модель цензурирования k -ого типа. Тем самым, фактически наблюдается случайная величина, распределенная как смесь с компонентами представленными в виде моделей цензурирования k -ого порядка.

Дадим формальное описание модели интервального цензурирования смешанного типа. Пусть K — неотрицательная целочисленная случайная величина. Обозначим через T треугольный набор случайных величин вида $\{T_{k,j}, j = 1, \dots, k; k = 1, \dots, +\infty\}$, таких, что $0 \equiv T_{k,0} < T_{k,1} < T_{k,2} < \dots < T_{k,k} < T_{k,k+1} \equiv +\infty$. Всюду далее будем предполагать, что случайные величи-

ны X и (K, T) независимы. Рассмотрим случайный вектор $Y = (\Delta_K, T_K, K)$, где T_k — k -я строка треугольного набора T , $\Delta_k = (\Delta_{k,1}, \dots, \Delta_{k,k+1})$ с индикатором $\Delta_{k,j} = 1_{(T_{k,j-1}, T_{k,j}]}(X)$. Другими словами, Y описывает разбиение числовой оси $[0, +\infty)$ на $K + 1$ случайных интервалов и указывает интервал, содержащий X .

Итак, на событии $\{K = k\}$ имеем:

$$(L, R) = \begin{cases} (-\infty, T_{k,1}), & X \leq T_{k,1}, \\ (T_{k,1}, T_{k,2}), & T_{k,1} < X \leq T_{k,2}, \\ \dots & \\ (T_{k,k}, +\infty), & T_{k,k} < X, \end{cases}$$

Величина X имеет смысл случайного времени до наступления некоторого события (например, времени элиминации объекта, времени *наработки на отказ*). Случайная величина K — это количество наблюдений за состоянием индивида, а случайный набор T_K — набор временных точек, в которых производится наблюдение за состоянием индивида.

Типичным примером проявления модели интервального цензурирования при сборе информации во время клинических испытаний является ситуация, когда контроль состояния производится в начале эксперимента, а потом происходит периодический контроль вплоть до момента прекращения эксперимента. Так, если Z_i обозначает отрезок времени между отдельными последовательными моментами контроля, а L — общую продолжительность эксперимента, то имеем

$$T_{k,j} = \sum_{i=1}^j Z_i, \quad K = \sup_{j \geq 1} \left\{ \sum_{i=1}^j Z_i < L \right\}. \quad (1.1)$$

1.2. Построение оценок

Обозначим через $\mathcal{L}(\xi)$ распределение случайной величины ξ . Нетрудно видеть, что условное распределение вектора Δ_K при условии $K = k$ и $T_K = t_k = (t_{k,1}, \dots, t_{k,k})$ является мультиномиальным:

$$\mathcal{L}(\Delta_K | K = k, T_K = t_k) \sim \text{Mult}(1, h_1, \dots, h_{k+1}). \quad (1.2)$$

Здесь $h_j = F_\theta(t_{k,j}) - F_\theta(t_{k,j-1})$ и всюду далее положим $t_{k,0} = 0$ и $t_{k,k+1} = +\infty$.

Рассмотрим измеримое пространство $(\mathcal{Y}, \mathcal{D})$ с

$$\mathcal{Y} = \bigcup_{k=1}^{+\infty} \{0, 1\}^{k+1} \times \mathbb{R}_+^k \times \{k\}. \quad (1.3)$$

\mathcal{D} — сигма-алгебра, порожденная всевозможными множествами вида $A_k \times B_k \times C_k$, $k = 1, \dots, +\infty$. Здесь $A_k \in \mathcal{A}_k$, где \mathcal{A}_k — сигма-алгебра, порожденная $\{\delta_k\}$, и δ_k обозначает вектор индикаторов $(\delta_{k,1}, \dots, \delta_{k,k+1})$, содержащий в точности один ненулевой элемент; $B_k \in \mathcal{B}_k$, \mathcal{B}_k — Борелевская сигма-алгебра подмножеств \mathbb{R}_+^k ; $C_k \in \sigma(\emptyset, \{k\})$. Таким образом, случайная величина Y является измеримым отображением в $(\mathcal{Y}, \mathcal{D})$.

Введем меру ν на \mathcal{D} следующим образом:

$$\nu(\{\delta_k\} \times B_k \times \{k\}) = \#(\delta_k) \mathbf{P}(K = k) \mathbf{P}(T_k \in B_k | K = k),$$

где $\#$ обозначает считающую меру на \mathcal{A}_k .

Ввиду независимости X и (K, T_K) распределение Y является абсолютно непрерывным относительно меры ν . Соответствующая плотность определяется как

$$p_\theta(y) = p_\theta(\delta_k, t_k, k) = \prod_{j=1}^{k+1} [F_\theta(t_{k,j}) - F_\theta(t_{k,j-1})]^{\delta_{k,j}}. \quad (1.4)$$

Пусть Y_1, \dots, Y_n — n независимых, одинаково распределенных случайных величин и $\mathcal{L}(Y_i) = \mathcal{L}(Y) = Q$. Обозначим $Y_i = (\Delta_{K^{(i)}}, T_{K^{(i)}}, K^{(i)})$. Всюду далее символом \mathbb{Q}_n будем обозначать эмпирическую меру, порожденную Y_1, \dots, Y_n .

Обозначим через m_θ логарифм плотности (1.4):

$$m_\theta(\delta_k, t_k, k) = \sum_{j=1}^{k+1} \delta_{k,j} \log [F_\theta(t_{k,j}) - F_\theta(t_{k,j-1})]. \quad (1.5)$$

Введем нормализованную логарифмическую функцию правдоподобия для θ :

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{K^{(i)}+1} \Delta_{K,j}^{(i)} \log [F_\theta(T_{K,j}^{(i)}) - F_\theta(T_{K,j-1}^{(i)})] = \int m_\theta d\mathbb{Q}_n. \quad (1.6)$$

Определение 1. *Оценкой максимального правдоподобия назовем такое значение $\hat{\theta}_n$ параметра θ , для которого имеет место равенство*

$$\int m_{\hat{\theta}_n} d\mathbb{Q}_n = \sup_{\theta \in \Theta} \int m_\theta d\mathbb{Q}_n. \quad (1.7)$$

Супремум в правой части формулы (1.7) иногда может не достигаться, однако допускать сколько угодно близкую аппроксимацию при помощи некоторой последовательности $\hat{\theta}_n$ (например, в случае, когда $\hat{\theta}_n$ вычисляется численно). Поэтому класс оценок максимального правдоподобия разумно расширить, включив туда такие последовательности $\hat{\theta}_n$, для которых выполняется

$$\sup_{\theta \in \Theta} \int m_\theta d\mathbb{Q}_n - \int m_{\hat{\theta}_n} d\mathbb{Q}_n \xrightarrow{\mathbf{P}} 0. \quad (1.8)$$

Такие оценки в литературе обычно называются *приближенными оценками максимального правдоподобия* [30], и именно их мы будем рассматривать всюду далее.

1.3. Состоятельность оценок

Нам понадобятся некоторые вспомогательные результаты относительно асимптотических свойств функции правдоподобия (1.6). В частности, мы покажем, что при наложении достаточно естественных условий функция правдоподобия $\int m_\theta d\mathbb{Q}_n$ сходится по вероятности к $\int m_\theta d\mathbb{Q}$ при n , стремящемся к бесконечности.

Хотя и кажется разумным ожидать, что приближенная точка максимума $\hat{\theta}_n$ функции правдоподобия $\int m_\theta dQ_n$ будет сходиться к точке максимума асимптотической функции правдоподобия $\int m_\theta dQ$, для того, чтобы получить состоятельные оценки, необходимо по меньшей мере показать, что функция $\int m_\theta dQ$ имеет единственную точку максимума. Более того, максимум должен достигаться на истинном значении θ_0 оцениваемого параметра.

Дальнейшее изложение будет разбито на два шага: доказательство асимптотической принадлежности оценки $\hat{\theta}_n$ некоторому множеству предельных точек и вывод условий, гарантирующих предельное множество $\{\theta_0\}$ (а, следовательно, и состоятельность).

1.3.1. Асимптотические свойства функции правдоподобия

Зафиксируем $\theta \in \Theta$. Используя закон больших чисел, получаем, что $\int m_\theta dQ_n \xrightarrow{\mathbf{P}} \int m_\theta dQ$, как только $\int m_\theta dQ < \infty$.

Теорема 1.1. Пусть $\mathbf{E}(K) < \infty$. Тогда для всех $\theta: \int m_\theta dQ < \infty$.

Доказательство. Используя (1.2), (1.5) и независимость X от (K, T_K) , можно получить следующее представление для $\int m_\theta dQ$:

$$\int m_\theta dQ = \sum_{k=1}^{\infty} \mathbf{P}(K = k) \mathbf{E}(f_{\theta,k}(T_{k,1}, \dots, T_{k,k}) | K = k),$$

где для $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = +\infty$

$$f_{\theta,k}(t_1, \dots, t_k) = \sum_{j=1}^{k+1} [F_0(t_j) - F_0(t_{j-1})] \log [F_\theta(t_j) - F_\theta(t_{j-1})]. \quad (1.9)$$

Всюду далее будем считать, что $0 \log 0 = 0$ и $\log 0 = -\infty$. Также обозначим через $F_0(x)$ функцию распределения $F_{\theta_0}(x)$.

Пусть $f_{F,k}$ обозначает выражение, аналогичное формуле (1.9), но с заменой функции распределения F_θ на произвольную функцию распределения F .

Тогда, для любого положительного k и для любого набора $t_1 < \dots < t_k$ мы имеем $\sup_{\theta \in \Theta} f_{\theta,k}(t_1, \dots, t_k) \leq \sup_{\mathcal{F}} f_{F,k}(t_1, \dots, t_k)$, где \mathcal{F} обозначает множество всех функций распределения на \mathbb{R}_+ .

Несложно видеть, что последний супремум достигается на функции $F \in \mathcal{F}$ тогда и только тогда, когда $F_0(t_j) = F(t_j)$ для $j = 1, \dots, k$. Таким образом, для любых k и $0 < t_1 < \dots < t_k < +\infty$ мы имеем

$$\sup_{\theta \in \Theta} f_{\theta,k}(t_1, \dots, t_k) = \sum_{j=1}^{k+1} [F_0(t_j) - F_0(t_{j-1})] \log [F_0(t_j) - F_0(t_{j-1})]. \quad (1.10)$$

Так как $|[F_0(t_j) - F_0(t_{j-1})] \log [F_0(t_j) - F_0(t_{j-1})]| \leq \max_{0 \leq p \leq 1} |p \log p| < 1$, то получаем

$$\begin{aligned} \left| \int m_{\theta} dQ \right| &\leq \int \left| \sup_{\theta \in \Theta} m_{\theta} \right| dQ \leq \\ &\leq \sum_{k=1}^{\infty} \mathbf{P}(K = k) \mathbf{E} \left(\left| \sup_{\theta \in \Theta} f_{\theta,k}(T_{k,1}, \dots, T_{k,k}) \right| \middle| K = k \right) \leq \\ &\leq \sum_{k=1}^{\infty} \mathbf{P}(K = k) \sum_{j=1}^{k+1} \max_{0 \leq p \leq 1} |p \log p| \leq \\ &\leq \sum_{k=1}^{\infty} (k+1) \mathbf{P}(K = k) = E(K) + 1 < \infty. \quad (1.11) \end{aligned}$$

■

Следствие 1.1.1. *Несложно видеть, что из формулы (1.11) следует, что для любого множества $B \subset \Theta$, такого, что отображение*

$$(\delta, t, k) \mapsto \sup_{\theta \in B} m_{\theta}(\delta, t, k)$$

оказывается измеримым, имеет место закон больших чисел в следующей форме:

$$\int \sup_{\theta \in B} m_{\theta} dQ_n \xrightarrow{\mathbf{P}} \int \sup_{\theta \in B} m_{\theta} dQ, \quad \text{при } n \rightarrow \infty.$$

1.3.2. Сходимость к предельному множеству

При доказательстве сходимости к предельному множеству мы придерживаемся общего подхода, развитого в работе [48] для данных без цензурирования.

Теорема 1.2. Пусть $\mathbf{E}(K) < \infty$. Предположим, что $F_\theta(x)$ непрерывна по θ для почти всех x . Предположим, что для любого достаточно малого открытого шара $B \subset \Theta$ отображение $(\delta, t, k) \mapsto \sup_{\theta \in B} m_\theta(\delta, t, k)$ измеримо.

Обозначим

$$\Theta_0 = \left\{ \theta^* \in \Theta : \int m_{\theta^*} dQ = \sup_{\theta} \int m_{\theta} dQ \right\}$$

и предположим, что множество Θ_0 не пусто. Тогда для любого $\varepsilon > 0$ и всякого компакта $S \subset \Theta$

$$\mathbf{P} \left(\text{dist} \left(\hat{\theta}_n, \Theta_0 \right) \geq \varepsilon, \hat{\theta}_n \in S \right) \rightarrow 0 \quad \text{при } n \rightarrow \infty.$$

Доказательство. Зафиксируем $\theta' \in \Theta$. Пусть $\{B_r(\theta')\}$ — убывающая последовательность открытых шаров, содержащих θ' , то есть,

$$\bigcap_{r=1}^{\infty} B_r(\theta') = \{\theta'\}.$$

Из непрерывности функции F следует, что

$$\sup_{B_r(\theta')} m_\theta \downarrow m_{\theta'}, \quad \text{при } r \rightarrow \infty.$$

Тогда теорема 1.1 и теорема о монотонной сходимости гарантируют, что

$$\int \sup_{B_r(\theta')} m_\theta dQ \rightarrow \int m_{\theta'} dQ, \quad \text{при } r \rightarrow \infty. \quad (1.12)$$

Для $\theta^* \in \Theta_0$ и $\theta' \notin \Theta_0$ имеем $\int m_{\theta'} dQ < \int m_{\theta^*} dQ$. Из формулы (1.12) получаем, что существует такая окрестность $U(\theta')$ точки θ' , для которой выполняется

$$\int \sup_{U(\theta')} m_\theta dQ < \int m_{\theta^*} dQ.$$

Отметим, что значение $\int m_{\theta^*} dQ$ (и поэтому окрестность $U(\theta')$) не зависит от конкретного выбора точки $\theta^* \in \Theta_0$.

Зафиксируем $\varepsilon > 0$. Обозначим $\bar{S}^\varepsilon = \{\theta \in S : \text{dist}(\theta, \Theta_0) \geq \varepsilon\}$. Покроем множество \bar{S}^ε окрестностями $U(\theta)$ каждой точки: $\bar{S}^\varepsilon \subset \bigcup_{\theta \in \bar{S}^\varepsilon} U(\theta)$. В силу компактности множества \bar{S}^ε существует подпокрытие $\theta_1, \dots, \theta_k \in \bar{S}^\varepsilon$, такое, что $\bar{S}^\varepsilon \subset \bigcup_{i=1}^k U(\theta_i)$. Отсюда выводим

$$\begin{aligned} \sup_{\theta \in \bar{S}^\varepsilon} \int m_\theta dQ_n &\leq \max_{i=1, \dots, k} \int \sup_{U(\theta_i)} m_\theta dQ_n \rightarrow \\ &\rightarrow \max_{i=1, \dots, k} \int \sup_{U(\theta_i)} m_\theta dQ < \int m_{\theta^*} dQ, \quad \text{п.н.} \end{aligned} \quad (1.13)$$

Теперь для $\hat{\theta}_n \in \bar{S}^\varepsilon$ имеем

$$\sup_{\theta \in \bar{S}^\varepsilon} \int m_\theta dQ_n \geq \int m_{\hat{\theta}_n} dQ_n. \quad (1.14)$$

Из определения $\hat{\theta}_n$ следует, что

$$\sup_{\theta \in \Theta} \int m_\theta dQ_n - \int m_{\hat{\theta}_n} dQ_n \xrightarrow{\mathbf{P}} 0, \quad \text{при } n \rightarrow \infty.$$

Отсюда, применяя усиленный закон больших чисел, получаем

$$\int m_{\theta^*} dQ - \int m_{\hat{\theta}_n} dQ_n \xrightarrow{\mathbf{P}} 0, \quad \text{при } n \rightarrow \infty. \quad (1.15)$$

Зафиксируем $\delta > 0$. Из формулы (1.14) выводим

$$\begin{aligned} \mathbf{P}(\hat{\theta}_n \in \bar{S}^\varepsilon) &\leq \mathbf{P}\left(\sup_{\theta \in \bar{S}^\varepsilon} \int m_\theta dQ_n - \int m_{\hat{\theta}_n} dQ_n \geq 0\right) \leq \\ &\leq \mathbf{P}\left(\sup_{\theta \in \bar{S}^\varepsilon} \int m_\theta dQ_n - \int m_{\hat{\theta}_n} dQ_n \geq 0, \left|\int m_{\hat{\theta}_n} dQ_n - \int m_{\theta^*} dQ\right| \leq \delta\right) + \\ &\quad + \mathbf{P}\left(\left|\int m_{\hat{\theta}_n} dQ_n - \int m_{\theta^*} dQ\right| > \delta\right) \leq \\ &\leq \mathbf{P}\left(\sup_{\theta \in \bar{S}^\varepsilon} \int m_\theta dQ_n - \int m_{\theta^*} dQ \geq -\delta\right) + \mathbf{P}\left(\left|\int m_{\hat{\theta}_n} dQ_n - \int m_{\theta^*} dQ\right| > \delta\right). \end{aligned}$$

Формула (1.13) гарантирует существование достаточно малого δ , для которого первая вероятность сходится к нулю. В таком случае вторая вероятность сходится к нулю в силу формулы (1.15). ■

Замечание. Условие непрерывности функции $F_\theta(x)$ по параметру θ для почти всех x может быть ослаблено: достаточно предполагать, что функция m_θ полунепрерывна сверху по θ для почти всех (δ, t, k) .

Замечание. Измеримость отображения $(\delta, t, k) \mapsto \sup_{\theta \in B} m_\theta(\delta, t, k)$ для всякого достаточно малого открытого шара $B \subset \Theta$ часто автоматически присутствует в приложениях. Например, это верно, при выполнении любого из следующих условий:

1. Функция $F_\theta(x)$ непрерывна по θ для п.в. x и пространство параметров Θ сепарабельно.
2. Семейство функций $\{m_\theta(\delta, t, k)\}$ равностепенно непрерывно для ν -п.в. (δ, t, k) и Θ сепарабельно.
3. Пространство параметров Θ является подмножеством числовой прямой \mathbb{R} , а функция m_θ непрерывна справа по θ .

1.3.3. Идентифицируемость

Теорема 1.2 предоставляет достаточные условия для того, чтобы оценка $\hat{\theta}_n$ была сколь угодно близка (по вероятности) к множеству

$$\Theta_0 = \left\{ \theta^* \in \Theta : \int m_{\theta^*} dQ = \sup_{\theta} \int m_{\theta} dQ \right\}.$$

Для состоятельности оценки необходима *идентифицируемость*: достаточные условия, гарантирующие, что множество $\Theta_0 = \{\theta_0\}$.

Для начала заметим, что в условиях теоремы 1.1, $\theta_0 \in \Theta_0$. Действительно,

$$\int m_{\theta_0} dQ \leq \sup_{\theta \in \Theta} \int m_{\theta} dQ \leq \int \sup_{\theta \in \Theta} m_{\theta} dQ. \quad (1.16)$$

Из (1.10) следует, в частности, что

$$\int \sup_{\theta \in \Theta} m_{\theta} dQ = \int m_{\theta_0} dQ,$$

отсюда мы получаем, что все неравенства в формуле (1.16) являются равенствами. Тогда из первого равенства

$$\int m_{\theta_0} dQ = \sup_{\theta \in \Theta} \int m_{\theta} dQ,$$

следует, что $\theta_0 \in \Theta_0$.

Обозначим через μ меру Шика (Schick) [55], определенную на Борелевской сигма-алгебре \mathcal{B} подмножеств \mathbb{R} :

$$\mu(B) = \sum_{k=1}^{+\infty} \mathbf{P}(K = k) \sum_{j=1}^k \mathbf{P}(T_{k,j} \in B \mid K = k), \quad B \in \mathcal{B}. \quad (1.17)$$

Мера μ может быть использована для характеристики всех точек максимума логарифмической функции правдоподобия (1.6): из формул (1.9) и (1.10) можно получить, что любое значение θ , максимизирующее функцию (1.6), удовлетворяет соотношению $F_{\theta} = F_0 \mu$ — п.н.. Как следствие этого, получаем следующее представление для множества предельных точек Θ_0 :

$$\Theta_0 = \{\theta^* \in \Theta : F_{\theta^*} = F_0 \mu \text{ — п.н.}\}.$$

Таким образом, выводим условие идентифицируемости, которое вместе с теоремой 1.2 гарантирует состоятельность оценки $\hat{\theta}_n$:

$$F_{\theta_1} = F_{\theta_2} \mu \text{ — п.н.} \Rightarrow \theta_1 = \theta_2. \quad (1.18)$$

Действительно, это условие гарантирует отсутствие значения $\theta' \neq \theta_0$, такого, что $F_{\theta'} = F_0 \mu$ — п.н., и, как следствие этого, $\Theta_0 = \{\theta_0\}$.

Стоит заметить, что условие идентифицируемости связывает модельную функцию распределения F с распределением величины (K, T_K) (задающей механизм цензурирования) посредством меры μ . Вообще говоря, проверка этого условия может быть весьма нетривиальной, так как обычно распределение величины (K, T_K) не известно. Однако, часто в приложениях условие идентифицируемости можно свести к соотношениям на носители F и (K, T_K) , проверка которых может быть проведена существенно проще (или же эти соотношения возможно обеспечить заранее за счет выбора процедуры проведения эксперимента и сбора данных).

В качестве примера, рассмотрим условие идентифицируемости в случае модели интервального цензурирования первого типа.

Напомним, что данные о состоянии содержат момент контроля состояния объекта T и метку состояния $\delta = 1_{[X \leq T]}$. X и T предполагаются независимыми. Запишем условие идентифицируемости параметра θ :

$$\mu \{x : F_{\theta_1}(x) \neq F_{\theta_2}(x)\} > 0, \quad \theta_1 \neq \theta_2,$$

откуда тривиально получаем

$$\mathbf{P}(T \in \{x : F_{\theta_1}(x) \neq F_{\theta_2}(x)\}) > 0, \quad \theta_1 \neq \theta_2,$$

включающее достаточно естественные условия на носители распределений T и X .

1.3.4. Строгая состоятельность оценок

В случае, когда выполняется условие идентифицируемости параметра, теорема 1.2 допускает модификацию, позволяющую получить строгую состоятельность (то есть сходимость п.н.) оценок. Для полноты изложения мы приведем доказательство этого результата целиком.

Теорема 1.3. Пусть выполняются условия теоремы 1.2 и условие идентифицируемости (1.18). Предположим, что пространство параметров Θ компактно. Тогда $\hat{\theta}_n \rightarrow \theta_0$, п.н. $n \rightarrow \infty$.

Доказательство. Как и в теореме 1.2 зафиксируем $\theta' \in \Theta$ и обозначим через $\{B_r(\theta')\}$ убывающую последовательность открытых шаров, содержащих θ' :

$$\bigcap_{r=1}^{\infty} B_r(\theta') = \{\theta'\}.$$

Из непрерывности функции F следует, что

$$\sup_{B_r(\theta')} m_\theta \downarrow m_{\theta'}, \quad \text{при } r \rightarrow \infty.$$

Тогда теорема 1.1 и теорема о монотонной сходимости гарантируют, что

$$\int \sup_{B_r(\theta')} m_\theta dQ \rightarrow \int m_{\theta'} dQ, \quad \text{при } r \rightarrow \infty. \quad (1.19)$$

Рассмотрим окрестность $B(\theta_0)$ точки θ_0 . Определим множество $\bar{B}(\theta_0) = \Theta \setminus B(\theta_0)$. Для $\theta' \in \bar{B}(\theta_0)$ условие идентифицируемости (1.18) гарантирует, что

$$\int m_{\theta'} dQ < \int m_{\theta_0} dQ.$$

Таким образом, объединяя с (1.19), имеем

$$\int \sup_{B_r(\theta')} m_\theta dQ \rightarrow \int m_{\theta'} dQ < \int m_{\theta_0} dQ, \quad \text{при } r \rightarrow \infty.$$

Следовательно, для всякого $\theta' \in \bar{B}(\theta_0)$ существует окрестность $U(\theta')$, что

$$\int \sup_{U(\theta')} m_\theta dQ < \int m_{\theta_0} dQ.$$

Отсюда имеем

$$\begin{aligned} \sup_{U(\theta')} \int m_\theta dQ_n &\leq \int \sup_{U(\theta')} m_\theta dQ_n \rightarrow \\ &\rightarrow \int \sup_{U(\theta')} m_\theta dQ < \int m_{\theta_0} dQ, \quad \text{п.н. при } n \rightarrow \infty. \end{aligned} \quad (1.20)$$

Компактность множества $\bar{B}(\theta_0)$ позволяет переписать формулу (1.20) в виде

$$\limsup_{n \rightarrow \infty} \sup_{\bar{B}(\theta_0)} \int m_\theta dQ_n < \int m_{\theta_0} dQ, \quad \text{п.н.}$$

Отсюда

$$\limsup_{n \rightarrow \infty} \left(\sup_{\bar{B}(\theta_0)} \int m_\theta dQ_n - \int m_{\theta_0} dQ \right) < 0, \quad \text{п.н.}$$

а так как из теоремы 1.1 следует, что $\int m_{\theta_0} dQ_n - \int m_{\theta_0} dQ \rightarrow 0$, то

$$\limsup_{n \rightarrow \infty} \left(\sup_{\bar{B}(\theta_0)} \int m_\theta dQ_n - \int m_{\theta_0} dQ_n \right) < 0, \quad \text{п.н.}$$

Определение (1.8) тогда гарантирует, что

$$\mathbf{P} \left(\hat{\theta}_n \in B(\theta_0) \text{ для почти всех } n \right) = 1,$$

откуда, в силу произвольности окрестности $B(\theta_0)$, следует, что

$$\hat{\theta}_n \rightarrow \theta_0, \quad \text{п.н. при } n \rightarrow \infty.$$

■

1.3.5. Случай информативного цензурирования

Предположение о независимости случайных величин X и (K, T_K) может оказаться достаточно сложным для проверки при наличии только данных без какой-либо информации о процедуре сбора данных. Если это предположение не выполняется, что механизм цензурирования называется *информативным* (см. обзор в монографии [61]). Кратко рассмотрим отличия информативного от неинформативного цензурирования в контексте результатов, полученных в предыдущих разделах.

Во-первых, функция $p_\theta(y)$ из формулы (1.4) уже не будет плотностью величины Y . Впрочем, мы по-прежнему можем использовать оценки, опре-

деленные формулой (1.7). В таком случае, функция m_θ будет «функцией-критерием», и оценки $\hat{\theta}_n$ можно рассматривать как приближенные М-оценки параметра θ_0 .

Формула (1.9) перестанет быть верной, но останется возможным получить неравенство вида

$$f_{\theta,k}(t_1, \dots, t_k) \leq \sum_{j=1}^{k+1} a_j^{(k)} \log [F_\theta(t_j) - F_\theta(t_{j-1})],$$

для некоторых $a_j^{(k)} \geq 0$, $\sum_{j=1}^{k+1} a_j^{(k)} = 1$. Как следствие этого, все результаты теоремы 1.1 останутся верными.

Теорема 1.2 напрямую не зависит от механизма цензурирования и поэтому останется верной и в случае информативного цензурирования. Основное различие между информативным и неинформативным цензурированием заключается в структуре множества Θ_0 , определенном в теореме 1.2. Теперь возможно, что $\theta_0 \notin \Theta_0$, и условие идентифицируемости (1.18) потребует изменения для учета зависимости X и (K, T_K) .

1.4. Асимптотическое распределение оценок

Имея состоятельную оценку $\hat{\theta}_n$, становится возможным применить дельта-метод и стандартные условия регулярности М-оценок [9, 42, 87] для установления достаточных условий асимптотической нормальности оценок параметров.

Так, в разделе 1.4.1 приводятся достаточные условия регулярности типа Крамера [87]. Однако, как видно из условия теоремы 1.4, такие условия труднопроверяемы на практике: фактически требуется установить равномерную ограниченность третьих частных производных функции $\theta \mapsto m_\theta$.

Эти условия можно существенно ослабить, применив методы теории эмпирических процессов [9, 67]. Финальным результатом тут является теоре-

ма 1.8. Фактически оказывается достаточным наличие лишь ограниченного градиента функции $\theta \mapsto \log(F_\theta(y) - F_\theta(x))$, $x < y$ в окрестности истинного значения неизвестного параметра, что вполне естественно.

Помимо этого, результаты, использующие теорию эмпирических процессов в таком виде, нам потребуются в главе 2 для установления асимптотических свойств оценок по минимуму расстояния Кульбака-Лейблера.

1.4.1. Условия регулярности типа Крамера

Для упрощения обозначений предположим, что $\theta \in \Theta \subset \mathbb{R}$. Предположим, что функция $\theta \mapsto t_\theta$ дифференцируема всюду и обозначим $\psi_\theta = \frac{\partial}{\partial \theta} t_\theta$.

Теорема 1.4. *Потребуем выполнения следующих условий:*

1. Пусть отображение $\theta \mapsto \psi_\theta(\delta, t, k)$ дважды непрерывно дифференцируемо для почти всех (δ, t, k) .
2. $\int \psi_{\theta_0}^2 dQ < \infty$ и $\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ \neq 0$.
3. Предположим, что найдется такая суммируемая (по распределению Q) функция $M(\delta, t, k)$, что в некоторой окрестности точки θ_0 :

$$\left| \frac{\partial^2}{\partial \theta^2} \psi_\theta(\delta, t, k) \right| \leq M(\delta, t, k).$$

4. $\int \psi_{\hat{\theta}_n} dQ_n = o_p(n^{-1/2})$.

Тогда для состоятельной приближенной ОМП $\hat{\theta}_n$ выполняется:

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N \left(0, \frac{\int \psi_{\theta_0}^2 dQ}{\left(\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ \right)^2} \right), \text{ при } n \rightarrow \infty.$$

Доказательство. Разложим функцию $\int \psi_{\hat{\theta}_n} d\mathbb{Q}_n$ в окрестности точки θ_0 по формуле Тейлора (перестановка местами операций интегрирования и дифференцирования возможна, так как \mathbb{Q}_n — эмпирическая мера):

$$\int \psi_{\hat{\theta}_n} d\mathbb{Q}_n = \int \psi_{\theta_0} d\mathbb{Q}_n + (\hat{\theta}_n - \theta_0) \int \frac{\partial}{\partial \theta} \psi_{\theta_0} d\mathbb{Q}_n + \frac{1}{2} (\hat{\theta}_n - \theta_0)^2 \int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\mathbb{Q}_n, \quad (1.21)$$

здесь величина $\tilde{\theta}_n$ расположена на сегменте прямой, соединяющей θ_0 и $\hat{\theta}_n$.

Из условия 2 и закона больших чисел следует, что

$$\int \frac{\partial}{\partial \theta} \psi_{\theta_0} d\mathbb{Q}_n \xrightarrow{\mathbf{P}} \int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ, \text{ при } n \rightarrow \infty.$$

Состоятельность оценки $\hat{\theta}_n$, условие 3 и закон больших чисел, в свою очередь, гарантируют ограниченность величины $\int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\mathbb{Q}_n$ по вероятности. С учетом всего этого из формулы (1.21) получаем:

$$\begin{aligned} \int \psi_{\hat{\theta}_n} d\mathbb{Q}_n - \int \psi_{\theta_0} d\mathbb{Q}_n &= \\ &= (\hat{\theta}_n - \theta_0) \int \frac{\partial}{\partial \theta} \psi_{\theta_0} d\mathbb{Q}_n + \frac{1}{2} (\hat{\theta}_n - \theta_0)^2 \int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\mathbb{Q}_n = \\ &= (\hat{\theta}_n - \theta_0) \left(\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ + o_P(1) + (\hat{\theta}_n - \theta_0) O_P(1) \right), \text{ при } n \rightarrow \infty. \end{aligned}$$

Состоятельность оценки $\hat{\theta}_n$ влечет за собой $(\hat{\theta}_n - \theta_0) O_P(1) = o_P(1)$. Из условия 2 следует, что при достаточно больших n величина $\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ + o_P(1)$ отлична от нуля. Отсюда, используя условие 4, выводим

$$\begin{aligned} \sqrt{n} (\hat{\theta}_n - \theta_0) &= \sqrt{n} \left(\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ + o_P(1) \right)^{-1} \left(\int \psi_{\hat{\theta}_n} d\mathbb{Q}_n - \int \psi_{\theta_0} d\mathbb{Q}_n \right) = \\ &= \sqrt{n} \left(\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ + o_P(1) \right)^{-1} \left(o_P(n^{-1/2}) - \int \psi_{\theta_0} d\mathbb{Q}_n \right). \quad (1.22) \end{aligned}$$

Так как точка θ_0 является точкой максимума отображения $\theta \mapsto m_\theta(\delta, t, k)$, то, очевидно, $\int \psi_{\theta_0} dQ = 0$. В таком случае из условия 2 следует, что случайная величина $\sqrt{n} \int \psi_{\theta_0} d\mathbb{Q}_n$ сходится по распределению к случайной величине,

имеющей нормальное распределение со средним 0 и дисперсией $\int \psi_{\theta_0}^2 dQ$, и, следовательно,

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N \left(0, \frac{\int \psi_{\theta_0}^2 dQ}{\left(\int \frac{\partial}{\partial \theta} \psi_{\theta_0} dQ \right)^2} \right), \text{ при } n \rightarrow \infty.$$

■

Замечание. Условие 4 фактически накладывает ограничение на то, насколько хорошо должна оценка $\hat{\theta}_n$ аппроксимировать точку максимума функции $\theta \mapsto \int t_\theta dQ_n$. Так, если бы $\hat{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} \int t_\theta dQ_n$, то тогда необходимо выполняется $\int \psi_{\hat{\theta}_n} dQ_n = 0$. Если же величина $\int \psi_{\hat{\theta}_n} dQ_n$ сходится к нулю недостаточно быстро, то, как видно из формулы 1.22, предельное распределение величины $\sqrt{n} (\hat{\theta}_n - \theta_0)$ может отличаться от нормального (при $\int \psi_{\hat{\theta}_n} dQ_n = O_p(n^{-1/2})$), либо оценка $\hat{\theta}_n$ будет сходиться со скоростью, меньшей, чем $n^{-1/2}$ (при $\int \psi_{\hat{\theta}_n} dQ_n = O_p(n^\alpha)$ с $\alpha > -1/2$).

1.4.2. Слабые условия асимптотической нормальности

Используемые обозначения и теоремы

В дальнейшем нам потребуется несколько определений и теорем из [67], в частности, определения скобочного ε -энтропийного числа, скобочной ε -энтропии и соответствующего энтропийного интеграла, теорема об оценивании модуля непрерывности эмпирического процесса и теорема о скорости сходимости M -оценок.

Рассмотрим нормированное пространство L вещественнозначных функций $f : \mathcal{X} \rightarrow \mathbb{R}$. Пусть \mathcal{F} — его подмножество.

Определение 2. Зафиксируем пару функций $l, u \in L$. Скобкой $[l, u]$ назовем множество всех функций $f \in L : l \leq f \leq u$. Соответственно, ε -скобка — это скобка $[l, u]$, для которой $\|u - l\| < \varepsilon$.

Определение 3. Скобочным ε -энтропийным числом $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ подмножества $\mathcal{F} \subset L$ назовем минимальное количество ε -скобок, необходимых для покрытия \mathcal{F} . Скобочная ε -энтропия — логарифм скобочного числа. В определении скобочного числа верхние и нижние границы скобок l, u не обязаны принадлежать \mathcal{F} , но должны иметь конечные нормы.

Определение 4. Энтропийным интегралом подмножества \mathcal{F} назовем величину

$$J_{[]}(\delta, \mathcal{F}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)} d\varepsilon$$

Определение 5. Эмпирическим процессом \mathbb{G}_n назовем меру $\sqrt{n}(\mathbb{Q}_n - Q)$.

На скорость сходимости любых состоятельных M-оценок влияет модуль непрерывности эмпирического процесса \mathbb{G}_n на классе функций-критериев $\mathcal{G} = \{g_\theta, \theta \in \Theta\}$. Это отражено в теореме 3.2.5 из [67]:

Теорема E1 (van der Vaart—Wellner). Пусть (Θ, ρ) — метрическое пространство. Предположим, что найдется такая константа C , что для любого достаточно малого δ выполняется

$$\sup_{\rho(\theta, \theta_0) < \delta} \int (g_\theta - g_{\theta_0}) dQ \leq -C\delta^2. \quad (1.23)$$

Предположим, что существует функция ϕ , такая, что $\phi(x)/x^\alpha$ убывает для некоторого $\alpha < 2$ и для любого n имеет место неравенство:

$$\mathbf{E}^* \sup_{\rho(\theta, \theta_0) < \delta} \int |g_\theta - g_{\theta_0}| d\mathbb{G}_n \leq C\phi(\delta). \quad (1.24)$$

Если для последовательность $\bar{\theta}_n$ имеет место соотношение

$$\int g_{\bar{\theta}_n} d\mathbb{Q}_n - \int g_{\theta_0} d\mathbb{Q}_n = O_P(r_n^{-2}),$$

и $\bar{\theta}_n$ сходится по вероятности к θ_0 , то $r_n \rho(\bar{\theta}_n, \theta_0) = O_p(1)$ для любой последовательности r_n , такой, что $r_n^2 \phi(1/r_n) \leq \sqrt{n}$ для любого n .

Оценить модуль непрерывности (1.24) процесса \mathbb{G}_n , например, можно при помощи максимальных неравенств, связывающих его с энтропийным интегралом. Мы будем использовать теорему 2.14.2 из [67]:

Теорема E2 (van der Vaart—Wellner). *Рассмотрим класс измеримых функций $\mathcal{F} \subset L^2(Q)$. Предположим, что существует измеримая мажоранта F класса \mathcal{F} . Тогда найдется такая константа C , что*

$$\mathbf{E} \sup_{f \in \mathcal{F}} \int |f| d\mathbb{G}_n \leq C J_{[]} (1, \mathcal{F}, L^2(Q)) \left(\int F^2 dQ \right)^{1/2}.$$

При доказательстве асимптотической нормальности оценок на одном из шагов мы воспользуемся теоремой 2.11.23 из [67]:

Теорема E3 (van der Vaart—Wellner). *Для каждого n обозначим через $\mathcal{F}_n = \{f_{n,t} : t \in T\}$ класс измеримых функций, индексированных равномерно ограниченным полуметрическим пространством (T, ρ) . Предположим, что найдутся такие измеримые функции F_n , что $\sup_{t \in T} |f_{n,t}| \leq F_n$. Пусть выполняются следующие условия:*

1. $\int F_n^2 dQ = O(1)$, $n \rightarrow \infty$.
2. $\int_{F_n > \eta\sqrt{n}} F_n^2 dQ \rightarrow 0$ при $n \rightarrow \infty$ для всякого $\eta > 0$.
3. $\sup_{\rho(s,t) < \delta_n} \int (f_{n,s} - f_{n,t})^2 dQ \rightarrow 0$, для всех $\delta_n \rightarrow 0$.

Кроме этого, пусть выполняется

$$\int_0^{\delta_n} \sqrt{\log N_{[]} \left(\varepsilon \|F_n\|_{L^2(Q)}, \mathcal{F}_n, L^2(Q) \right)} d\varepsilon \rightarrow 0, \quad \text{для всех } \delta_n \rightarrow 0. \quad (1.25)$$

Тогда последовательность процессов $\left\{ \int f_{n,t} d\mathbb{G}_n : t \in T \right\}$ является плотной в $L^\infty(T)$.

Скорость сходимости ОМП

Условие (1.23) теоремы E1 достаточно естественно: если θ_0 является точкой максимума функции $\theta \mapsto \int m_\theta dQ$, то градиент функции должен обращаться в ноль, а матрица вторых производных быть отрицательно определенной. Например, оно выполняется, если функция $\theta \mapsto \int m_\theta dQ$ допускает разложение в ряд Тейлора в окрестности точки θ_0 до второго члена:

$$\int (m_\theta - m_{\theta_0}) dQ = \frac{1}{2} (\theta - \theta_0)^T \Sigma_0 (\theta - \theta_0) + o(|\theta - \theta_0|^2), \quad \theta \rightarrow \theta_0.$$

В нашем случае ОМП $\hat{\theta}_n$ для оценивания модуля непрерывности (1.24) на семействе $\{m_\theta, \theta \in \Theta\}$ теорему E2 необходимо применить к классам функций

$$\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : \|\theta - \theta_0\| < \delta\}.$$

Теорема 1.5. Пусть выполняется:

1. Функция $\theta \mapsto \int m_\theta dQ$ допускает разложение в ряд Тейлора до второго члена.
2. Пространство параметров Θ является подмножеством евклидова пространства \mathbb{R}^p .
3. Отображение $\theta \mapsto \log(F_\theta(x) - F_\theta(y))$ дифференцируемо в некоторой окрестности точки θ_0 для $\mu \times \mu$ -п.в. пар (x, y) , $x < y$ с равномерно ограниченным градиентом.
4. Для состоятельной приближенной оценки максимального правдоподобия $\hat{\theta}_n$ выполняется

$$\int m_{\hat{\theta}_n} dQ_n - \int m_{\theta_0} dQ_n = O_p(n^{-1}).$$

Тогда $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$.

Доказательство. Условия 3 теоремы гарантирует наличие мажоранты для класса функций \mathcal{M}_δ : несложно видеть, что найдется функция $M(\delta, t, k)$ с $\int M^2 dQ < \infty$, что

$$\sup_{f \in \mathcal{M}_\delta} |f| \leq M\delta. \quad (1.26)$$

Тогда, из теоремы E2 имеем

$$\mathbf{E} \sup_{f \in \mathcal{M}_\delta} \int |f| d\mathbb{G}_n \leq \delta J_{\square} (1, \mathcal{M}_\delta, L^2(Q)) \|M\|_{L^2(Q)}.$$

Покажем, что энтропийный интеграл

$$J_{\square} (1, \mathcal{M}_\delta, L^2(Q)) = \int_0^1 \sqrt{1 + \log N_{\square}(\varepsilon, \mathcal{M}_\delta, L^2(Q))} d\varepsilon \quad (1.27)$$

конечен. Для этого достаточно вычислить скобочное число $N_{\square}(\varepsilon, \mathcal{M}_\delta, L^2(Q))$.

Из формулы (1.26) и определения класса \mathcal{M}_δ следует, что

$$-M \|\theta_1 - \theta_2\| + (m_{\theta_2} - m_{\theta_0}) \leq (m_{\theta_1} - m_{\theta_0}) \leq M \|\theta_1 - \theta_2\| + (m_{\theta_2} - m_{\theta_0}). \quad (1.28)$$

Покроем множество $\Theta_\delta = \{\theta \in \Theta : \|\theta - \theta_0\| < \delta\}$ шарами радиуса r , где

$$r = \frac{\varepsilon}{2 \|M\|_{L^2(Q)}}$$

и рассмотрим ε -скобки вида $[m_\theta - m_{\theta_0} - rM, m_\theta - m_{\theta_0} + rM]$. Условие (1.28) гарантирует, что количество скобок такого вида, необходимое для покрытия множества \mathcal{M}_δ , совпадает с количеством шаров радиуса r , необходимых для покрытия множества Θ_δ . Θ_δ можно вписать в куб радиуса 2δ , поэтому требуется не более, чем $\left(\frac{2\delta}{r\sqrt{d}}\right)^d = \left(\frac{4\delta \|M\|_{L^2(Q)}}{\varepsilon\sqrt{d}}\right)^d$ кубов со стороной r/\sqrt{d} , чтобы покрыть Θ_δ . Ясно, что каждый такой куб покрывается шаром радиуса r . Таким образом, получаем следующую оценку на $N_{\square}(\varepsilon, \mathcal{M}_\delta, L^2(Q))$:

$$N_{\square}(\varepsilon, \mathcal{M}_\delta, L^2(Q)) \leq \left(\frac{4\delta \|M\|_{L^2(Q)}}{\varepsilon\sqrt{d}}\right)^d = C \left(\frac{\delta}{\varepsilon}\right)^d. \quad (1.29)$$

Подставим полученную оценку в энтропийный интеграл (1.27):

$$\begin{aligned}
J_{[]} (1, \mathcal{M}_\delta, L^2(Q)) &= \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{M}_\delta, L^2(Q))} d\varepsilon \leq \\
&\leq \int_0^1 \sqrt{1 + \log C + d \log \left(\frac{\delta}{\varepsilon} \right)} d\varepsilon \leq C' + \sqrt{d} \int_0^1 \sqrt{\log \left(\frac{\delta}{\varepsilon} \right)} d\varepsilon = \\
&= C' + \sqrt{d} \int_\delta^{+\infty} \frac{\sqrt{\log \varepsilon'}}{\varepsilon'^2} d\varepsilon' < +\infty
\end{aligned}$$

Условие (1.24) выполняется с функцией $\phi(\delta) = \delta$. Условие (1.23) тривиально выполняется вследствие возможности разложения по Тейлору функции $\theta \mapsto \int m_\theta dQ$. Таким образом, применяя теорему E1, получаем требуемую скорость сходимости оценки $\hat{\theta}_n$: $\sqrt{n} (\hat{\theta}_n - \theta_0) = O_p(1)$. ■

Асимптотическая нормальность ОМП

Рассмотрим выражение $n \int (m_{\hat{\theta}_n} - m_{\theta_0}) dQ_n$. Выводим:

$$\begin{aligned}
n \int (m_{\hat{\theta}_n} - m_{\theta_0}) dQ_n &= \\
&= n \int (m_{\hat{\theta}_n} - m_{\theta_0}) dQ + n \int (m_{\hat{\theta}_n} - m_{\theta_0}) d(Q_n - Q) = \\
&= n \int (m_{\hat{\theta}_n} - m_{\theta_0}) dQ + \sqrt{n} \int (m_{\hat{\theta}_n} - m_{\theta_0}) dG_n. \quad (1.30)
\end{aligned}$$

Пусть функция $\theta \mapsto \int m_\theta dQ$ допускает разложение в ряд Тейлора в окрестности точки θ_0 до второго члена с невырожденной матрицей вторых частных производных Σ_0 . Так как точка θ_0 является точкой максимума функции $\theta \mapsto \int m_\theta dQ$, то из формулы (1.30) получаем

$$\begin{aligned}
n \int (m_{\hat{\theta}_n} - m_{\theta_0}) dQ_n &= \frac{1}{2} \sqrt{n} (\hat{\theta}_n - \theta_0)^T \Sigma_0 \sqrt{n} (\hat{\theta}_n - \theta_0) + \\
&+ \sqrt{n} \int (m_{\hat{\theta}_n} - m_{\theta_0}) dG_n + o_P(n \|\hat{\theta}_n - \theta_0\|^2), \quad n \rightarrow \infty. \quad (1.31)
\end{aligned}$$

Предположим, что функция m_θ стохастически дифференцируема в окрестности точки θ_0 по мере \mathbb{G}_n , т.е. найдется такая функция m'_{θ_0} , что для всякой последовательности $\hat{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$, $n \rightarrow \infty$ имеет место

$$\int \left(m_{\hat{\theta}_n} - m_{\theta_0} \right) d\mathbb{G}_n = \left(\hat{\theta}_n - \theta_0 \right)^T \int m'_{\theta_0} d\mathbb{G}_n + o_p(n^{-1/2}), \quad n \rightarrow \infty. \quad (1.32)$$

Тогда, обозначив $z_n = \sqrt{n} \left(\hat{\theta}_n - \theta_0 \right)$, из формулы (1.31) получаем

$$\begin{aligned} n \int \left(m_{\hat{\theta}_n} - m_{\theta_0} \right) d\mathbb{Q}_n &= \frac{1}{2} z_n^T \Sigma_0 z_n + \sqrt{n} z_n^T \int m'_{\theta_0} d\mathbb{G}_n + o_p(1 + n \|\hat{\theta}_n - \theta_0\|^2) = \\ &= \frac{1}{2} z_n^T \Sigma_0 z_n + z_n^T \Sigma_0 \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n + \frac{1}{2} \left(\int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0^{-1} \Sigma \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n - \\ &\quad - \frac{1}{2} \left(\int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0^{-1} \Sigma \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n + o_p(1 + n \|\hat{\theta}_n - \theta_0\|^2) = \\ &= \frac{1}{2} \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0 \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right) - \\ &\quad - \frac{1}{2} \left(\int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0^{-1} \Sigma \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n + o_p(1 + n \|\hat{\theta}_n - \theta_0\|^2), \quad n \rightarrow \infty. \end{aligned} \quad (1.33)$$

Заметим, что последнее слагаемое имеет вид, аналогичный первому слагаемому в формуле (1.31). Запишем это выражение с заменой $\hat{\theta}_n$ на величину

$$\tilde{\theta}_n = \theta_0 - \frac{1}{\sqrt{n}} \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n.$$

Очевидно, что $\tilde{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$, $n \rightarrow \infty$, поэтому условие стохастической дифференцируемости (1.32) верно и для последовательности $\tilde{\theta}_n$. Из формулы (1.31) тогда имеем

$$\begin{aligned} n \int \left(m_{\tilde{\theta}_n} - m_{\theta_0} \right) d\mathbb{Q}_n &= \frac{1}{2} \left(\int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n - \\ &\quad - \sqrt{n} \int \left(m_{\tilde{\theta}_n} - m_{\theta_0} \right) d\mathbb{G}_n + o_p(1 + n \|\tilde{\theta}_n - \theta_0\|^2) = \\ &= -\frac{1}{2} \left(\int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n + o_p(1 + n \|\tilde{\theta}_n - \theta_0\|^2), \quad n \rightarrow \infty. \end{aligned} \quad (1.34)$$

Подставляя выражение из формулы (1.34) в формулу (1.33), получаем

$$\begin{aligned} n \int (m_{\hat{\theta}_n} - m_{\theta_0}) d\mathbb{Q}_n - n \int (m_{\tilde{\theta}_n} - m_{\theta_0}) d\mathbb{Q}_n &= \\ &= \frac{1}{2} \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0 \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right) + \\ &\quad + o_p(1 + n\|\hat{\theta}_n - \theta_0\|^2 + n\|\tilde{\theta}_n - \theta_0\|^2), \quad n \rightarrow \infty. \end{aligned} \quad (1.35)$$

Помимо условия стохастической дифференцируемости (1.32) потребуем выполнения следующих условий:

1. $\int m_{\hat{\theta}_n} d\mathbb{Q}_n - \int m_{\theta_0} d\mathbb{Q}_n = o_p(n^{-1}), \quad n \rightarrow \infty.$
2. $\sqrt{n} (\hat{\theta}_n - \theta_0) = O_p(1).$
3. $\sqrt{n} (\tilde{\theta}_n - \theta_0) = O_p(1).$

Из определения приближенной оценки максимального правдоподобия $\hat{\theta}_n$ и предположения 1 следует, что при достаточно больших n имеет место неравенство

$$\int (m_{\hat{\theta}_n} - m_{\theta_0}) d\mathbb{Q}_n \geq \int (m_{\tilde{\theta}_n} - m_{\theta_0}) d\mathbb{Q}_n.$$

Тогда, в формуле (1.35) имеем:

$$\frac{1}{2} \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right)^T \Sigma_0 \left(z_n + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right) + o_p(1) \geq 0.$$

Однако, так как θ_0 является точкой максимума функции $\theta \mapsto \int m_\theta dQ$, то матрица Σ_0 является отрицательно определенной. Следовательно,

$$\left\| \sqrt{n} (\hat{\theta}_n - \theta_0) + \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n \right\| \xrightarrow{\mathbf{P}} 0, \quad n \rightarrow \infty.$$

Применяя центральную предельную теорему к величине $\Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n$ в предположении существования ковариационной матрицы $\int m'_{\theta_0} (m'_{\theta_0})^T dQ$, полу-

чаем

$$\begin{aligned} \Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n &= \Sigma_0^{-1} \sqrt{n} \left(\int m'_{\theta_0} d\mathbb{Q}_n - \int m'_{\theta_0} dQ \right) \xrightarrow{d} \\ &\xrightarrow{d} N \left(0, \Sigma_0^{-1} \int m'_{\theta_0} (m'_{\theta_0})^T dQ \Sigma_0^{-1} \right), \end{aligned} \quad (1.36)$$

следовательно, случайная величина $\sqrt{n} (\hat{\theta}_n - \theta_0)$ асимптотически имеет нормальное распределение со средним 0 и ковариационной матрицей Σ , где

$$\Sigma = \Sigma_0^{-1} \int m'_{\theta_0} (m'_{\theta_0})^T dQ \Sigma_0^{-1}.$$

Помимо этого формула (1.36) позволяет получить ограниченность по вероятности для последовательности $\sqrt{n} (\tilde{\theta}_n - \theta_0)$. Действительно,

$$\sqrt{n} (\tilde{\theta}_n - \theta_0) = -\Sigma_0^{-1} \int m'_{\theta_0} d\mathbb{G}_n.$$

Ограниченность по вероятности теперь следует из слабой сходимости величины $\int m'_{\theta_0} d\mathbb{G}_n$.

Таким образом, фактически доказана следующая теорема

Теорема 1.6. *Пусть выполняются следующие условия:*

1. *Функция $\theta \mapsto \int t_\theta dQ$ допускает разложение в ряд Тейлора в окрестности точки θ_0 с невырожденной матрицей вторых производных Σ_0 .*
2. *Функция t_θ стохастически дифференцируема в окрестности точки θ_0 по мере $\mathbb{G}_n = \sqrt{n} (\mathbb{Q}_n - Q)$, то есть найдется такая функция $m'_{\theta_0} \in L^2(Q)$, что для всякой последовательности $\tilde{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$, $n \rightarrow \infty$ выполняется*

$$\int (m_{\tilde{\theta}_n} - m_{\theta_0}) d\mathbb{G}_n = (\tilde{\theta}_n - \theta_0)^T \int m'_{\theta_0} d\mathbb{G}_n + o_p(n^{-1/2}), \quad n \rightarrow \infty.$$

3. *Приближенная оценка максимального правдоподобия $\hat{\theta}_n$ состоятельна и имеет скорость сходимости \sqrt{n} : $\sqrt{n} (\hat{\theta}_n - \theta_0) = O_p(1)$.*

4. Приближенная оценка максимального правдоподобия $\hat{\theta}_n$ аппроксимирует точку максимума логарифмической функции правдоподобия $\theta \mapsto \int m_\theta d\mathbb{Q}_n$ с порядком, большим, чем n^{-1} :

$$\int m_{\hat{\theta}_n} d\mathbb{Q}_n - \int m_{\theta_0} d\mathbb{Q}_n = o_p(n^{-1}), \quad n \rightarrow \infty$$

Тогда приближенная оценка максимального правдоподобия $\hat{\theta}_n$ асимптотически нормальна:

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N \left(0, \Sigma_0^{-1} \int m'_{\theta_0} (m'_{\theta_0})^T d\mathbb{Q} \Sigma_0^{-1} \right).$$

Скорость сходимости для величины $\hat{\theta}_n$ можно получить, применив теорему 1.5. Условия 1 и 4 являются почти необходимыми. Наибольшие проблемы представляет проверка условия стохастической дифференцируемости.

Теорема 1.7. *Предположим, что выполняются следующие условия:*

1. *Пространство параметров Θ является подмножеством евклидова пространства \mathbb{R}^p .*
2. *Отображение $\theta \mapsto \log (F_\theta(x) - F_\theta(y))$ дифференцируемо в некоторой окрестности точки θ_0 для $\mu \times \mu$ -п.в. пар (x, y) , $x < y$ с равномерно ограниченным градиентом.*

Тогда $\theta \mapsto m_\theta$ стохастически дифференцируема в окрестности точки θ_0 , то есть для всякой последовательности $\tilde{\theta}_n \xrightarrow{\mathbf{P}} \theta_0$, $n \rightarrow \infty$ выполняется

$$\int (m_{\tilde{\theta}_n} - m_{\theta_0}) d\mathbb{G}_n = (\tilde{\theta}_n - \theta_0)^T \int m'_{\theta_0} d\mathbb{G}_n + o_p(n^{-1/2}), \quad n \rightarrow \infty.$$

Доказательство. Рассмотрим случайный процесс $\xi_t^{(n)}$:

$$\xi_t^{(n)} = \int [\sqrt{n} (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0}) - t^T m'_{\theta_0}] d\mathbb{G}_n. \quad (1.37)$$

Покажем, что $\xi_t^{(n)} \xrightarrow{\mathbf{P}} 0$, $n \rightarrow \infty$ равномерно на множествах вида $\|t\| < C$.

Сходимость по вероятности к константе эквивалентна слабой сходимости, поэтому достаточно показать, что для любой константы C процесс $\xi_t^{(n)} \xrightarrow{d} 0$ как элемент пространства $L^\infty (\|t\| < C)$. Для этого, в свою очередь, достаточно показать сходимость конечномерных распределений процесса и плотность семейства распределений $\{\xi_t^{(n)}\}$.

Из определения процесса (1.37) с очевидностью следует, что $\mathbf{E}\xi_t^{(n)} = 0$. Для дисперсии $\mathbf{D}\xi_t^{(n)}$ имеем выражение

$$\mathbf{D}\xi_t^{(n)} = \int [\sqrt{n} (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0}) - t^T m'_{\theta_0}]^2 dQ.$$

Из условия 2 теоремы тривиально следует дифференцируемость функции $\theta \mapsto m_\theta$. Отсюда для подынтегрального выражения получаем

$$\sqrt{n} (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0}) - t^T m'_{\theta_0} = \sqrt{n} \left(\frac{t^T}{\sqrt{n}} m'_{\theta_0} + o(n^{-1/2}) \right) - t^T m'_{\theta_0} = o(1).$$

Таким образом, подынтегральное выражение поточечно сходится к нулю.

Аналогично теореме 1.5 выводим, что условие 2 теоремы гарантирует наличие функции $M(\delta, t, k)$ с $\int M^2 dQ < \infty$, что

$$\int n (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0})^2 dQ \leq t^2 \int M^2 dQ < \infty.$$

Применяя теорему Лебега о доминируемой сходимости получаем, что для любого t , $\mathbf{D}\xi_t^{(n)} \rightarrow 0$, $n \rightarrow \infty$. Поэтому все конечномерные распределения процесса сходятся по вероятности к нулю.

Осталось показать плотность семейства распределений $\{\xi_t^{(n)}\}$. Применим теорему ЕЗ к классам функций

$$\mathcal{F}_n = \{ \sqrt{n} (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0}) : |t| < C \}.$$

Положим $F_n = C \cdot M$. Аналогично теореме 1.5 имеем $\sup_{f \in \mathcal{F}_n} |f| \leq F_n$ и первые два условия теоремы ЕЗ выполняются. Третье условие также имеет

место:

$$\begin{aligned} \sup_{|s-t|<\delta_n} \int (\sqrt{n} (m_{\theta_0+s/\sqrt{n}} - m_{\theta_0}) - \sqrt{n} (m_{\theta_0+t/\sqrt{n}} - m_{\theta_0}))^2 dQ &\leq \\ &\leq \sup_{|s-t|<\delta_n} \int M^2 |s-t|^2 dQ = C' \delta_n^2 \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Осталось убедиться в убывании энтропийного интеграла (1.25). Фактически, это уже было проделано в теореме 1.5.

Рассмотрим класс функций $\mathcal{F}'_n = \frac{1}{\sqrt{n}} \mathcal{F}_n$ с измеримой мажорантой $F'_n = \frac{1}{\sqrt{n}} F_n$. Ясно, что соответствующие скобочные энтропийные числа классов \mathcal{F}_n и \mathcal{F}'_n совпадают. Подставляя $\delta = C/\sqrt{n}$ и $\mathcal{M}_\delta = \mathcal{F}'_n$ в формулу (1.29), получаем

$$N_{[]} \left(\varepsilon \|F'_n\|_{L^2(Q)}, \mathcal{F}'_n, L^2(Q) \right) \leq C' \left(\frac{C/\sqrt{n}}{\varepsilon \|F_n\|_{L^2(Q)}/\sqrt{n}} \right)^2 \leq C'' \left(\frac{1}{\varepsilon} \right)^d.$$

Отсюда

$$\begin{aligned} \int_0^{\delta_n} \sqrt{\log N_{[]} \left(\varepsilon \|F'_n\|_{L^2(Q)}, \mathcal{F}'_n, L^2(Q) \right)} d\varepsilon &\leq \int_0^{\delta_n} \sqrt{\log C'' + d \log \left(\frac{1}{\varepsilon} \right)} \varepsilon \leq \\ &\leq \delta_n \log C'' + \sqrt{d} \int_{\delta_n}^{\infty} \frac{\sqrt{\log \varepsilon'}}{\varepsilon'^2} d\varepsilon' \rightarrow 0, \quad \delta_n \rightarrow 0. \end{aligned}$$

■

Наконец, объединяя условия теорем 1.7, 1.6, 1.5 и 1.3, получаем следующую теорему, предоставляющую достаточные условия для асимптотической нормальности приближенных оценок максимального правдоподобия в случае интервального цензурирования смешанного типа.

Теорема 1.8. *Пусть выполняются следующие условия:*

1. *Пространство параметров Θ является компактным подмножеством Евклидова пространства \mathbb{R}^p . Истинное значение неизвестного параметра θ_0 принадлежит внутренней части Θ .*

2. Функция распределения $F_\theta(x)$ непрерывна по θ для почти всех x .
3. Параметр θ идентифицируем: для меры μ из (1.17) имеет место соотношение: $F_{\theta_1} = F_{\theta_2} \mu - \text{п.в.} \Rightarrow \theta_1 = \theta_2$.
4. $\mathbf{E}(K) < \infty$.
5. Отображение $\theta \mapsto \log(F_\theta(x) - F_\theta(y))$ дифференцируемо в окрестности точки θ_0 для $\mu \times \mu$ -п.в. пар (x, y) , $x < y$. Обозначим соответствующую производную функции t_θ через t'_θ .
6. Градиент этого отображения равномерно ограничен в окрестности точки θ_0 для $\mu \times \mu$ -п.в. пар (x, y) , $x < y$.
7. Функция $\theta \mapsto \int t_\theta dQ$ допускает разложение в ряд Тейлора в окрестности точки θ_0 с невырожденной матрицей вторых производных Σ_0 .
8. Приближенная оценка максимального правдоподобия $\hat{\theta}_n$ аппроксимирует точку максимума логарифмической функции правдоподобия $\theta \mapsto \int t_\theta dQ_n$ с порядком, большим, чем n^{-1} :

$$\int t_{\hat{\theta}_n} dQ_n - \int t_{\theta_0} dQ_n = o_p(n^{-1}), \quad n \rightarrow \infty$$

Тогда приближенная оценка максимального правдоподобия $\hat{\theta}_n$ строго состоятельна и асимптотически нормальна:

$$\hat{\theta}_n \rightarrow \theta_0, \quad \text{п.н. при } n \rightarrow \infty.$$

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N \left(0, \Sigma_0^{-1} \int t'_{\theta_0} (t'_{\theta_0})^T dQ \Sigma_0^{-1} \right), \quad \text{при } n \rightarrow \infty.$$

Замечание. Условие б теоремы 1.8 может быть несколько ослаблено: вместо равномерной ограниченности в окрестности точки θ_0 градиента отображения $\theta \mapsto \log(F_\theta(x) - F_\theta(y))$ для $\mu \times \mu$ -п.в. пар (x, y) , $x < y$ достаточно

требовать его суммируемости с квадратом относительно всех двумерных проекций распределения Q .

Оценки по минимуму расстояния Кульбака-Лейблера

Пусть, как и раньше, X_1, \dots, X_n — последовательность независимых одинаково распределенных случайных величин с функцией распределения F_0 . Допустим, что у F_0 существует плотность f_0 относительно некоторой доминирующей меры ν . Будем производить оценивание в предположении параметрической модели для функции распределения F_0 , т.е. $F_0 = F_{\theta_0}$ для некоторого $\theta_0 \in \Theta$.

Обозначим через $I(f_\theta; f_0)$ *расстояние Кульбака-Лейблера* между плотностями f_θ и f_0 [40]:

$$I(f_\theta; f_0) = \int \log \frac{f_0}{f_\theta} dF_0, \quad (2.1)$$

Известно [1], что стандартные оценки максимального правдоподобия $\hat{\theta}_n$ обладают свойством минимизации расстояния Кульбака-Лейблера:

$$I(f_{\hat{\theta}_n}, f_{\theta_0}) \rightarrow \min_{\theta \in \Theta} I(f_\theta, f_{\theta_0}), \quad n \rightarrow \infty. \quad (2.2)$$

Это свойство сохраняется и становится крайне важным в том случае, когда предположение о параметрической модели для F_0 не выполняется (т.е. не существует такого $\theta_0 \in \Theta$, что $F_0 = F_{\theta_0}$): среди всех возможных оценок, оценки максимального правдоподобия являются оптимальными в смысле расстояния Кульбака-Лейблера (КЛ-оптимальные) [1, 8, 70].

Несложно видеть, что оценки типа максимального правдоподобия (1.7) при оценивании по выборке с интервальным цензурированием свойством (2.2) не обладают (точнее, они лишь минимизируют расстояние Кульбака-Лейблера между $\frac{dQ}{d\nu}$ и параметрической плотностью (1.4)).

В данной главе будут построены оценки по выборке с интервальным цензурированием, обладающие свойством КЛ-оптимальности (2.2), и рассмотрены их асимптотические свойства.

Наличие таких оценок особенно важно для приложений: как правило, при анализе реальных данных мы не можем быть уверенными в том, что предполагаемая параметрическая модель точна, даже более того, можно исходно предполагать, что она не точна. В этом случае наличие робастных оценок, которые минимизируют некоторое расстояние между истинным и предполагаемым распределением данных, позволяет получить более достоверные результаты.

2.1. Построение оценок

Как и раньше, будем предполагать, что случайная величина X непосредственно не наблюдается. Вместо этого предположим наличие механизма интервального цензурирования смешанного типа $Y = (\Delta_K, T_K, K)$. Оценивание будем производить в предположении параметрической модели для плотности f_0 , однако, будем допускать, что это предположение может и не выполняться. Из формулы (2.1) имеем:

$$\operatorname{argmin}_{\theta \in \Theta} I(f_\theta; f_0) = \operatorname{argmin}_{\theta \in \Theta} \int \log \frac{f_0}{f_\theta} dF_0 = \operatorname{argmax}_{\theta \in \Theta} \int \log f_\theta dF_0.$$

Как и прежде, пусть Y_1, \dots, Y_n — n независимых, одинаково распределенных случайных величин и $\mathcal{L}(Y_i) = \mathcal{L}(Y)$. Обозначим $Y_i = (\Delta_{K^{(i)}}, T_{K^{(i)}}, K^{(i)})$. Пусть \hat{F}_n — некоторая оценка для функции распределения F_0 по Y_1, \dots, Y_n . Оценим величину $I(f_\theta; f_0)$ посредством следующего выражения:

$$\hat{I}_n(f_\theta, f_0) = \int \log \frac{f_0}{f_\theta} d\hat{F}_n$$

Введем оценку $\tilde{\theta}_n$ как точку минимума $\hat{I}_n(f_\theta, f_0)$:

$$\tilde{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} \hat{I}_n(f_\theta, f_0) = \operatorname{argmax}_{\theta \in \Theta} \int f_\theta(x) d\hat{F}_n(x). \quad (2.3)$$

Несложно видеть, что оценки $\hat{\theta}_n$ и $\tilde{\theta}_n$ совпадают при оценивании по полным данным, так как в таком случае в качестве \hat{F}_n можно взять эмпирическую функцию распределения для X_1, \dots, X_n .

В качестве оценки \hat{F}_n будем использовать оценку максимального правдоподобия для F_0 :

$$\hat{F}_n = \operatorname{argmax}_{F \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{K^{(i)}+1} \Delta_{K,j}^{(i)} \log [F(T_{K,j}^{(i)}) - F(T_{K,j-1}^{(i)})] = \operatorname{argmax}_{F \in \mathcal{F}} \int m_F d\mathbb{Q}_n, \quad (2.4)$$

здесь \mathcal{F} — класс всех функций распределения с носителем $[0, \infty)$, а логарифм плотности m_F вводится аналогично формуле (1.5):

$$m_F(\delta_k, t_k, k) = \sum_{j=1}^{k+1} \delta_{k,j} \log [F(t_{k,j}) - F(t_{k,j-1})].$$

2.2. Оценивание линейных функционалов относительно мер

Как видно из теорем 1.1 и 1.3 необходимым условием для состоятельности оценок параметров является наличие законов больших чисел для линейных функционалов относительно мер вида

$$\int c(x) d\hat{F}_n(x) \xrightarrow{\mathbf{P}} \int c(x) dF_0(x), \quad \text{при } n \rightarrow \infty. \quad (2.5)$$

В качестве функции $c(x)$ при этом выступает либо логарифм некоторой плотности: $c(x) = \log f_\theta(x)$, либо его супремум по некоторому шару B_{θ_0} : $c(x) = \sup_{\theta \in B_{\theta_0}} \log f_\theta(x)$. Функция \hat{F}_n тут — некоторая оценка функции распределения f_0 по выборке с интервальным цензурированием Y_1, \dots, Y_n .

Общая состоятельность гладких функционалов вида (2.5) изучалась в работе [60] в случае случайного правого цензурирования. В этом случае в качестве естественной оценки \hat{F}_n использовалась оценка Каплана-Мейера [34]. Случай интервального цензурирования первого типа (модель единичного контроля текущего состояния) рассматривался в работах [23, 26]. Детальному рассмотрению случая интервального цензурирования второго типа посвящены работы [18–20].

Случай интервального цензурирования смешанного типа не сильно отличается от случая интервального цензурирования второго типа. Оказалось, что результаты из [18] достаточно несложно обобщаются на случай произвольного числа наблюдений за состоянием (то есть интервального цензурирования смешанного типа). Это и будет проделано в текущем разделе.

Для начала нам потребуются некоторые определения и результаты из работ [23], [19, 20] и [9].

2.2.1. Используемые обозначения и теоремы

Функционалы, дифференцируемые вдоль путей

Пусть \mathcal{P} — класс распределений на измеримом пространстве $(\mathcal{X}, \mathcal{B})$. Предположим, что \mathcal{P} доминируется сигма-конечной мерой l . Зафиксируем меру $P \in \mathcal{P}$. Обозначим через p плотность P относительно меры l .

Рассмотрим некоторое однопараметрическое семейство мер $\{P_t\} \subset \mathcal{P}$, где $t \in (0, \delta)$ для некоторого $\delta > 0$. Обозначим через p_t соответствующую плотность относительно меры l .

Определение 6. Семейство распределений $\{P_t\}$ назовем дифференцируемым в смысле Хеллингера или регулярным [50, 66], если найдется такая функ-

ция $g \in L^2(P)$, что

$$\int \left(\frac{\sqrt{p_t} - \sqrt{p}}{t} - \frac{1}{2}g\sqrt{p} \right)^2 dl \rightarrow 0, \quad \text{при } t \rightarrow 0. \quad (2.6)$$

Регулярность в смысле Хеллинджера можно рассматривать как L^2 -версию дифференцируемости $t \mapsto \log p_t(x)$ в точке $t = 0$ (при $p_0 = p$). Действительно, при фиксированном x имеем

$$\lim_{t \rightarrow 0} \frac{\sqrt{p_t} - \sqrt{p_0}}{t} = \frac{1}{2\sqrt{p_0}} \frac{d}{dt} p_t \Big|_{t=0} = \frac{1}{2}\sqrt{p_0} \left(\frac{d}{dt} \log p_t \Big|_{t=0} \right).$$

Определение 7. Рассмотрим всевозможные однопараметрические семейства $\{P_t\} \subset \mathcal{P}$. Касательным пространством $T(P)$ к мере P назовем множество всех $g \in L^2(P)$, для которых выполняется условие (2.6). Будем предполагать семейство мер \mathcal{P} настолько богатым, что $T(P)$ является линейным пространством.

Замечание. В работе [9] показано, что $T(P) \subset L^2_0(P)$, то есть $\int g^2 dP = 0$.

Рассмотрим абстрактный функционал $K : \mathcal{P} \rightarrow \mathbb{R}$.

Определение 8. Функционал K назовем дифференцируемым вдоль путей в точке P , если найдется такое непрерывное линейное отображение $K'_P : T(P) \rightarrow \mathbb{R}$, что для любого регулярного семейства $\{P_t\}$:

$$\lim_{t \rightarrow 0} \frac{K(P_t) - K(P)}{t} = K'_P(g),$$

где логарифмическая производная g определяется условием (2.6).

Касательное пространство $T(P)$ является подпространством гильбертова пространства $L^2(P)$, поэтому линейный непрерывный функционал K'_P может быть продолжен до линейного непрерывного функционала \overline{K}'_P на всем $L^2(P)$. По теореме Рисса-Фреше найдется элемент $\kappa_P \in L^2(P)$, что

$$\overline{K}'_P(g) = \langle \kappa_P, g \rangle, \quad \forall g \in T(P). \quad (2.7)$$

Определение 9. Градиентом функционала K в точке P называется элемент $k_P \in L^2(P)$, удовлетворяющий соотношению (2.7).

Конечно, продолжение отображения K'_P можно строить на любое подпространство $H \subset L^2(P)$, лишь бы $\overline{T(P)} \subset H$. Каждому такому продолжению соответствует свой градиент, поэтому, вообще говоря, в своем определении градиент не единственный.

Определение 10. Каноническим градиентом называется градиент \tilde{k}_P , полученный продолжением линейного функционала K'_P с $T(P)$ на $\overline{T(P)}$.

Ясно, что канонический градиент единственен. Кроме этого, он обладает наименьшей нормой, так как может быть получен ортогональной проекцией любого другого градиента на $\overline{T(P)}$.

Определение 11. Оценка $\hat{K}_n = \hat{k}(X_1, \dots, X_n)$ значения функционала $K(P)$ называется регулярной, если найдется такая случайная величина Z , что для всякого регулярного по Хеллинджеру семейства $\{P_t\} \subset \mathcal{P}$ и для всякой последовательности $\eta_n = O(n^{-1/2})$ выполняется:

$$\sqrt{n} \left(\hat{K}_n - K(P_{\eta_n}) \right) \xrightarrow{d} Z, \quad \text{при } n \rightarrow \infty,$$

и $\mathcal{L}(X_1) = \dots = \mathcal{L}(X_n) = P_{\eta_n}$.

Распределение случайной величины Z определяется известной теоремой о свертке Хаека–Ле Кама (см. [9], теорема 2.3.1).

Теорема E4 (Hájek–Le Cam). Предположим, что

1. Функционал K дифференцируем вдоль путей в точке $P \in \mathcal{P}$.
2. \hat{K}_n является регулярной оценкой значения функционала $K(P)$.

Тогда найдутся такие независимые случайные величины Z_0 и Δ_0 , что:

1. $\mathcal{L}(Z) = \mathcal{L}(Z_0 + \Delta_0)$.
2. $\mathcal{L}(Z_0) = N(0, \|\tilde{\kappa}_P\|_P^2)$.

Таким образом, $\|\tilde{\kappa}_P\|_P^2$ является наименьшей возможной дисперсией, которой может достигать регулярная оценка. В таком случае $\mathcal{L}(Z) = \mathcal{L}(Z_0)$ и оценка \hat{K}_n называется *асимптотически эффективной*.

Далее в основном нас будут интересовать линейные функционалы вида

$$K(F) = \int c(x) dF(x)$$

для некоторой функции c . Согласно теореме А.5.2 из [9], достаточным условием дифференцируемости таких функционалов в точке $F \in \mathcal{F}$ является

$$\int c^2(x) dF(x) < \infty. \quad (2.8)$$

Например, это условие выполняется для ограниченной функции $c(x)$.

Функционалы, заданные неявно

Пусть распределение Q определяется некоторым другим распределением P . Рассмотрим задачу оценивания некоторого функционала K в точке P . Будем считать, что функционал $K(P)$ задан неявно в виде некоторого функционала $\Xi(Q)$ относительно индуцированной меры Q . Асимптотическая теория оценивания функционалов в таком случае рассматривалась в работах [9, 66]. Предположим, что отображение Q может быть получено в виде $Q = PC^{-1}$ для некоторого отображения \mathcal{C} .

Отображение \mathcal{C} индуцирует не только распределение Q , но и регулярные семейства, как следует из теоремы А.5.5 из [9]:

Теорема Е5 (Bickel–Klaassen–Ritov–Wellner). Пусть задан класс распределений \mathcal{P} , абсолютно непрерывный относительно меры l , на некотором

пространстве $(\mathcal{X}, \mathcal{B})$. Пусть семейство $\{P_t\}$ регулярно в точке $P \in \mathcal{P}$:

$$\int \left(\frac{\sqrt{p_t} - \sqrt{p}}{t} - \frac{1}{2}g\sqrt{p} \right)^2 dl \rightarrow 0, \quad \text{при } t \rightarrow 0,$$

$g \in L_0^2(l)$. Пусть P является распределением некоторой случайной величины X , а $\mathcal{C} : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{Z}, \mathcal{C})$ — измеримое отображение. Пусть мера $l\mathcal{C}^{-1}$ — сигма-конечная. Предположим, что индуцированные распределения $Q_t = P_t\mathcal{C}^{-1}$ и $Q = P\mathcal{C}^{-1}$ на $(\mathcal{Z}, \mathcal{C})$ абсолютно непрерывны относительно меры $l\mathcal{C}^{-1}$ с плотностями q_t и q соответственно. Тогда семейство $\{Q_t\}$ регулярно и выполняется

$$\int \left(\frac{\sqrt{q_t} - \sqrt{q}}{t} - \frac{1}{2}\bar{g}\sqrt{q} \right)^2 dl\mathcal{C}^{-1} \rightarrow 0, \quad \text{при } t \rightarrow 0,$$

при этом $\bar{g}(z) = \mathbf{E}(g(X)|\mathcal{C} = z)$.

Таким образом отображение \mathcal{C} индуцирует оператор между касательными пространствами $T(P)$ и $T(Q)$:

$$A_P : g(\cdot) \mapsto \mathbf{E}(g(X)|\mathcal{C} = \cdot). \quad (2.9)$$

Очевидно, что оператор A является линейным и непрерывным.

Достаточные условия дифференцируемости для Ξ в точке Q дает следующая теорема из [66]:

Теорема Е6 (van der Vaart). Пусть $\mathcal{Q} = \mathcal{P}\mathcal{C}^{-1}$ является образом набора распределений \mathcal{P} при измеримом преобразовании \mathcal{C} . Предположим, что функционал $\Xi : \mathcal{Q} \rightarrow \mathbb{R}$ может быть представлен в виде $\Xi(Q_P) = K(P)$, где функционал $K(P)$ дифференцируем в точке P с каноническим градиентом $\tilde{\kappa}_P$.

Функционал Ξ дифференцируем в точке $Q_P \in \mathcal{Q}$ тогда и только тогда, когда

$$\tilde{\kappa}_P \in \text{Im}(A_P^*),$$

где A_P^* — оператор, сопряженный к оператору A_P , определенному в (2.9). В таком случае канонические градиенты $\tilde{\xi}_{Q_P}$ оператора Ξ в точке Q_P и $\tilde{\kappa}_P$ оператора K в точке P связаны соотношением

$$\tilde{\kappa}_P = A_P^* \tilde{\xi}_{Q_P}. \quad (2.10)$$

2.2.2. Вычисление информационных границ в случае интервального цензурирования

Модель интервального цензурирования смешанного типа является типичным примером модели с потерей информации: вместо n независимых одинаково распределенных случайных величин X_1, \dots, X_n с функцией распределения F наблюдается менее информативные случайные величины Y_1, \dots, Y_n с $Y_i = (\Delta_{K^{(i)}}, T_{K^{(i)}}, K^{(i)})$.

Как видно, случайная величина Y получена из случайной величины (X, T, K) при помощи измеримого преобразования \mathcal{C} :

$$\mathcal{C}(x, t, k) = (\delta_k, t_k, k).$$

Случайная величина Y задана на измеримом пространстве $(\mathcal{Y}, \mathcal{D})$, определенном формулой (1.3). Введем на \mathcal{D} меру ν' :

$$\nu'(\{\delta_k\} \times B_k \times \{k\}) = \#(\delta_k) \lambda_k(B_k),$$

где $\#$ обозначает считающую меру, а λ_k — k — мерную меру Лебега.

Тогда при введенных выше допущениях распределение Q случайной величины Y индуцируется распределением F и имеет следующую плотность $q_{F,H}$ относительно меры ν' (здесь $p_k = \mathbf{P}(K = k)$):

$$q_{F,H}(\delta_k, t_k, k) = p_k h_k(t_k) \prod_{i=1}^{k+1} [F(t_{k,i}) - F(t_{k,i-1})]^{\delta_{k,i}}. \quad (2.11)$$

Потребуем дополнительно выполнения следующих условий:

1. Функция распределения F принадлежит классу функций распределения \mathcal{F} , абсолютно непрерывных относительно меры Лебега λ и имеющих ограниченный носитель: $\text{supp } F \subset [0, M]$, $F \in \mathcal{F}$. Обозначим через f плотность F .
2. Случайная величина $T = (T_1, \dots)$ принадлежит классу распределений $\mathcal{H} = \bigcup_{k=1}^{\infty} \mathcal{H}_k$, где \mathcal{H}_k представляет собой класс распределений, абсолютно непрерывных относительно k -мерной меры Лебега. Обозначим соответствующую плотность через h_k . Как и раньше будем предполагать независимость X от распределения $H = (K, T_K)$.
3. Функция распределение F непрерывна и для всех $a < b$ из $0 < F(a) < F(b) < 1$ следует $\mu((a, b)) > 0$, где μ — мера, определенная формулой (1.17).

Касательное пространство $T(P)$ определяется всеми регулярными семействами мер, которые можно получить из классов распределений \mathcal{F} и \mathcal{H} . Используя пример 3.2.1 из [9] можно показать, что в приведенных выше допущениях на семейства \mathcal{F} и \mathcal{H} настолько большие, что $T(F) = L_0^2(F)$ и $T(H_k) = L_0^2(H_k)$ для любого k : зафиксируем $g \in L_0^2(F)$, тогда семейство $\{F_t\}$ с плотностями

$$f_t(x) = \frac{f(1 + tg(x))}{\int_{1+tg(x) \geq 0} f(1 + tg(x)) d\lambda}$$

регулярно с логарифмической производной g . Аналогичным образом строится регулярное семейство для произвольного $e_k \in L_0^2(H_k)$. Отсюда, пользуясь независимостью, можно показать, что касательное пространство $T(P)$ имеет следующий вид:

$$T(P) = L_0^2(F) + \bigcup_{k=0}^{\infty} L_0^2(H_k).$$

Тогда оператор $A_{F,H} : T(P) \rightarrow T(Q)$, определенный формулой (2.9), будет действовать по правилу

$$(A_{F,H}(g + e))(\delta_k, t_k, k) = \mathbf{E}(g(X) + e(T_K) | (\Delta_K, T_K, K) = (\delta_k, t_k, k)).$$

Оператор $A_{F,H}$ можно представить в виде суммы операторов $A^{(1)} + A^{(2)}$:

$$A_{F,H}(g + e) = \mathbf{E}(g(X) | (\Delta_K, T_K, K)) + \mathbf{E}(e(T_K) | (\Delta_K, T_K, K)) = A^{(1)}g + A^{(2)}e.$$

Очевидно, что $A^{(2)}$ является тождественным оператором, т.е. $A^{(2)}e = e$, оператор $A^{(1)}$ в силу независимости X и (K, T_K) действует по правилу:

$$(A^{(1)}g)(\delta_k, t_k, k) = \sum_{i=1}^{k+1} \frac{\delta_{k,i} \int_{t_{k,i-1}}^{t_{k,i}} g(x) dF(x)}{F(t_{k,i}) - F(t_{k,i-1})}, \quad g \in L_0^2(F). \quad (2.12)$$

Несложно видеть, что $\text{Im } A^{(1)}$ и $\text{Im } A^{(2)}$ являются ортогональными подпространствами $L^2(Q_{F,H})$.

Отметим, что оператор $A^{(1)}$ имеет вид, аналогичный оператору L^1 из [19], а при $\mathbf{P}(K = 2) = 1$ в точности совпадает с ним (в таком случае $T(P) = L_0^2(F) + L_0^2(H_2)$).

В случае интервального цензурирования смешанного типа функционал $K(F)$ непосредственно наблюдать невозможно, так как случайная величина X не наблюдается. Вместо этого доступен функционал $\Xi(Q) = K(F)$ относительно индуцированной меры Q . Для установления дифференцируемости функционала $\Xi(Q)$ в точке Q воспользуемся теоремой Еб.

Для начала заметим, что пространства $\text{Im } A^{(1)}$ и $\text{Im } A^{(2)}$ ортогональны, поэтому $A_{F,H}^* = (A^{(1)})^* + (A^{(2)})^*$. Кроме этого $\tilde{\kappa}_{F,H} \in L_0^2(F)$, и, следовательно, для всякого элемента $b \in \text{Im } (A^{(2)})^*$ выполняется: $\langle \tilde{\kappa}_{F,H}, b \rangle = 0$. Таким образом из условия (2.10) выводим:

$$\tilde{\kappa}_{F,H} = A_{F,H}^* \tilde{\xi}_Q = (A^{(1)})^* \tilde{\xi}_Q + (A^{(2)})^* \tilde{\xi}_Q = (A^{(1)})^* \tilde{\xi}_Q.$$

Сопряженный оператор $(A^{(1)})^*$, действующий из $L^2(Q)$ допускает представление в виде условного математического ожидания:

$$(A^{(1)})^*b(x) = \mathbf{E}(b(\Delta_K, T_K, K)|X = x) = \sum_{k=1}^{\infty} p_k \mathbf{E}(b_k(\Delta_k, T_k, k)|X = x).$$

Для вычисления математического ожидания $\mathbf{E}(b_k(\Delta_k, T_k, k)|X = x)$ рассмотрим три случая:

1. $X < T_{k,1}$. В таком случае $\Delta_k = (1, 0, \dots, 0)$, и выводим:

$$\begin{aligned} E_k^{(1)}(x) &= \mathbf{E}(b_k(\Delta_k, T_k, k)|X = x, T_{k,1} > x) = \\ &= \int_x^S \int_{t_1}^S \cdots \int_{t_{k-1}}^S b_k(\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1. \end{aligned}$$

2. $T_{k,i-1} < X < T_{k,i}$ для некоторого $1 < i \leq k$. Здесь $\Delta_k = (0, \dots, 1, \dots, 0)$ и имеем:

$$\begin{aligned} E_k^{(i)}(x) &= \mathbf{E}(b_k(\Delta_k, T_k, k)|X = x, T_{k,i-1} < x < T_{k,i}) = \\ &= \int_0^x \int_{t_1}^x \cdots \int_{t_i=t_{i-1}}^x \int_{t_{i+1}=x}^S \int_{t_{i+1}}^S \cdots \int_{t_{k-1}}^S b_k(\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1. \end{aligned}$$

3. $X > T_{k,k}$. При этом $\Delta_k = (0, \dots, 0, 1)$ и

$$\begin{aligned} E_k^{(k+1)}(x) &= \mathbf{E}(b_k(\Delta_k, T_k, k)|X = x, T_{k,k} < x) = \\ &= \int_0^x \int_{t_1}^x \cdots \int_{t_{k-1}}^x b_k(\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1. \end{aligned}$$

Соединяя все вместе, имеем:

$$(A^{(1)})^*b(x) = \mathbf{E}(b(\Delta_K, T_K, K)|X = x) = \sum_{k=1}^{\infty} p_k \sum_{i=1}^{k+1} E_k^{(i)}(x). \quad (2.13)$$

Всюду далее будем предполагать, что $\tilde{\xi}_Q \in \text{Im } A^{(1)}$. Это означает, что необходимо исследовать разрешимость следующего операторного уравнения относительно $g \in L_0^2(F)$:

$$\tilde{\kappa}_F(x) = \left[(A^{(1)})^* A^{(1)} \right] g(x) \quad F - \text{п.в.} \quad (2.14)$$

Оператор $(A^{(1)})^* A^{(1)}$ такого вида носит название *информационного оператора* [18–20, 23] и, как будет показано далее, имеет достаточно специальную структуру.

Случай интервального цензурирования первого типа

В качестве примера, когда уравнение (2.14) разрешимо явно, приведем случай интервального цензурирования первого типа, детально разобранный в [26]. В этом частном случае интервального цензурирования смешанного типа имеем $\mathbf{P}(K = 1) = 1$. Наблюдаемая величина Y состоит из момента контроля состояния объекта T и метки состояния $\Delta = 1_{[X \leq T]}$. X и T предполагаются независимыми. Условие Шика в таком случае превратится в условие доминирования распределения F распределением H случайной величины T .

При этом оператор $A^{(1)}$ примет вид

$$\left(A^{(1)} g \right) (t, \delta) = \frac{\delta \int_0^t g dF}{F(t)} + \frac{(1 - \delta) \int_t^S g dF}{1 - F(t)},$$

сопряженным к нему будет оператор $(A^{(1)})^*$:

$$\left((A^{(1)})^* b \right) (x) = \mathbf{E} (b(\Delta, T) | X = x) = \int_x^S b(1, t) h(t) dt + \int_0^x b(0, t) h(t) dt,$$

где $h(t)$ — плотность случайной величины T .

Обозначая

$$\varphi(x) = \int_x^S g(t) dF(t),$$

выпишем уравнение (2.14):

$$\int_0^x \frac{\varphi(t)}{1-F(t)} h(t) dt - \int_x^S \frac{\varphi(t)}{F(t)} h(t) dt = \tilde{\kappa}_F(x), \quad F - \text{п.н.}$$

Предположим, что $\tilde{\kappa}_F(x)$ непрерывно дифференцируем, тогда, дифференцируя обе части уравнения по x , получаем

$$\frac{\varphi(x)}{1-F(x)} h(x) + \frac{\varphi(x)}{F(x)} h(x) = \tilde{\kappa}'_F(x),$$

откуда, в предположении $h > 0$, получаем

$$\varphi(x) = \tilde{\kappa}'_F(x) \frac{F(x)(1-F(x))}{h(x)},$$

и, наконец, используя тот факт, что $a \in L_0^2(F)$, т.е. $\int_0^S a(t) dF(t) = 0$, получаем в явном виде канонический градиент:

$$\begin{aligned} \tilde{\xi}_Q(t, \delta) &= A^{(1)}g(t, \delta) = \frac{\delta \int_0^t g dF}{F(t)} + \frac{(1-\delta) \int_t^S g dF}{1-F(t)} = \\ &= -\frac{\delta \varphi(t)}{F(t)} + \frac{(1-\delta)\varphi(t)}{1-F(t)} = \begin{cases} -\tilde{\kappa}'_F(t) \frac{1-F(t)}{h(t)}, & \delta = 1, \\ \tilde{\kappa}'_F(t) \frac{F(t)}{h(t)}, & \delta = 0. \end{cases} \end{aligned} \quad (2.15)$$

При этом минимально возможная дисперсия для оценивания функционала $\Xi(Q)$ равна

$$\left\| \tilde{\xi}_Q \right\|_Q^2 = \int_0^S (\tilde{\kappa}'_F(t))^2 \frac{F(x)(1-F(x))}{h(x)} dx.$$

Случай интервального цензурирования смешанного типа

Случай интервального цензурирования смешанного типа (как и частный случай интервального цензурирования второго типа, разобранный в [18–20]) достаточно сильно отличается от интервального цензурирования первого типа. Так, как будет видно дальше, из информационного уравнения (2.14) невозможно будет получить канонический градиент в явном виде. Вместо этого мы

получим интегральное уравнение второго рода относительно введенной выше функции $\varphi(x)$. Мы покажем, что большинство результатов работы [19] при этом переносятся на случай интервального цензурирования смешанного типа с некоторыми модификациями.

Как и раньше, для $g \in L_0^2(F)$ обозначим

$$\varphi(x) = \int_x^S g(t) dF(t).$$

Перепишем оператор $A^{(1)}$ в терминах функции φ . Для этого заметим, что $\varphi(0) = \varphi(S) = 0$. Поэтому, используя $t_{k,0} = 0, t_{k,k+1} = S$, получаем

$$\begin{aligned} (A^{(1)}g)(\delta_k, t_k, k) &= \sum_{i=1}^{k+1} \frac{\delta_{k,i} \int_{t_{k,i-1}}^{t_{k,i}} g(x) dF(x)}{F(t_{k,i}) - F(t_{k,i-1})} = \\ &= \frac{\delta_{k,1} \int_{t_{k,0}}^{t_{k,1}} g(x) dF(x)}{F(t_{k,1}) - F(t_{k,0})} + \sum_{i=2}^k \frac{\delta_{k,i} \int_{t_{k,i-1}}^{t_{k,i}} g(x) dF(x)}{F(t_{k,i}) - F(t_{k,i-1})} + \frac{\delta_{k,k+1} \int_{t_{k,k}}^{t_{k,k+1}} g(x) dF(x)}{F(t_{k,k+1}) - F(t_{k,k})} = \\ &= -\frac{\delta_{k,1}\varphi(t_{k,1})}{F(t_{k,1})} - \sum_{i=2}^k \frac{\delta_{k,i}(\varphi(t_{k,i}) - \varphi(t_{k,i-1}))}{F(t_{k,i}) - F(t_{k,i-1})} + \frac{\delta_{k,k+1}\varphi(t_{k,k})}{1 - F(t_{k,k})}. \end{aligned} \quad (2.16)$$

Выпишем информационное уравнение (2.14). Для начала вычислим величины $E_k^{(i)}(x)$ в случае, когда $b_k(\delta_k, t_k, k) = (A^{(1)}g)(\delta_k, t_k, k)$.

Для $E_k^{(1)}$ из формулы (2.16) с учетом того, что $\Delta_k = (1, 0, \dots, 0)$, получаем:

$$\begin{aligned}
E_k^{(1)}(x) &= \mathbf{E} \left(\left(A^{(1)}g \right) (\Delta_k, T_k, k) | X = x, T_{k,1} > x \right) = \\
&= \int_x^S \int_{t_1}^S \cdots \int_{t_{k-1}}^S \left(A^{(1)}g \right) (\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= - \int_x^S \int_{t_1}^S \cdots \int_{t_{k-1}}^S \frac{\varphi(t_1)}{F(t_1)} h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= - \int_x^S \frac{\varphi(t_1)}{F(t_1)} \int_{t_1}^S \cdots \int_{t_{k-1}}^S h(t_1, \dots, t_k) dt_k \dots dt_1 = - \int_x^S \frac{\varphi(t_1)}{F(t_1)} h_1^{(k)}(t_1) dt_1,
\end{aligned}$$

где $h_1^{(k)}$ — условная плотность $T_{k,1}$ при $K = k$.

Далее, вычислим $E_k^{(i)}$, $1 < i \leq k$. В этом случае $\Delta_k = (0, \dots, 1, 0, \dots, 0)$.

$$\begin{aligned}
E_k^{(i)}(x) &= \mathbf{E} \left(\left(A^{(1)}g \right) (\Delta_k, T_k, k) | X = x, T_{k,i-1} < x < T_{k,i} \right) = \\
&= \int_0^x \int_{t_1}^x \cdots \int_{t_i=t_{i-1}}^x \int_{t_{i+1}=x}^S \int_{t_{i+1}}^S \cdots \int_{t_{k-1}}^S \left(A^{(1)}g \right) (\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= - \int_0^x \int_{t_1}^x \cdots \int_{t_i=t_{i-1}}^x \int_{t_{i+1}=x}^S \int_{t_{i+1}}^S \cdots \int_{t_{k-1}}^S \frac{\varphi(t_{k,i}) - \varphi(t_{k,i-1})}{F(t_{k,i}) - F(t_{k,i-1})} h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= - \int_0^x \int_x^S \frac{\varphi(t_i) - \varphi(t_{i-1})}{F(t_i) - F(t_{i-1})} h_{i,i-1}^{(k)}(t_{i-1}, t_i) dt_i dt_{i-1}.
\end{aligned}$$

Здесь $h_{i,i-1}^{(k)}$ — условная плотность $(T_{k,i}, T_{k,i-1})$ при $K = k$.

Аналогично для $E_k^{(k+1)}$ имеем $\Delta_k = (0, \dots, 0, 1)$. Поэтому,

$$\begin{aligned}
E_k^{(k+1)}(x) &= \mathbf{E} \left(\left(A^{(1)} g \right) (\Delta_k, T_k, k) | X = x, T_{k,k} < x \right) = \\
&= \int_0^x \int_{t_1}^x \cdots \int_{t_{k-1}}^x \left(A^{(1)} g \right) (\Delta_k, t_k, k) h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= \int_0^x \int_{t_1}^x \cdots \int_{t_{k-1}}^x \frac{\varphi(t_k)}{1 - F(t_k)} h(t_1, \dots, t_k) dt_k \dots dt_1 = \\
&= \int_0^x \int_{t_{k-1}}^x \frac{\varphi(t_k)}{1 - F(t_k)} h_{k,k-1}^{(k)}(t_{k-1}, t_k) dt_k dt_{k-1},
\end{aligned}$$

где $h_{k,k-1}^{(k)}$ — условная плотность $(T_{k,k}, T_{k,k-1})$ при $K = k$.

Таким образом, получаем следующее информационное уравнение относительно функции φ :

$$\begin{aligned}
\tilde{\kappa}_F(x) &= \sum_{k=1}^{\infty} p_k \left(- \int_x^S \frac{\varphi(u)}{F(u)} h_1^{(k)}(u) du - \right. \\
&\quad - \sum_{i=2}^k \int_0^x \int_x^S \frac{\varphi(v) - \varphi(u)}{F(v) - F(u)} h_{i,i-1}^{(k)}(u, v) dv du + \\
&\quad \left. + \int_0^x \int_u^x \frac{\varphi(v)}{1 - F(v)} h_{k,k-1}^{(k)}(u, v) dv du, \right) ,
\end{aligned}$$

откуда, переставляя местами суммирование и интегрирование, получаем

$$\begin{aligned}
\tilde{\kappa}_F(x) &= - \int_x^S \frac{\varphi(u)}{F(u)} \bar{h}_1(u) dt_1 - \\
&\quad - \int_0^x \int_x^S \frac{\varphi(v) - \varphi(u)}{F(v) - F(u)} \bar{h}(u, v) dv du + \int_0^x \int_u^x \frac{\varphi(v)}{1 - F(v)} \bar{h}_{\infty, \infty-1}(u, v) dv du, \quad (2.17)
\end{aligned}$$

где

$$\begin{aligned}\bar{h}_1(u) &= \sum_{k=1}^{\infty} p_k h_1^{(k)}(u), \\ \bar{h}_{i,i-1}(u, v) &= \sum_{k=1}^{\infty} p_k h_{i,i-1}^{(k)}(u, v), \\ \bar{h}(u, v) &= \sum_{i=2}^k \bar{h}_{i,i-1}(u, v), \\ \bar{h}_{\infty, \infty-1}(u, v) &= \sum_{k=1}^{\infty} p_k h_{k,k-1}^{(k)}(u, v).\end{aligned}$$

Несложно видеть, что \bar{h}_1 является плотностью первого момента контроля состояния, $\bar{h}_{i,i-1}$ совместной плотностью i -ого и $i - 1$ -ого момента контроля состояния, а $\bar{h}_{\infty, \infty-1}(u, v)$ — совместная плотность последнего и предпоследнего момента контроля состояния.

Предположим, что градиент $\tilde{\kappa}_F$ всюду непрерывно дифференцируем. Продифференцируем уравнение (2.17):

$$\begin{aligned}\tilde{\kappa}'_F(x) &= \frac{\varphi(x)}{F(x)} \bar{h}_1(x) + \int_0^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \bar{h}(t, x) dt - \\ &\quad - \int_x^S \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \bar{h}(x, t) dt + \frac{\varphi(x)}{1 - F(x)} \bar{h}_{\infty}(x), \quad (2.18)\end{aligned}$$

где $\bar{h}_{\infty}(x) = \sum_{k=1}^{\infty} p_k h_k^{(k)}(x)$ — плотность последнего момента контроля состояния. Обозначая

$$\bar{d}_F(x) = \frac{F(x)(1 - F(x))}{\bar{h}_1(x)(1 - F(x)) + \bar{h}_{\infty}(x)F(x)},$$

из (2.18) получаем

$$\varphi(x) + \bar{d}_F(x) \left(\int_0^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \bar{h}(t, x) dt - \int_x^S \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) = \tilde{\kappa}'(x) \bar{d}_F(x). \quad (2.19)$$

Полученное интегральное уравнение похоже по своей структуре на интегральное уравнение (2.5) из [18].

Уравнение (2.19) является интегральным уравнением второго рода:

$$\varphi(x) - \int_0^S K_F(x, t) \varphi(t) dt = a(x). \quad (2.20)$$

Ядро интегрального оператора имеет вид

$$K_F(x, t) = \frac{\bar{d}_F(x) \bar{D}_F(x, t)}{1 + \bar{d}_F(x) \int_0^S \bar{D}_F(x, t) dt}, \quad (2.21)$$

где

$$\bar{D}_F(x, t) = \begin{cases} \frac{\bar{h}(t, x)}{F(x) - F(t)}, & t \leq x, \\ \frac{\bar{h}(x, t)}{F(t) - F(x)}, & t \geq x. \end{cases} \quad (2.22)$$

Правая часть $a(x)$ при этом равна

$$a(x) = \frac{\tilde{\kappa}'(x) \bar{d}_F(x)}{1 + \bar{d}_F(x) \int_0^S \bar{D}_F(x, t) dt}. \quad (2.23)$$

Отметим, что вообще говоря, оператор

$$\varphi(\cdot) \mapsto \int_0^S K(\cdot, t) dt$$

не является компактным без введения дополнительных ограничений на функцию $\bar{h}(t, x)$: из формулы (2.22) видно, что знаменатель функции \bar{D}_F может

быть сколько угодно мал, если t близко к x , а $\bar{h}(t, x)$ отлична от нуля. В терминах моментов контроля состояния это соответствует случаю, когда два последовательных момента контроля состояния могут быть сколь угодно близки с вероятностью, отличной от нуля, и, следовательно, фактически достаточно точно известна реализация случайной величины X .

Стоит отметить, что условие отделимости

$$\mathbf{P}(T_{k,i} - T_{k,i-1} < \varepsilon) = 0, \quad 1 \leq i \leq k \quad (2.24)$$

для некоторого ε является достаточно естественным в реальных приложениях: как правило, два последовательных момента контроля состояния всегда отделены ненулевым расстоянием во времени.

Достаточные условия разрешимости уравнения вида (2.20) с ядром (2.21) были приведены в работе [19]. Эти результаты слабо зависят от конкретного вида функций \bar{D}_F , \bar{d}_F , $\bar{h}(t, x)$, что позволяет после дополнительной модификации применить их не только в оригинальном случае интервального цензурирования второго типа, но и в случае интервального цензурирования смешанного типа.

Потребуем выполнения следующих условий:

- S1) Функции \bar{h}_1 и \bar{h}_∞ непрерывны и $\bar{h}_1(x) + \bar{h}_\infty(x) > 0$ на $[0, S]$.
- S2) Функция $\bar{h}(u, v)$ непрерывна.
- S3) Функция распределения F с $\text{supp } F = [0, M]$ непрерывна, либо кусочно-постоянная с конечным числом скачков τ_1, \dots, τ_m на $[0, M]$. Введем множество $\mathcal{D} = \{\tau_0, \tau_1, \dots, \tau_m, \tau_\infty\}$, где $\tau_1 < \dots < \tau_m$ — точки скачков, а τ_0, τ_∞ — концы носителя F .
- S4) $\mathbf{P}(T_{k,i} - T_{k,i-1} < \varepsilon) = 0$, где $\varepsilon = 1/2M$.
- S5) Функция $\tilde{\kappa}'(x)$ непрерывна.

Теорема 2.1. Пусть выполняются условия $S1-S5$. Тогда интегральный оператор с ядром (2.21) компактен на банаховом пространстве $D([\tau_0, \tau_\infty], \|\cdot\|_\infty)$ càdlàg функций на интервале $[\tau_0, \tau_\infty]$.

Более того, однородное интегральное уравнение

$$\varphi(x) - \int_{\tau_0}^{\tau_\infty} K_F(x, t)\varphi(t) dt = 0$$

имеет только тривиальное решение, и, следовательно, решение уравнение (2.20) существует и притом единственно.

Доказательство. См. приложение А. ■

Замечание. Условие $S2$ можно ослабить (и тогда условие $S1$ уже не будет следствием из условия $S2$): достаточно для любого x из интервала $[\tau_0, \tau_\infty]$ требовать существования такой последовательности x_j , что

$$\int_{\tau_0}^{\tau_\infty} |K(x, t) - K(x_j, t)| dt \rightarrow 0, \quad \text{при } x_j \rightarrow x.$$

В таком случае в качестве функции $\bar{h}(u, v)$ можно брать функцию, равную нулю, при $0 < v - u < \varepsilon$, и константе иначе.

Не всякое решение уравнения (2.20) абсолютно-непрерывно относительно функции F . Для того, чтобы нашлась такая функция $g(t) \in L_0^2(F)$, что

$$\varphi(x) = \int_x^M g(t) dF(t),$$

необходимо потребовать выполнения следующих условий:

- L1) Первые частные производные $\frac{\partial \bar{h}}{\partial u}(u, v)$ и $\frac{\partial \bar{h}}{\partial v}(u, v)$ существуют, кроме, быть может, конечного числа точек, где существуют левые и правые производные. Частные производные должны быть равномерно ограничены по (u, v) .

L2) Функция $\tilde{\kappa}'_F$ всюду дифференцируема, кроме, быть может, конечного числа точек, где существуют левые и правые производные. Производная должна быть ограничена равномерно по x .

Теорема 2.2. Пусть выполняются условия S1-S5 и L1-L2. Кроме этого, пусть плотность f отделена от 0: $f > C > 0$. Тогда информационное уравнение

$$\tilde{\kappa}_F(x) = \left[(A^{(1)})^* A^{(1)} \right] g(x) \quad F - n.v.. \quad (2.25)$$

с операторами $(A^{(1)})^*$ и $A^{(1)}$, определенными в (2.12), (2.13), имеет и притом единственное решение.

Доказательство. Смотри приложение А. ■

Из формулы (2.16) получаем, что канонический градиент $\tilde{\xi}_F$ при этом удовлетворяет соотношению

$$\tilde{\xi}_F(\delta_k, t_k, k) = -\frac{\delta_{k,1}\varphi_*(t_{k,1})}{F(t_{k,1})} - \sum_{i=2}^k \frac{\delta_{k,i}(\varphi_*(t_{k,i}) - \varphi_*(t_{k,i-1}))}{F(t_{k,i}) - F(t_{k,i-1})} + \frac{\delta_{k,k+1}\varphi_*(t_{k,k})}{1 - F(t_{k,k})}, \quad (2.26)$$

где

$$\varphi_*(x) = \int_x^M g_*(t) dF(t),$$

а $g_*(t)$ — решение уравнения (2.25).

2.2.3. ОМП функционалов в случае цензурирования

Итак, в случае интервального цензурирования смешанного типа интересующий функционал $K(F_0) = \int c(x) dF_0(x)$ задан неявно в виде $\Xi(Q)$. В качестве оценки $\Xi(Q)$ будем использовать величину $\Xi(Q_n) = \int c(x) d\hat{F}_n(x)$, где Q_n — эмпирическая мера, индуцированная наблюдениями Y_1, \dots, Y_n , а \hat{F}_n — некоторая оценка функции распределения F_0 , построенная по Y_1, \dots, Y_n .

В качестве оценки \hat{F}_n будем рассматривать оценку максимального правдоподобия для F_0 , а именно

$$\hat{F}_n = \operatorname{argmax}_{F \in \mathcal{F}} \sum_{i=1}^n \log q_{F,H} \left(\Delta_{K^{(i)}}, T_{K^{(i)}}, K^{(i)} \right), \quad (2.27)$$

где плотность $q_{F,H}$ задана формулой (2.11):

$$q_{F,H}(\delta_k, t_k, k) = p_k h_k(t_k) \prod_{i=1}^{k+1} [F(t_{k,i}) - F(t_{k,i-1})]^{\delta_{k,i}}.$$

Несложно видеть, что распределение цензурирования (K, T_K) не влияет на точку максимума, и мы приходим к выражению, аналогичному (2.4):

$$\hat{F}_n = \operatorname{argmax}_{F \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{K^{(i)}+1} \Delta_{K,j}^{(i)} \log \left[F(T_{K,j}^{(i)}) - F(T_{K,j-1}^{(i)}) \right]$$

Нашим основным инструментом для доказательства асимптотических свойств оценок (2.3) будет следующая теорема.

Теорема 2.3. Пусть функция распределения F_0 имеет плотность f отделенную от нуля. Пусть $\mathbf{E}(K) < \infty$ и выполняются условия S1-S5 и L1-L2.

Тогда для ОМП \hat{F}_n имеет место соотношение:

$$\sqrt{n} \left(\int c d\hat{F}_n - \int c dF_0 \right) = \sqrt{n} \left(\int \tilde{\xi}_F dQ_n - \int \tilde{\xi}_F dQ \right) + o_p(1), \text{ при } n \rightarrow \infty.$$

Доказательство. См. приложение Б. ■

2.3. Состоятельность оценок

Для установления асимптотических свойств оценок $\tilde{\theta}_n$, определенных в (2.3) нам потребуется применять теорему 2.3 к функционалам вида

$$F(\cdot) \mapsto \int \log p_\theta(\cdot) dF(\cdot). \quad (2.28)$$

Перепишем условия S5 и L2 на канонический градиент $\tilde{\kappa}_F$ функционала (2.28) в терминах функции $\log p_\theta$. Известно (см. [9], пример 3.3.2 и предложение A.5.2), что функционал вида (2.28) дифференцируем в точке F_0 как только

$$\int (\log p_\theta)^2 dF_0 < \infty$$

и при этом канонический градиент $\tilde{\kappa}_{F_0}$ равен

$$\tilde{\kappa}_{F_0}(x) = \log p_\theta(x) - \int \log p_\theta dF_0.$$

таким образом, условие S5 о непрерывности производной превращается в условие непрерывности функции

$$l_\theta(x) = \frac{p'_\theta(x)}{p_\theta(x)},$$

а условие L2 — в требование дифференцируемости функции $l_\theta(x)$ по x для всякого θ .

Определим величину θ^* :

$$\theta^* = \operatorname{argmin}_{\theta \in \Theta} \int \log \frac{f_0}{f_\theta} dF_0.$$

Теперь КЛ-состоятельность оценок $\tilde{\theta}_n$ можно получить при помощи достаточно несложной модификации теоремы 1.3.

Теорема 2.4. Пусть условия теоремы 2.3 выполняются как для функции $\log p_\theta(x)$, так и для функции $\sup_{\theta \in B} \log p_\theta(x)$ для всякого достаточно малого шара $B \subset \Theta$. Кроме этого, потребуем выполнения следующих условий:

1. Пространство параметров Θ — компакт.
2. Функция $\theta \mapsto \log p_\theta$ полунепрерывна сверху для F_0 -н.в. x .
3. Для достаточно малого шара $B \subset \Theta$ и всякого $\theta \in \Theta$ выполняется

$$\int \sup_{\theta \in B} \log p_\theta(x) dF(x) < \infty$$

4. Для всякой величины $\theta \in \Theta, \theta \neq \theta^*$ имеет место:

$$\int \log \frac{f_{\theta^*}}{f_{\theta}} dF_0 > 0.$$

Тогда оценка $\tilde{\theta}_n$, определенная как

$$\tilde{\theta}_n = \operatorname{argmax}_{\theta \in \Theta} \int f_{\theta}(x) d\hat{F}_n(x).$$

КЛ-состоятельна:

$$\tilde{\theta}_n \rightarrow \theta^*, \quad \text{н.н. } n \rightarrow \infty.$$

Доказательство. Как и в теореме 1.3 зафиксируем $\theta' \in \Theta$ и обозначим через $\{B_r(\theta')\}$ убывающую последовательность открытых шаров, содержащих θ' :

$$\bigcap_{r=1}^{\infty} B_r(\theta') = \{\theta'\}.$$

Из условия 2 следует, что

$$\sup_{B_r(\theta')} \log p_{\theta} \downarrow \log p_{\theta'}, \quad \text{при } r \rightarrow \infty.$$

Тогда условие 3 и теорема о монотонной сходимости гарантируют, что

$$\int \sup_{B_r(\theta')} \log p_{\theta} dF_0 \rightarrow \int \log p_{\theta'} dF_0, \quad \text{при } r \rightarrow \infty. \quad (2.29)$$

Рассмотрим окрестность $B(\theta^*)$ точки θ^* . Определим множество $\bar{B}(\theta^*) = \Theta \setminus B(\theta^*)$. Для $\theta' \in \bar{B}(\theta^*)$ условие идентифицируемости 4 гарантирует, что

$$\int \log p_{\theta'} dF_0 < \int \log p_{\theta^*} dF_0.$$

Таким образом, объединяя с (2.29), имеем

$$\int \sup_{B_r(\theta')} \log p_{\theta} dF_0 \rightarrow \int \log p_{\theta'} dF_0 < \int \log p_{\theta^*} dF_0, \quad \text{при } r \rightarrow \infty.$$

Следовательно, для всякого $\theta' \in \bar{B}(\theta^*)$ существует окрестность $U(\theta')$, что

$$\int \sup_{U(\theta')} \log p_\theta dF_0 < \int \log p_{\theta^*} dF_0.$$

Применяя теорему 2.3, отсюда имеем

$$\begin{aligned} \sup_{U(\theta')} \int \log p_\theta d\hat{F}_n &\leq \int \sup_{U(\theta')} \log p_\theta d\hat{F}_n \rightarrow \\ &\rightarrow \int \sup_{U(\theta')} \log p_\theta dF_0 < \int \log p_{\theta^*} dF_0, \quad \text{п.н. при } n \rightarrow \infty. \end{aligned} \quad (2.30)$$

Компактность множества $\bar{B}(\theta^*)$ позволяет переписать формулу (2.30) в виде

$$\limsup_{n \rightarrow \infty} \sup_{\bar{B}(\theta_0)} \int \log p_\theta d\hat{F}_n < \int \log p_{\theta^*} dF_0, \quad \text{п.н.}$$

Отсюда

$$\limsup_{n \rightarrow \infty} \left(\sup_{\bar{B}(\theta^*)} \int \log p_\theta d\hat{F}_n - \int \log p_{\theta^*} dF_0 \right) < 0, \quad \text{п.н.}$$

а так как из теоремы 2.3 следует, что $\int \log p_{\theta^*} d\hat{F}_n - \int \log p_{\theta^*} dF_0 \rightarrow 0$, то

$$\limsup_{n \rightarrow \infty} \left(\sup_{\bar{B}(\theta^*)} \int \log p_\theta d\hat{F}_n - \int \log p_{\theta^*} d\hat{F}_n \right) < 0, \quad \text{п.н.}$$

Определение (2.3) тогда гарантирует, что

$$\mathbf{P} \left(\tilde{\theta}_n \in B(\theta^*) \text{ для почти всех } n \right) = 1,$$

из чего, в силу произвольности окрестности $B(\theta^*)$, следует, что

$$\tilde{\theta}_n \rightarrow \theta^*, \quad \text{п.н. при } n \rightarrow \infty.$$

■

Следствие 2.4.1. *Предположим, что параметрическая модель для F_0 точна, то есть найдется $\theta_0 \in \Theta$, что $f_0 = f_{\theta_0}$. Тогда:*

1. $\theta^* = \theta_0$.

2. Условие идентифицируемости 4 совпадет с условием идентифицируемости для обычных оценок максимального правдоподобия для полных данных, т.е. $p_{\theta_1}(x) \neq p_{\theta_2}(x)$ при $\theta_1 \neq \theta_2$ для F_0 -п.в. x .

2.4. Асимптотическое распределение оценок

Теорема 2.3 предоставляет возможность перенесения стандартных результатов об асимптотическом распределении оценок максимального правдоподобия на случай ОМКЛ $\tilde{\theta}_n$. В данном разделе это будет сначала продемонстрировано на условиях типа Крамера (по аналогии с разделом 1.4.1). После этого, теорема 2.3 будет использована для распространения результатов раздела 1.4.2 на случай ОМКЛ.

2.4.1. Условия типа Крамера

Как и в разделе 1.4.1 предположим, что $\theta \in \Theta \subset \mathbb{R}$. Предположим, что функция $\theta \mapsto \log p_\theta$ дифференцируема всюду и обозначим $\psi_\theta = \frac{\partial}{\partial \theta} \log p_\theta$.

Теорема 2.5. *Потребуем выполнения следующих условий:*

1. Пусть отображение $\theta \mapsto \psi_\theta(x)$ дважды непрерывно дифференцируемо для F_0 -почти всех x .
2. Предположим, что найдется такая суммируемая (по распределению F_0) функция $M(x)$, что в некоторой окрестности точки θ^* :

$$\left| \frac{\partial^2}{\partial \theta^2} \psi_\theta(x) \right| \leq M(x).$$

3. $\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 \neq 0$ и условия теоремы 2.3 выполняются для функций $\psi_{\theta^*}(x)$, $\frac{\partial \psi_{\theta^*}}{\partial \theta}(x)$ и $M(x)$ (или же для $\sup_{\theta \in B} \frac{\partial^2 \psi_\theta}{\partial \theta^2}(x)$ для некоторой окрестности B точки θ^*)

Тогда для состоятельной ОМКЛ $\tilde{\theta}_n$ выполняется:

$$\sqrt{n} \left(\tilde{\theta}_n - \theta^* \right) \xrightarrow{d} N \left(0, \frac{\int \tilde{\xi}_{\theta^*}^2 dQ}{\left(\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 \right)^2} \right), \text{ при } n \rightarrow \infty,$$

где $\tilde{\xi}_{\theta^*}$ — канонический градиент, соответствующий линейному оператору с функцией $s(x) = \psi_{\theta^*}$.

Доказательство. Разложим функцию $\int \psi_{\tilde{\theta}_n} d\hat{F}_n$ в окрестности точки θ^* по формуле Тейлора (перестановка местами операций интегрирования и дифференцирования возможна, так как функция \hat{F}_n в силу леммы 2.10 может быть выбрана кусочно-постоянной):

$$\int \psi_{\tilde{\theta}_n} d\hat{F}_n = \int \psi_{\theta^*} d\hat{F}_n + \left(\tilde{\theta}_n - \theta^* \right) \int \frac{\partial}{\partial \theta} \psi_{\theta^*} d\hat{F}_n + \frac{1}{2} \left(\tilde{\theta}_n - \theta^* \right)^2 \int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\hat{F}_n, \quad (2.31)$$

здесь величина $\tilde{\theta}_n$ расположена на сегменте прямой, соединяющей θ^* и $\tilde{\theta}_n$.

Из условия 3 и теоремы 2.3 следует, что

$$\int \frac{\partial}{\partial \theta} \psi_{\theta^*} d\hat{F}_n \xrightarrow{\mathbf{P}} \int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0, \text{ при } n \rightarrow \infty.$$

Состоятельность оценки $\tilde{\theta}_n$, условие 2 и теорема 2.3, в свою очередь, гарантируют ограниченность величины $\int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\hat{F}_n$ по вероятности. С учетом всего этого из формулы (2.31) получаем:

$$\begin{aligned} \int \psi_{\tilde{\theta}_n} d\hat{F}_n - \int \psi_{\theta^*} d\hat{F}_n &= \\ &= \left(\tilde{\theta}_n - \theta^* \right) \int \frac{\partial}{\partial \theta} \psi_{\theta^*} d\hat{F}_n + \frac{1}{2} \left(\tilde{\theta}_n - \theta^* \right)^2 \int \frac{\partial^2}{\partial \theta^2} \psi_{\tilde{\theta}_n} d\hat{F}_n = \\ &= \left(\tilde{\theta}_n - \theta^* \right) \left(\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 + o_P(1) + \left(\tilde{\theta}_n - \theta^* \right) O_p(1) \right), \text{ при } n \rightarrow \infty. \end{aligned}$$

Состоятельность оценки $\tilde{\theta}_n$ влечет за собой $\left(\tilde{\theta}_n - \theta^* \right) O_p(1) = o_p(1)$. Из условия 3 следует, что при достаточно больших n величина $\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 + o_P(1)$

отлична от нуля. Отсюда, учитывая, что $\int \psi_{\tilde{\theta}_n} d\hat{F}_n = 0$, выводим

$$\begin{aligned} \sqrt{n} (\tilde{\theta}_n - \theta^*) &= \sqrt{n} \left(\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 + o_P(1) \right)^{-1} \left(\int \psi_{\tilde{\theta}_n} d\hat{F}_n - \int \psi_{\theta^*} d\hat{F}_n \right) = \\ &= -\sqrt{n} \left(\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 + o_P(1) \right)^{-1} \int \psi_{\theta^*} d\hat{F}_n \quad (2.32) \end{aligned}$$

Так как точка θ^* является точкой максимума отображения $\theta \mapsto \log p_\theta$, то, очевидно, $\int \psi_{\theta^*} dF_0 = 0$. В таком случае из условия 2 следует, что случайная величина $\sqrt{n} \int \psi_{\theta^*} d\hat{F}_n$ сходится по распределению к случайной величине, имеющей нормальное распределение со средним 0 и дисперсией $\int \tilde{\xi}_{\theta^*}^2 dQ$, и, следовательно,

$$\sqrt{n} (\tilde{\theta}_n - \theta^*) \xrightarrow{d} N \left(0, \frac{\int \tilde{\xi}_{\theta^*}^2 dQ}{\left(\int \frac{\partial}{\partial \theta} \psi_{\theta^*} dF_0 \right)^2} \right), \text{ при } n \rightarrow \infty.$$

■

2.4.2. Слабые условия асимптотической нормальности

Как было отмечено выше, теорема 2.3 позволяет перенести результаты раздела 1.4.2 на случай ОМКЛ $\tilde{\theta}_n$. В частности, теорему E1 оказывается возможным вместо процесса $c(\cdot) \mapsto \sqrt{n} \left(\int c(x) d\hat{F}_n - \int c(x) dF_0 \right)$ применить к процессу $\mathbb{G}_n = \sqrt{n} (\mathbb{Q}_n - Q)$, тем самым получив результат, аналогичный теореме 1.5.

Для начала нам потребуется вспомогательный результат относительно дифференцируемости канонического градиента $\tilde{\xi}_\theta$ по параметру θ .

Лемма 2.6. *Обозначим через $\tilde{\kappa}_\theta$ канонический градиент функционала*

$$F(x) \mapsto \int c_\theta(x) dF(x)$$

в точке F_0 . Пусть выполняются условия S1-S5. Предположим, что функция $\tilde{\kappa}'_\theta$ дифференцируема по параметру θ в окрестности точки θ_0 и производная $\frac{\partial \tilde{\kappa}'_\theta}{\partial \theta}(x)$ непрерывна для F_0 -п.в. x . Более того, предположим, что в

окрестности точки θ_0 производная $\frac{\partial \tilde{\kappa}'_\theta}{\partial \theta}(x)$ равномерно ограничена по θ для F_0 -п.в. x .

Тогда канонический градиент $\tilde{\xi}_\theta$ дифференцируем по θ в окрестности точки θ_0 .

Доказательство. Из формулы (2.26) следует, что дифференцируемость по параметру θ канонического градиента $\tilde{\xi}_\theta$ равносильно дифференцируемости по параметру θ решения φ_θ информационного уравнения (2.19).

Пусть $R(x, t)$ — резольвента интегрального оператора с ядром (2.21). Тогда, из теоремы 2.1 следует, что функция $\varphi_\theta(x)$ допускает представление:

$$\varphi_\theta(x) = a_\theta(x) - \int_0^S R(x, t) a_\theta(t) dt,$$

где функция $a_\theta(x)$ определена формулой (2.23):

$$a_\theta(x) = \frac{\tilde{\kappa}'_\theta(x) \bar{d}_{F_0}(x)}{1 + \bar{d}_{F_0}(x) \int_0^S \bar{D}_{F_0}(x, t) dt}.$$

Результат леммы теперь следует из непрерывности интегрального оператора с ядром $R(x, t)$. ■

Замечание. *Всюду далее в качестве функции $s_\theta(x)$ будет выступать логарифм плотности $\log p_\theta(x)$. В связи с этим, условия леммы фактически определяют поведение функции $l_\theta(x) = \frac{p'_\theta(x)}{p_\theta(x)}$ и ее частных производных по θ в окрестности точки θ_0 .*

Лемма 2.6 позволяет применить теорему E1 для установления скорости сходимости оценок параметров $\tilde{\theta}_n$: для этого достаточно всюду в условиях теоремы 1.5 заменить функцию m_θ на $\tilde{\xi}_\theta$, где $\tilde{\xi}_\theta$ — канонический градиент функционала

$$F(x) \mapsto \int \log p_\theta(x) dF(x)$$

в точке F_0 . Другими словами, всюду, где требуется оценить модуль непрерывности процесса

$$\theta \mapsto \sqrt{n} \int \log p_\theta d(\hat{F}_n - F_0),$$

производится замена на эмпирический процесс вида

$$\theta \mapsto \sqrt{n} \int \tilde{\xi}_\theta d(Q_n - Q).$$

Таким образом, получаем следующую теорему

Теорема 2.7. Пусть функция $\theta \mapsto \int \log p_\theta dF_0$ допускает разложение в ряд Тейлора до второго члена. Предположим, что пространство параметров Θ является подмножеством евклидова пространства \mathbb{R}^p . Пусть для функционала $F(x) \mapsto \int \log p_\theta(x) dF(x)$ выполняются условия теоремы 2.3. Предположим, что функция $l_\theta(x) = \frac{p'_\theta(x)}{p_\theta(x)}$ дифференцируема по параметру θ в окрестности точки θ^* и производная $\frac{\partial}{\partial \theta} \left(\frac{p'_\theta(x)}{p_\theta(x)} \right)$:

1. Непрерывна по x для любого θ из некоторой окрестности θ^* F_0 -п.в.
2. Непрерывна по θ в окрестности θ^* для F_0 -п.в. x .
3. Равномерно ограничена по θ в окрестности точки θ^* .

$$\text{Тогда } \sqrt{n} (\tilde{\theta}_n - \theta_0) = O_p(1).$$

Доказательство. Следует из леммы 2.6 и теоремы 1.5. ■

Аналогичным образом устанавливается асимптотическая нормальность. Так же, как и в формуле (1.30) имеем:

$$\begin{aligned} n \int (\log p_{\tilde{\theta}_n} - \log p_{\theta^*}) d\hat{F}_n &= \\ &= n \int (\log p_{\tilde{\theta}_n} - \log p_{\theta^*}) dF_0 + n \int (\log p_{\tilde{\theta}_n} - \log p_{\theta^*}) d(\hat{F}_n - F_0). \end{aligned} \quad (2.33)$$

Пусть функция $\theta \mapsto \int \log p_\theta dF_0$ допускает разложение в ряд Тейлора в окрестности точки θ^* до второго члена с невырожденной матрицей вторых частных производных Σ^* . Так как точка θ^* является точкой максимума функции $\theta \mapsto \int \log p_\theta dF_0$, то из формулы (2.33) и теоремы 2.3 получаем

$$\begin{aligned} n \int (\log p_{\tilde{\theta}_n} - \log p_{\theta^*}) d\hat{F}_n &= \frac{1}{2} \sqrt{n} (\tilde{\theta}_n - \theta^*)^T \Sigma^* \sqrt{n} (\tilde{\theta}_n - \theta^*) + \\ &+ \sqrt{n} \int (\log p_{\tilde{\theta}_n} - \log p_{\theta^*}) d(\hat{F}_n - F_0) + o_P(n \|\tilde{\theta}_n - \theta_0\|^2) = \\ &= \frac{1}{2} \sqrt{n} (\tilde{\theta}_n - \theta^*)^T \Sigma^* \sqrt{n} (\tilde{\theta}_n - \theta^*) + \sqrt{n} \int (\tilde{\xi}_{\tilde{\theta}_n} - \tilde{\xi}_{\theta^*}) d(\mathbb{Q}_n - Q) + \\ &+ o_P(n \|\tilde{\theta}_n - \theta_0\|^2 + \sqrt{n} |\log p_{\tilde{\theta}_n} - \log p_{\theta^*}|), \quad n \rightarrow \infty. \end{aligned} \quad (2.34)$$

Заметим, что в силу леммы 2.6 условие стохастической дифференцируемости (1.32) применимо к функции

$$\theta \mapsto \int (\tilde{\xi}_\theta - \tilde{\xi}_{\theta^*}) d(\mathbb{Q}_n - Q).$$

Тогда, аналогично формулам (1.33), (1.34), (1.35) получаем, что

$$\left\| \sqrt{n} (\tilde{\theta}_n - \theta^*) + (\Sigma^*)^{-1} \sqrt{n} \int \tilde{\xi}'_{\theta^*} d(\mathbb{Q}_n - Q) \right\| \xrightarrow{\mathbf{P}} 0, \quad n \rightarrow \infty.$$

Применяя центральную предельную теорему к величине

$$(\Sigma^*)^{-1} \sqrt{n} \int \tilde{\xi}'_{\theta^*} d(\mathbb{Q}_n - Q),$$

получаем

$$(\Sigma^*)^{-1} \sqrt{n} \left(\int \tilde{\xi}'_{\theta^*} d\mathbb{Q}_n - \int \tilde{\xi}'_{\theta^*} dQ \right) \xrightarrow{d} N \left(0, (\Sigma^*)^{-1} \int \tilde{\xi}'_{\theta^*} (\tilde{\xi}'_{\theta^*})^T dQ (\Sigma^*)^{-1} \right),$$

следовательно, случайная величина $\sqrt{n} (\tilde{\theta}_n - \theta_0)$ асимптотически имеет нормальное распределение со средним 0 и ковариационной матрицей $\tilde{\Sigma}$, где

$$\tilde{\Sigma} = (\Sigma^*)^{-1} \int \tilde{\xi}'_{\theta^*} (\tilde{\xi}'_{\theta^*})^T dQ (\Sigma^*)^{-1}.$$

Объединяя все условия, получаем следующую теорему об асимптотическом распределении ОМКЛ $\tilde{\theta}_n$:

Теорема 2.8. Пусть выполняются следующие условия:

1. Пространство параметров Θ является компактным подмножеством Евклидова пространства \mathbb{R}^p . Величина θ^* принадлежит внутренности Θ .
2. Выполняются условия теоремы 2.4.
3. Функция $\theta \mapsto \int \log p_\theta dF_0$ допускает разложение в по Тейлору в окрестности точки θ^* до второго члена с невырожденной матрицей вторых производных Σ^* .
4. Предположим, что функция $l_\theta(x) = \frac{p'_\theta(x)}{p_\theta(x)}$ дифференцируема по параметру θ в окрестности точки θ^* и производная $\frac{\partial}{\partial \theta} \left(\frac{p'_\theta(x)}{p_\theta(x)} \right)$:
 - 1) Непрерывна по переменной x для любого θ из некоторой окрестности θ^* F_0 -п.в.
 - 2) Непрерывна по переменной θ в окрестности θ^* для F_0 -п.в. x .
 - 3) Равномерно ограничена по переменной θ в окрестности точки θ^* .

Тогда ОМКЛ $\tilde{\theta}_n$ строго состоятельна и асимптотически нормальна:

$$\sqrt{n} \left(\tilde{\theta}_n - \theta^* \right) \xrightarrow{d} N \left(0, (\Sigma^*)^{-1} \int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ (\Sigma^*)^{-1} \right).$$

Следствие 2.8.1. Предположим, что параметрическая модель для F_0 точна, то есть найдется $\theta_0 \in \Theta$, что $f_0 = f_{\theta_0}$. Тогда в условиях теоремы:

1. $\theta^* = \theta_0$.
2. Матрица Σ^* совпадает с информационной матрицей Фишера:

$$\Sigma^* = - \int \frac{\partial^2}{\partial \theta^2} \log p_{\theta_0}(x) dF_0(x).$$

2.5. Вычисление непараметрической оценки функции распределения.

Для нахождения оценки \hat{F}_n непосредственно из уравнения (2.4) необходимо максимизировать логарифм функции правдоподобия $\int m_F dQ_n$ по функциональному пространству \mathcal{F} , тем самым, задача оптимизации (2.4) является бесконечномерной.

Следуя [64], сведем сложную бесконечномерную задачу к гораздо более простой конечномерной задаче. Кроме того, одновременно получим условия существования решения и выделим класс функций распределения, в котором его следует искать.

2.5.1. Редукция.

Определение 12. *Интервал H назовем внутренним подинтервалом (в англоязычной литературе иногда используется термин «максимальное пересечение» — maximal intersection) для набора интервалов $\{I_j\}_{j=1}^n$, если выполняются следующие условия:*

1. *H представим в виде пересечения интервалов $\bigcap_{j \in \alpha} I_j$ для некоторого набора индексов $\alpha \subseteq \{1, \dots, n\}$.*
2. *Для всякого номера j либо $H \cap I_j = H$, либо $H \cap I_j = \emptyset$.*

Заметим, что если H, H' — два различных внутренних интервала, то $H \cap H' = \emptyset$.

В работе [17] рассматривается теоретико-графовый подход к построению набора всех внутренних подинтервалов для данного набора интервалов $\{I_j\}_{j=1}^n$. Показано, что внутренние подинтервалы соответствуют максимальным кликам графа пересечений для данного набора интервалов.

Поведение оценки максимального правдоподобия \hat{F}_n описывается леммами из [17].

Лемма 2.9. Пусть $\mathcal{H} = \{H_1, \dots, H_m\}$ — набор всех внутренних подинтервалов для $\{I_j\}_{j=1}^n$. Тогда любая функция распределения из \mathcal{F} , возрастающая вне \mathcal{H} , не может быть оценкой максимального правдоподобия для функции распределения F .

Лемма 2.10. При фиксированных значениях $\mathbf{P}_F(H_i)$ функция правдоподобия (1.6) не зависит от поведения функции распределения F внутри каждого интервала H_i .

Эти леммы имеют очень важный смысл, с точки зрения процедуры вычисления оценки \hat{F}_n . Лемма 2.9 позволяет свести сложную бесконечномерную оптимизационную задачу (2.4) по всему пространству \mathcal{F} к аналогичной, но уже по некоторому подпространству \mathcal{F} . Лемма 2.10, с одной стороны, фактически означает, что оценкой максимального правдоподобия для F является целый класс эквивалентности функций распределения и нет никакой дополнительной информации для того, чтобы выбрать конкретного представителя из этого класса, с другой — позволяет при построении оценки \hat{F}_n ограничиться поиском исключительно в классе скачкообразных на \mathcal{H} функций и тем самым получить конечномерную оптимизационную задачу.

Действительно, имеется ровно m интервалов, где функция \hat{F}_n может производить скачки, потому для ее определения достаточно найти величину этих скачков.

Для $i = 1, \dots, m$ положим $p_i = \mathbf{P}_F(H_i)$ и $p = (p_1, \dots, p_m)$. Тогда для $j = 1, \dots, n$ имеем

$$\mathbf{P}_F(I_j) = \sum_{i=1}^m p_i 1_{\{H_i \subseteq I_j\}} = \sum_{i=1}^m p_i \alpha_{ij}.$$

После этого можно переписать выражение для логарифма функции правдоподобия (1.6) в терминах величин p_i и α_{ij} :

$$l_n(F) = l_n(p_1, \dots, p_m) = \sum_{j=1}^n \log \mathbf{P}_F(I_j) = \sum_{j=1}^n \log \left[\sum_{i=1}^m p_i \alpha_{ij} \right].$$

Таким образом, вычисление оценки максимального правдоподобия \hat{F}_n можно свести к следующей двухшаговой процедуре:

1. *Редукция.* Вычисление набора \mathcal{H} всех внутренних подинтервалов. Это можно сделать за $O(n \log n)$ операций и $O(n)$ памяти [43].
2. *Оптимизация.* Решение конечномерной оптимизационной задачи

$$\hat{p} = \operatorname{argmax}_{\mathcal{P}} l_n(p_1, \dots, p_m) = \operatorname{argmax}_{p \in \mathcal{P}} l_n(p), \quad (2.35)$$

здесь

$$\mathcal{P} = \left\{ p \in \mathbb{R}^m : p_i \geq 0, i = 1, \dots, m; \sum_{i=1}^m p_i = 1 \right\}.$$

Таким образом, получаем m -мерную выпуклую задачу оптимизации.

После шага редукции несложно доказать теорему о существовании решения оптимизационной задачи (построению решения посвящён раздел 2.5.2).

Теорема 2.11. *Величина \hat{p} , определенная уравнением (2.35), существует.*

Доказательство. Полагая $\log(0) = -\infty$, продолжим логарифм по непрерывности на расширенную числовую прямую $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty\}$. Тогда логарифм функции правдоподобия $l_n(p)$ является непрерывной функцией, заданной на непустом компакте $\mathcal{P} \subset \mathbb{R}^m$ и принимающей значения в $\bar{\mathbb{R}}$. Тем самым существует максимум $l_n(p)$ на \mathcal{P} , и величина \hat{p} , являющаяся точкой максимума $l_n(p)$, в действительности существует. ■

2.5.2. Оптимизация. EM-алгоритм.

Для решения задачи оптимизации (2.35) будем применять EM-алгоритм [11]. В силу лемм 2.9, 2.10 «полные данные» представляют собой знание, какому интервалу H_i принадлежит наблюдение с номером j . Это, естественно, не эквивалентно знанию точного времени отказа X_j : в некотором смысле механизм цензурирования диктует выбор информативной модели для полных данных.

Обозначим $\beta_{ij} = 1_{\{X_j \in H_i\}}$ и $n_i = \sum_{j=1}^n \beta_{ij}$. Тогда естественной моделью для полных данных является мультиномиальная модель на \mathcal{H} с вектором вероятностей (p_1, \dots, p_m) . Запишем логарифм (ненаблюдаемой) функции правдоподобия для полных данных $l_n^c(p_1, \dots, p_m)$:

$$l_n^c(p) = l_n^c(p_1, \dots, p_m) = \sum_{i=1}^m n_i \log p_i, \quad p = (p_1, \dots, p_m) \in \mathcal{P}. \quad (2.36)$$

Общая схема EM-алгоритма.

EM-алгоритм решает задачу максимизации логарифма функции правдоподобия для неполных данных (2.35) неявно, применяя для этого логарифм функции правдоподобия для полных данных (2.36). Так как величины n_i не наблюдаются, следовательно, не наблюдается и $l_n^c(p)$, то используется условное математическое ожидание $l_n^c(p)$ по отношению к наблюдаемым случайным величинам α_{ij} . Приведем $(k + 1)$ -тую итерацию EM-алгоритма:

E-шаг. Вычислить функцию $Q(p; p^{(k)})$, где

$$Q(p; p^{(k)}) = \mathbf{E}_{p^{(k)}}(l_n^c(p) | \mathcal{H}).$$

M-шаг. Выбрать в качестве следующего приближения $p^{(k+1)}$ точку максимума функции $Q(p; p^{(k)})$:

$$Q(p^{(k+1)}; p^{(k)}) \geq Q(p; p^{(k)}), \quad \forall p \in \mathcal{P}.$$

Всюду далее символом $\mathbf{E}_{p^{(k)}}(\cdot)$ будет обозначаться математическое ожидание с использованием вектора параметров $p^{(k)}$.

Е- и М-шаги алгоритма последовательно выполняются до тех пор, пока последовательность $p^{(k)}$ не достигнет стационарной точки.

Теорема 2.12. *Применение процедуры EM-алгоритма приводит к увеличению значения функции правдоподобия, т.е.*

$$l_n(p^{(k+1)}) \geq l_n(p^{(k)}), \quad k = 0, 1, 2, \dots$$

Доказательство. Следует из общей теории EM-алгоритма [11, 45]. ■

EM-алгоритм для оценивания \hat{F}_n .

Для того чтобы применить EM-алгоритм к максимизации функции $l_n^c(p)$, необходимо получить выражение для функции $Q(p; p^{(k)})$.

Заметим, что $l_n^c(p)$ линейна по n_i , которые, в свою очередь, являются простой суммой β_{ij} . Из определения β_{ij} непосредственно имеем

$$\mathbf{E}_{p^{(k)}}(\beta_{ij} | \mathcal{H}) = \mathbf{E}_{p^{(k)}}(1_{\{X_j \in H_i\}} | \mathcal{H}) = \mathbf{P}(X_j \in H_i | X_j \in I_j) = \frac{\alpha_{ij} p_i^{(k)}}{\sum_{i=1}^m \alpha_{ij} p_i^{(k)}}.$$

Отсюда, суммируя по $j = 1, \dots, n$, получаем

$$\mu_i(p^{(k)}) = \mathbf{E}_{p^{(k)}}(n_i | \mathcal{H}) = \sum_{j=1}^n \mathbf{E}_{p^{(k)}}(\beta_{ij} | \mathcal{H}) = \sum_{j=1}^n \left[\frac{\alpha_{ij} p_i^{(k)}}{\sum_{i=1}^m \alpha_{ij} p_i^{(k)}} \right].$$

После этих приготовлений несложно вывести явное выражение для функции $Q(p; p^{(k)})$

$$Q(p; p^{(k)}) = \mathbf{E}_{p^{(k)}}(l_n^c(p) | \mathcal{H}) = \sum_{i=1}^m \mu_i(p^{(k)}) \log p_i.$$

М-шаг EM-алгоритма заключается в максимизации функции $Q(p; p^{(k)})$ по пространству \mathcal{P} . Этот максимум несложно получить в явной форме. Рассмотрим функцию

$$f(p_1, \dots, p_n, \lambda) = \sum_{i=1}^m \mu_i(p^{(k)}) \log p_i - \lambda \sum_{i=1}^m p_i.$$

Приравняем частные производные $f(p_1, \dots, p_n, \lambda)$ по p_1, \dots, p_m к нулю:

$$\frac{\partial f}{\partial p_i} = \frac{\mu_i(p^{(k)})}{p_i} - \lambda = 0,$$

откуда

$$p_i = \frac{\mu_i(p^{(k)})}{\lambda}.$$

Суммируя p_i по $i = 1, \dots, m$, находим

$$1 = \sum_{i=1}^m p_i = \sum_{i=1}^m \frac{\mu_i(p^{(k)})}{\lambda},$$

что дает

$$\lambda = \sum_{i=1}^m \mu_i(p^{(k)}) = n.$$

Окончательно для пересчета p_i имеем явную формулу

$$p_i^{(k+1)} = \frac{\mu_i(p^{(k)})}{n} = \frac{1}{n} \sum_{j=1}^n \left[\frac{\alpha_{ij} p_i^{(k)}}{\sum_{i=1}^m \alpha_{ij} p_i^{(k)}} \right]. \quad (2.37)$$

Замечание. EM-алгоритм, вообще говоря, не всегда сходится к \hat{F}_n . В работе [16] приведены условия типа Куна–Таккера, позволяющие проверить, действительно ли вектор (p_1, \dots, p_m) является решением задачи (2.35).

EM-алгоритм был выбран из-за простоты описания и реализации, он хорошо работал в интересующих нас случаях и на практике всегда сходился к оценке максимального правдоподобия \hat{F}_n .

Существуют другие методы решения оптимизационной задачи (2.35). Среди наиболее часто используемых отметим метод ICM (Iterative Convex Minorant) [31].

Информационные критерии типа Акайке

Процедура выбора адекватной статистической модели встает при анализе практически любых данных. Как правило, мы не можем быть уверенным в том, что предполагаемая нами модель на самом деле *точна* для данных. Вместо этого мы можем лишь предполагать, что выбранная модель *адекватна*, то есть хорошо аппроксимирует интересующие нас характеристики распределения данных.

Процедура выбора статистической модели для данных обычно производится в два этапа:

1. Выбор предполагаемого семейства распределений модели.
2. Оценивание параметров распределения модели.

Естественно эти два этапа взаимосвязаны: выбор сложного и «богатого» семейства распределений на первом шаге потенциально может позволить аппроксимировать распределение данных с большой точностью. В то же время увеличение количества степеней свободы, как правило, приводит к существенному увеличению дисперсии оценок параметров и усложняет второй шаг. Напротив, упрощение модели позволяет получить более точные (в смысле дисперсии) оценки параметров, однако, такая модель может не вполне адекватно отражать требуемые характеристики имеющегося набора данных.

В предыдущих главах был рассмотрен вопрос оценивания параметров параметрической модели (в условиях неполных данных в виде интервального цензурирования смешанного типа), то есть фактически второй шаг в рассматриваемой процедуре построения статистической модели. Для выпол-

нения первого шага необходимо наличие некоторого критерия, по которому можно было бы сравнивать разные модели.

Одним из стандартных способов сравнения моделей является информационный критерий Акайке [1] и его модификации [35, 36]. Пусть, как и раньше, X_1, \dots, X_n — последовательность независимых одинаково распределенных случайных величин с функцией распределения F_0 . Допустим, что у F_0 существует плотность f_0 относительно некоторой доминирующей меры ν . Как и прежде обозначим через $I(g; f_0)$ расстояние Кульбака-Лейблера [40] между плотностями g и f_0 :

$$I(g; f_0) = \int \log \frac{f_0}{g} dF_0.$$

Известно, что $I(g; f_0) \geq 0$ и что $I(g; f_0) = 0$ тогда и только тогда, когда $g = f_0$ п.н. Отметим, что вообще говоря, сходимость распределений по расстоянию Кульбака-Лейблера эквивалентно сходимость соответствующих распределений по вариации. Таким образом, при сравнении двух моделей g_1 и g_2 и для выбора наилучшей достаточно выбрать ту, для которой расстояние $I(g_i; f_0)$ будет наименьшим.

Величина $\int f_0 dF_0$ не зависит от плотности g (и, вообще говоря, при анализе данных неизвестна, так как содержит неизвестное распределение F_0), поэтому вместо сравнения величин $I(g; f_0)$ достаточно рассматривать величины $-\int g dF_0$, так как, очевидно

$$I(g; f_0) = \int \log \frac{f_0}{g} dF_0 = \int f_0 dF_0 - \int g dF_0.$$

Как правило предполагается параметрическая модель для плотности g , то есть $g \in \mathcal{G} = \{g_\theta : \theta \in \Theta\}$ и рассматривается величина вида

$$I_0(g; f_0) = \int g_{\hat{\theta}_n} dF_0,$$

где $\hat{\theta}_n$ — некоторая оценка неизвестного параметра θ .

Таким образом, задача выбора адекватной параметрической модели формально становится задачей выбора параметрического семейства \mathcal{G} при оценке параметра с помощью оценки $\hat{\theta}_n$.

Функция распределения F_0 обычно неизвестна, поэтому на практике расстояние $I_0(g; f_0)$ оценивают, используя некоторую оценку \hat{F}_n для F_0 . Известно [1], что величина $\int g_{\hat{\theta}_n} d\hat{F}_n$ является смещенной оценкой для $\int g_{\hat{\theta}_n} dF_0$. Более того, это смещение зависит от параметрического класса \mathcal{G} . Поэтому, вообще говоря, сравнивать $\int g_{\hat{\theta}_n}^{(1)} d\hat{F}_n$ и $\int g_{\hat{\theta}_n}^{(2)} d\hat{F}_n$, где $g_{\theta}^{(1)} \in \mathcal{G}_1$, $g_{\theta}^{(2)} \in \mathcal{G}_2$ и $\mathcal{G}_1 \neq \mathcal{G}_2$, нельзя.

Информационный критерий Акайке определяется как асимптотически несмещенная оценка величины $\int g_{\hat{\theta}_n} dF_0$. Самый простой способ получить такую оценку — это оценить и скорректировать смещение $\int g_{\hat{\theta}_n} d\hat{F}_n$. Далее это будет сделано для ОМП $\hat{\theta}_n$ и ОМКЛ $\tilde{\theta}_n$.

3.1. Информационные критерии в случае ОМКЛ

Информационный критерий типа Акайке был введен выше как асимптотически несмещенная оценка величины $\tilde{I}_0 = \int \log f_{\tilde{\theta}_n} dF_0$. Так как функция распределения F_0 неизвестна, то в качестве оценки \tilde{I}_0 будем использовать $\hat{I}_n = \int \log f_{\tilde{\theta}_n} d\hat{F}_n$, где \hat{F}_n — непараметрическая оценка функции распределения, определенная формулой (2.4). Оценим смещение величины $D = \hat{I}_n - \tilde{I}_0$.

Имеем $D = D_1 + D_2 + D_3$, где

$$\begin{aligned} D_1 &= \int \log f_{\tilde{\theta}_n} d\hat{F}_n - \int \log f_{\theta^*} d\hat{F}_n, \\ D_2 &= \int \log f_{\theta^*} d\hat{F}_n - \int \log f_{\theta^*} dF_0, \\ D_3 &= \int \log f_{\theta^*} dF_0 - \int \log f_{\tilde{\theta}_n} dF_0. \end{aligned}$$

Оценим величины D_1 , D_2 , D_3 по отдельности.

Раскладывая функцию $\theta \mapsto \int \log f_\theta dF_0$ вблизи точки θ^* по формуле Тейлора, имеем:

$$\begin{aligned} \int \log f_{\tilde{\theta}_n} dF_0 &= \int \log f_{\theta^*} dF_0 + \left(\tilde{\theta}_n - \theta^*\right)^T \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right) \Big|_{\theta=\theta^*} dF_0 + \\ &+ \frac{1}{2} \left(\tilde{\theta}_n - \theta^*\right)^T \int \left(\frac{\partial^2}{\partial \theta^2} \log f_\theta \right) \Big|_{\theta=\theta^*} dF_0 \left(\tilde{\theta}_n - \theta^*\right) + o_P(\|\tilde{\theta}_n - \theta^*\|^2) = \\ &= \int \log f_{\theta^*} dF_0 + \frac{1}{2} \left(\tilde{\theta}_n - \theta^*\right)^T \Sigma^* \left(\tilde{\theta}_n - \theta^*\right) + o_P(\|\tilde{\theta}_n - \theta^*\|^2), \quad n \rightarrow \infty. \end{aligned}$$

Таким образом, математическое ожидание $\mathbf{E}D_3$ может быть аппроксимировано посредством

$$\begin{aligned} \mathbf{E}D_3 &= \\ &= \mathbf{E} \left(\int \log f_{\theta^*} dF_0 - \int \log f_{\tilde{\theta}_n} dF_0 \right) \sim -\frac{1}{2} \mathbf{E} \left[\left(\tilde{\theta}_n - \theta^*\right)^T \Sigma^* \left(\tilde{\theta}_n - \theta^*\right) \right] = \\ &= -\frac{1}{2} \mathbf{E} \left[\text{tr} \left(\Sigma^* \left(\tilde{\theta}_n - \theta^*\right) \left(\tilde{\theta}_n - \theta^*\right)^T \right) \right] = \\ &= -\frac{1}{2} \text{tr} \left(\Sigma^* \mathbf{E} \left(\tilde{\theta}_n - \theta^*\right) \left(\tilde{\theta}_n - \theta^*\right)^T \right). \quad (3.1) \end{aligned}$$

Так как оценка $\tilde{\theta}_n$ состоятельна, то из теоремы 2.8 можно получить выражение для аппроксимации ковариационной матрицы $\mathbf{E} \left(\tilde{\theta}_n - \theta^*\right) \left(\tilde{\theta}_n - \theta^*\right)^T$:

$$\mathbf{E} \left(\tilde{\theta}_n - \theta^*\right) \left(\tilde{\theta}_n - \theta^*\right)^T \sim \frac{1}{n} (\Sigma^*)^{-1} \int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*}\right)^T dQ (\Sigma^*)^{-1}.$$

Подставляя полученное выражение в формулу (3.1), получаем

$$\mathbf{E}D_3 \sim -\frac{1}{2n} \text{tr} \left(\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*}\right)^T dQ (\Sigma^*)^{-1} \right). \quad (3.2)$$

Перейдем к оценке величины D_1 . Имеем:

$$\begin{aligned}
D_1 &= \int \log f_{\tilde{\theta}_n} d\hat{F}_n - \int \log f_{\theta^*} d\hat{F}_n = \\
&= - \left(\tilde{\theta}_n - \theta^* \right)^T \int \frac{\partial}{\partial \theta} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n - \\
&\quad - \frac{1}{2} \left(\tilde{\theta}_n - \theta^* \right)^T \int \frac{\partial^2}{\partial \theta^2} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n \left(\tilde{\theta}_n - \theta^* \right) + o_P(\|\tilde{\theta}_n - \theta^*\|^2) = \\
&= -\frac{1}{2} \left(\tilde{\theta}_n - \theta^* \right)^T \Sigma^* \left(\tilde{\theta}_n - \theta^* \right) + o_P(\|\tilde{\theta}_n - \theta^*\|^2), \quad n \rightarrow \infty.
\end{aligned}$$

Таким образом, аналогично формуле (3.2) имеем:

$$\mathbf{E}D_1 \sim -\frac{1}{2n} \operatorname{tr} \left(\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ \left(\Sigma^* \right)^{-1} \right). \quad (3.3)$$

Объединяя формулы (3.2) и (3.3), получаем:

$$\mathbf{E}D = \mathbf{E}D_2 - \frac{1}{n} \operatorname{tr} \left(\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ \left(\Sigma^* \right)^{-1} \right). \quad (3.4)$$

Величина $\mathbf{E}D_2$ является смещением оценки $\int \log f_{\theta^*} d\hat{F}_n$ и, вообще говоря, отлична от нуля. Ее оценить можно применив процедуру складного ножа (jackknife) [52, 63] к функционалу

$$F(x) \mapsto \int \log f_{\tilde{\theta}_n}(x) dF(x),$$

вычисляемому в точке \hat{F}_n . Опишем процедуру подробнее.

Пусть, как и прежде, Y_1, \dots, Y_n — n независимых одинаково распределенных случайных величин с распределением Q , индуцированным механизмом интервального цензурирования (T_K, K) . Для $1 \leq i \leq n$ обозначим через $\hat{F}_n^{(i)}$ — оценку функции распределения F_0 , построенную по наблюдениям $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$. Тогда оценкой смещения $\mathbf{E}D_2$ будет служить величина

$$\mathbf{E}\tilde{D}_2 = (n-1) \left(\frac{1}{n} \sum_{i=1}^n \int \log f_{\tilde{\theta}_n} d\hat{F}_n^{(i)} - \int \log f_{\tilde{\theta}_n}(x) d\hat{F}_n \right).$$

Осталось построить оценки для матриц Σ^* и $\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ$. Для первой все достаточно тривиально: состоятельной оценкой матрицы Σ^* служит матрица

$$\tilde{\Sigma}^* = \int \frac{\partial^2}{\partial \theta^2} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n.$$

Поскольку оценка $\tilde{\Sigma}^*$ смещена, то логично скорректировать смещение при помощи процедуры jackknife. В таком случае в качестве оценки матрицы Σ^* будет выступать матрица $\tilde{\Sigma}_{JK}^*$ с

$$\tilde{\Sigma}_{JK}^* = n \int \frac{\partial^2}{\partial \theta^2} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n - \frac{n-1}{n} \sum_{i=1}^n \int \frac{\partial^2}{\partial \theta^2} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n^{(i)}.$$

Для оценивания $\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ$ достаточно заметить, что $\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ$ пропорциональна дисперсии оценки $\int \frac{\partial}{\partial \theta} \log f_{\theta} \Big|_{\theta=\tilde{\theta}_n} d\hat{F}_n$, которую можно состоятельно оценить при помощи процедуры бутстреп [12, 13].

Отдельно остановимся на случае интервального цензурирования первого типа. Здесь канонический градиент $\tilde{\xi}_{\theta^*}$ можно выписать явно и, следовательно, получить оценку дисперсии $\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ$ в явном виде.

3.1.1. Случай интервального цензурирования первого типа

Напомним, что в случае интервального цензурирования первого типа возможен только один момент контроля состояния T и $\mathbf{P}(K=1)=1$. Таким образом, наблюдаемая величина Y состоит из момента контроля состояния объекта T и метки состояния $\Delta = 1_{[X \leq T]}$. В соответствии с формулой (2.15) канонический градиент $\tilde{\xi}_F$ имеет вид:

$$\tilde{\xi}_F(t, \delta) = \begin{cases} -\tilde{\kappa}'_F(t) \frac{1-F(t)}{h(t)}, & \delta = 1, \\ \tilde{\kappa}'_F(t) \frac{F(t)}{h(t)}, & \delta = 0, \end{cases}$$

где $\tilde{\kappa}_F$ — канонический градиент функционала $F(x) \mapsto \int c(x) dF(x)$, равный (в точке F_0) $c(x) - \int c(x) dF_0(x)$. В нашем случае функция $c(x)$ равна

$\frac{\partial}{\partial \theta} \log f_{\theta}(x) \Big|_{\theta=\theta^*}$, откуда получаем:

$$\tilde{\xi}'_{\theta^*}(t, \delta) = \begin{cases} -\frac{\partial}{\partial \theta} \log \frac{f'_{\theta}(t)}{f_{\theta}(t)} \Big|_{\theta=\theta^*} \frac{1-F(t)}{h(t)}, & \delta = 1, \\ \frac{\partial}{\partial \theta} \log \frac{f'_{\theta}(t)}{f_{\theta}(t)} \Big|_{\theta=\theta^*} \frac{F(t)}{h(t)}, & \delta = 0. \end{cases}$$

Для простоты дальнейшего изложения предположим, что $\theta \in \Theta \subset \mathbb{R}$. В таком случае в качестве оценки величины $\int \tilde{\xi}'_{\theta^*} \left(\tilde{\xi}'_{\theta^*} \right)^T dQ = \int \left(\tilde{\xi}'_{\theta^*} \right)^2 dQ$ можно использовать $\int \left(\tilde{\xi}'_{\theta^*} \right)^2 dQ_n$. Оценим градиент $\tilde{\xi}'_{\theta^*}(t, \delta)$ величиной

$$\hat{\xi}'_{\theta^*}(t, \delta) = \begin{cases} -\frac{\partial}{\partial \theta} \log \frac{f'_{\theta}(t)}{f_{\theta}(t)} \Big|_{\theta=\hat{\theta}_n} \frac{1-\hat{F}_n(t)}{\hat{h}_n(t)}, & \delta = 1, \\ \frac{\partial}{\partial \theta} \log \frac{f'_{\theta}(t)}{f_{\theta}(t)} \Big|_{\theta=\hat{\theta}_n} \frac{\hat{F}_n(t)}{\hat{h}_n(t)}, & \delta = 0, \end{cases} \quad (3.5)$$

где $\hat{h}_n(t)$ — некоторая оценка плотности $h(t)$, построенная по T_1, \dots, T_n (см., например, [22]). Наконец, объединяя все вместе, получаем следующую оценку для $\int \left(\tilde{\xi}'_{\theta^*} \right)^2 dQ$:

$$\begin{aligned} \int \left(\hat{\xi}'_{\theta^*} \right)^2 dQ_n &= \frac{1}{n} \sum_{i=1}^n \left(\hat{\xi}'_{\theta^*}(T_i, \Delta_i) \right)^2 = \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial}{\partial \theta} \log \frac{f'_{\theta}(T_i)}{f_{\theta}(T_i)} \Big|_{\theta=\hat{\theta}_n} \right)^2 \frac{\delta_i \left(1 - \hat{F}_n(T_i) \right)^2 + (1 - \delta_i) \hat{F}_n^2(T_i)}{\hat{h}_n^2(T_i)}. \end{aligned}$$

3.2. Информационные критерии в случае ОМП

Как и раньше в качестве оценки величины $\hat{I}_0 = \int \log f_{\hat{\theta}_n} dF_0$ будем использовать величину $\hat{I}_n = \int \log f_{\hat{\theta}_n} d\hat{F}_n$. Оценим смещение величины $D = \hat{I}_n - \hat{I}_0$.

Имеем $D = D_1 + D_2 + D_3$, где

$$\begin{aligned} D_1 &= \int \log f_{\hat{\theta}_n} d\hat{F}_n - \int \log f_{\theta_0} d\hat{F}_n, \\ D_2 &= \int \log f_{\theta_0} d\hat{F}_n - \int \log f_{\theta_0} dF_0, \\ D_3 &= \int \log f_{\theta_0} dF_0 - \int \log f_{\hat{\theta}_n} dF_0. \end{aligned}$$

Оценим величины D_1, D_2, D_3 по отдельности.

Раскладывая функцию $\theta \mapsto \int \log f_\theta dF_0$ вблизи точки θ_0 по формуле Тейлора, имеем:

$$\begin{aligned} \int \log f_{\hat{\theta}_n} dF_0 &= \\ &= \int \log f_{\theta_0} dF_0 + (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right) \Big|_{\theta=\theta_0} dF_0 + \\ &+ \frac{1}{2} (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial^2}{\partial \theta^2} \log f_\theta \right) \Big|_{\theta=\theta_0} dF_0 (\hat{\theta}_n - \theta_0) + o_P(\|\hat{\theta}_n - \theta_0\|^2), n \rightarrow \infty. \end{aligned}$$

Таким образом получаем следующее выражение для D_3 :

$$\begin{aligned} D_3 &= \int \log f_{\theta_0} dF_0 - \int \log f_{\hat{\theta}_n} dF_0 = \\ &= - (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right) \Big|_{\theta=\theta_0} dF_0 - \\ &- \frac{1}{2} (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial^2}{\partial \theta^2} \log f_\theta \right) \Big|_{\theta=\theta_0} dF_0 (\hat{\theta}_n - \theta_0) + o_P(\|\hat{\theta}_n - \theta_0\|^2), n \rightarrow \infty. \end{aligned}$$

Аналогично, раскладывая функцию $\theta \mapsto \int \log f_\theta d\hat{F}_n$ вблизи θ_0 , получаем

$$\begin{aligned} D_1 &= \int \log f_{\hat{\theta}_n} d\hat{F}_n - \int \log f_{\theta_0} d\hat{F}_n = \\ &= (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right) \Big|_{\theta=\theta_0} d\hat{F}_n + \\ &+ \frac{1}{2} (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial^2}{\partial \theta^2} \log f_\theta \right) \Big|_{\theta=\theta_0} d\hat{F}_n (\hat{\theta}_n - \theta_0) + o_P(\|\hat{\theta}_n - \theta_0\|^2), n \rightarrow \infty. \end{aligned}$$

Отсюда имеем

$$\mathbf{E}(D_1 + D_3) \sim (\hat{\theta}_n - \theta_0)^T \int \left(\frac{\partial}{\partial \theta} \log f_\theta \right) \Big|_{\theta=\theta_0} d(\hat{F}_n - F_0). \quad (3.6)$$

Обозначим

$$u_i(\theta_0) = \sum_{j=1}^{K^{(i)}+1} \Delta_{K,j}^{(i)} \frac{\partial}{\partial \theta} \log \left[F_{\theta}(T_{K,j}^{(i)}) - F_{\theta}(T_{K,j-1}^{(i)}) \right] \Big|_{\theta=\theta_0}.$$

Тогда, из теоремы 1.8 для ОМП $\hat{\theta}_n$ имеем:

$$\hat{\theta}_n - \theta_0 = -\Sigma_0^{-1} \frac{1}{n} \sum_{i=1}^n u_i(\theta_0) + o_p(1/\sqrt{n}), \quad n \rightarrow \infty.$$

Обозначим

$$v_i(\theta_0) = \tilde{\xi}_{\theta_0}' \left(\Delta_K^{(i)}, T_K^{(i)}, K^{(i)} \right) - \int \tilde{\xi}_{\theta_0}' dQ.$$

Из теоремы 2.3 следует, что

$$\int \left(\frac{\partial}{\partial \theta} \log f_{\theta} \right) \Big|_{\theta=\theta_0} d(\hat{F}_n - F_0) = \frac{1}{n} \sum_{i=1}^n v_i(\theta_0) + o_p(1/\sqrt{n}), \quad n \rightarrow \infty.$$

Тогда во введенных обозначениях из формулы (3.6) получаем следующую аппроксимацию для $\mathbf{E}(D_1 + D_3)$:

$$\mathbf{E}(D_1 + D_3) \sim -\frac{1}{n^2} \mathbf{E} \left(\sum_{i=1}^n u_i^T(\theta_0) \Sigma_0^{-1} v_i(\theta_0) \right) = -\frac{1}{n} \mathbf{E} (u_i^T(\theta_0) \Sigma_0^{-1} v_i(\theta_0)).$$

Несложно видеть, что $\mathbf{E}v_i(\theta_0) = 0$ и $\mathbf{E}u_i(\theta_0) = 0$. Поэтому

$$-\frac{1}{n} \mathbf{E} (u_i^T(\theta_0) \Sigma_0^{-1} v_i(\theta_0)) = -\frac{1}{n} \text{tr} (\Sigma_0^{-1} \mathbf{E} (v_i(\theta_0) u_i^T(\theta_0))).$$

Величина $\mathbf{E}D_2$ является смещением оценки $\int \log f_{\theta_0} d\hat{F}_n$. Как и прежде ее можно оценить при помощи процедуры jackknife:

$$\mathbf{E}\hat{D}_2 = (n-1) \left(\frac{1}{n} \sum_{i=1}^n \int \log f_{\hat{\theta}_n} d\hat{F}_n^{(i)} - \int \log f_{\hat{\theta}_n}(x) d\hat{F}_n \right).$$

В качестве оценки матрицы Σ_0 можно взять

$$\hat{\Sigma}_0 = \int \frac{\partial^2}{\partial \theta^2} \log f_{\theta} \Big|_{\theta=\hat{\theta}_n} d\hat{F}_n,$$

либо матрицу $\hat{\Sigma}_0$ со скорректированным смещением посредством применения процедуры jackknife:

$$\hat{\Sigma}_{0,JK} = n \int \frac{\partial^2}{\partial \theta^2} \log f_\theta \Big|_{\theta=\hat{\theta}_n} d\hat{F}_n - \frac{n-1}{n} \sum_{i=1}^n \int \frac{\partial^2}{\partial \theta^2} \log f_\theta \Big|_{\theta=\hat{\theta}_n} d\hat{F}_n^{(i)}.$$

Величину $\mathbf{E}(v_i(\theta_0)u_i^T(\theta_0))$ так же, как и раньше, можно оценить при помощи процедуры бутстреп.

Остановимся на случае интервального цензурирования первого типа.

3.2.1. Случай интервального цензурирования первого типа

Ситуация здесь во многом аналогична изложенной в разделе 3.1.1. Из формулы (3.5) следует, что в качестве оценки градиента $\tilde{\xi}_{\theta_0}$ можно использовать

$$\tilde{\xi}'_{\theta_0}(t, \delta) = \begin{cases} -\frac{\partial}{\partial \theta} \log \frac{f'_\theta(t)}{f_\theta(t)} \Big|_{\theta=\hat{\theta}_n} \frac{1 - \hat{F}_n(t)}{\hat{h}_n(t)}, & \delta = 1, \\ \frac{\partial}{\partial \theta} \log \frac{f'_\theta(t)}{f_\theta(t)} \Big|_{\theta=\hat{\theta}_n} \frac{\hat{F}_n(t)}{\hat{h}_n(t)}, & \delta = 0, \end{cases}$$

где $\hat{h}_n(t)$ — некоторая оценка плотности $h(t)$, построенная по T_1, \dots, T_n . Предположим, что $\theta \in \Theta \subset \mathbb{R}$. Тогда, оценивая ковариацию $\int v_i(\theta_0)u_i(\theta_0) dQ$ интегралом по эмпирической мере \mathbb{Q}_n , получаем

$$\begin{aligned} \int v_i(\theta_0)u_i(\theta_0) d\mathbb{Q}_n &= \int \tilde{\xi}_{\theta_0} u_i(\theta_0) d\mathbb{Q}_n - \int \tilde{\xi}_{\theta_0} d\mathbb{Q}_n \int u_i(\theta_0) d\mathbb{Q}_n \sim \\ &\sim -\frac{1}{n} \left(\sum_{i=1}^n \frac{\partial}{\partial \theta} \left(\log \frac{f'_\theta(T^{(i)})}{f_\theta(T^{(i)})} + F_\theta(T^{(i)}) \right) \Big|_{\theta=\hat{\theta}_n} \frac{\Delta^{(i)} + (1 - 2\Delta^{(i)})\hat{F}_n(T^{(i)})}{\hat{h}_n(T^{(i)})} + \right. \\ &+ \left. \left(\sum_{i=1}^n (1 - 2\Delta^{(i)}) \frac{\partial}{\partial \theta} \log \frac{f'_\theta(T^{(i)})}{f_\theta(T^{(i)})} \Big|_{\theta=\hat{\theta}_n} \frac{\Delta^{(i)} + (1 - 2\Delta^{(i)})\hat{F}_n(T^{(i)})}{\hat{h}_n(T^{(i)})} \right) \times \right. \\ &\quad \left. \times \sum_{i=1}^n \left(2\Delta^{(i)} - 1 \right) \frac{\partial}{\partial \theta} F_\theta(T^{(i)}) \Big|_{\theta=\hat{\theta}_n} \right). \end{aligned}$$

Моделирование

Моделирование производилось в пакете R версии 2.11.1 [53] на четырех-процессорном компьютере под управлением ОС Linux. Ввиду значительной вычислительной сложности задачи (время, необходимое для получения одной реализации оценок параметров, достигало нескольких минут для достаточно больших объемов выборки n) моделирование производилось параллельно на всех процессорах компьютера. Для организации процедуры распределения задач по отдельным процессорам использовались пакеты snow и Rmpi [62, 71]. Для процедуры оптимизации с учетом ограничений на пространство параметров использовался алгоритм L-BFGS-B [10, 72].

Оценки параметров исследовались по нескольким характеристикам:

1. Асимптотические свойства: состоятельность, скорость сходимости к предельному распределению.
2. Свойства при фиксированном объеме выборки: дисперсия, среднеквадратическое отклонение.
3. Поведение при оценивании в неточно заданной модели.

Далее приведены результаты моделирования для ОМП $\hat{\theta}_n$ и ОМКЛ $\tilde{\theta}_n$ для ряда стандартных специальных моделей широко используемых при анализе данных типа времени жизни.

В качестве механизма цензурирования была выбрана модель, задаваемая формулой (1.1). Очевидно, что вышеобозначенные свойства оценок сильно зависят от того, насколько много информации о выборке теряется. В частности, ясно, что чем больше величина $\mathbf{P}(X > L)$, где L — общая продолжительность наблюдений, тем больше потеря информации о выборке. В связи с

этим процедура моделирования производилась дважды с $\mathbf{P}(X > L) = 0.2$ и $\mathbf{P}(X > L) = 0.4$. Величины Z_i имели экспоненциальное распределение со сдвинутым носителем (это необходимо для выполнения условий теоремы 2.4). Параметр экспоненциального распределения подбирался так, чтобы $\mathbf{E}K = 10$.

4.1. Модели данных типа времени жизни

Параметрические модели распределений типа времени жизни как правило задаются в терминах *функции риска* (hazard rate) [81] $h(t)$:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t \leq X < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)},$$

здесь X — случайная величина, имеющая смысл времени наступления до какого-либо события, $F(t)$ — ее функция распределения, а $f(t)$ — плотность.

Функция риска показывает вероятность отказа в момент t при условии, что отказа не было до этого момента: так, величина $h(t)\Delta t$ приближенно совпадает с вероятностью возникновения отказа на промежутке $[t, t + \Delta t)$, при условии, что он не произошел раньше.

Задав функцию конкретную $h(t)$, несложно получить выражения для функции распределения и плотности:

$$F(t) = 1 - \exp\left(-\int_0^t h(t) dt\right),$$

$$f(t) = h(t) \exp\left(-\int_0^t h(t) dt\right).$$

4.1.1. Распределения, связанные с экспоненциальным

Экспоненциальное распределение характеризуется постоянной функцией риска:

$$h(t) = \lambda, \quad t \geq 0.$$

Как правило, наличие такой функции риска означает, что событие отказа возникает из-за каких-либо внешних причин (например, несчастного случая), а не связана с характеристиками самого индивида.

Распределение Вейбулла характеризуется полиномиальной функцией риска следующего вида:

$$h(t) = \lambda\beta(\lambda t)^{\beta-1}, \quad t \geq 0,$$

где $\lambda > 0$, $\beta > 0$ — параметры. Отсюда имеем следующую функцию распределения:

$$F(t) = 1 - e^{-(\lambda t)^\beta}, \quad t > 0.$$

Данная модель является достаточно гибкой, так, функция риска монотонно возрастает при $\beta > 1$, постоянна при $\beta = 1$, и монотонно убывает при $\beta < 1$. В приложения типичные значения параметра β лежат в пределах от 0.5 до 3 [41].

Обобщенное гамма-распределение было предложено в статье [58] как распределение, включающее в качестве частного случая многие стандартные: экспоненциальное, Вейбулла, гамма, лог-нормальное и пр. Его плотность имеет вид

$$f(t) = \frac{b}{a\Gamma(k)} \left(\frac{t}{a}\right)^{bk-1} e^{-(t/a)^b}, \quad t > 0.$$

Однако, данная (естественная) параметризация не очень удобна для задач оценивания. В частности, в такой параметризации дисперсия неограниченно возрастает с ростом параметра k . В связи с этим в работе [51] была предложена альтернативная параметризация, более удобная для построения оценок.

Введем следующие обозначения: $\lambda = \frac{1}{\sqrt{k}}$, $\mu = \log a + \frac{\log k}{b}$, $\sigma = \frac{1}{b\sqrt{k}}$. В этой параметризации имеем:

$$f(t) = \begin{cases} \frac{\lambda}{\sigma\Gamma(\frac{1}{\lambda^2})t} \exp\left(\frac{\log t - \mu}{\lambda\sigma} - \frac{1}{\lambda^2} e^{\frac{\lambda(\log t - \mu)}{\sigma}}\right), & \lambda > 0 \\ \frac{1}{\sqrt{2\pi\sigma^2}t} \exp\left(-\frac{\log t - \mu}{2\sigma}\right), & \lambda = 0. \end{cases}$$

4.1.2. Модель Гомперца-Макегама

Модель Гомперца-Макегама [21, 44] была предложена как адекватная модель, описывающая закон смертности в человеческой популяции на отрезке 35-80 лет. Предполагается, что функция риска $h(t)$ содержит как постоянную компоненту, отражающую убыль популяции за счет независимых факторов (несчастные случаи и т.п.), так и экспоненциальную, отражающую смертность за счет естественных факторов старения. Таким образом, приходим к следующей функции риска:

$$h(t) = \gamma + \alpha e^{\beta t}, \quad t > 0$$

и функции распределения

$$F(t) = 1 - \exp\left(-\gamma t - \frac{\alpha}{\beta} (e^{\beta t} - 1)\right), \quad t > 0$$

Типичные значения характерных для человеческой популяции параметров лежат в пределах (t – время в годах) [5, 32]: $0.000001 < \alpha < 0.01$, $1.08 < e^{\beta} < 1.12$, $0.001 < \gamma < 0.003$.

4.1.3. Модель ExpCos А.Г. Барта

Специальная модель кривой дожития в виде произведения экспоненты на косинус была впервые предложена в работе [76]. Формальным основанием для использования этой модели в биометрии служит разработанный в [80] лагранжево-гамильтоновыи механизм для биологической систематики,

согласно которому компоненты в виде произведения экспоненты на косинус являются решением системы, описывающей взаимодействие факторов органа и организма. Модель применялась для описания динамики хронического гломерулонефрита [76], гипертонической болезни [6], раневых процессов [75], хронического генерализованного пародонтита [86].

Модель представляется удобной для интерпретации экспериментаторами реальных данных, так как позволяет учесть сложный дискретно-непрерывный характер процессов патогенеза-саногенеза, выражающийся в наличии чередующихся периодов обострения и ремиссии. Кроме этого, существует модификация модели, включающая управление через процедуру двойного обращения [77].

Функция распределения для этой модели имеет вид

$$F(t) = 1 - e^{-\eta t} \cos\left(\frac{\pi}{2\tau}t\right), \quad 0 \leq t \leq \tau.$$

4.2. Асимптотические свойства ОМП

Как было отмечено выше, при моделировании нас будет интересовать состоятельность оценок и скорость сходимости к предельному распределению. В силу неравенства Чебышева для этого достаточно установить зависимость дисперсии оценок от объема выборки n , в частности, достаточно проверить, что дисперсия ОМП $\hat{\theta}_n$ стремится к нулю как только n стремится к бесконечности.

Для этого построим графики зависимости выборочного стандартного отклонения оценок $\hat{\theta}_n$ от объема выборки n (выборка оценок имела объем $m = 100$). Из теоремы 1.8 следует, что оценки параметров имеют типичную скорость сходимости порядка $O_P(1/\sqrt{n})$, поэтому для построения зависимостей оценки были центрированы и нормированы на \sqrt{n} (для сравнения на

рис. 4.1(а), изображена зависимость стандартного отклонения оценок параметров без дополнительной нормировки).

4.2.1. Распределение Вейбулла

Моделирование производилось для двух наборов параметров: $\lambda = 1, \beta = 0.7$ и $\lambda = 1, \beta = 1.5$. Таким образом мы учли все возможные формы функции риска.

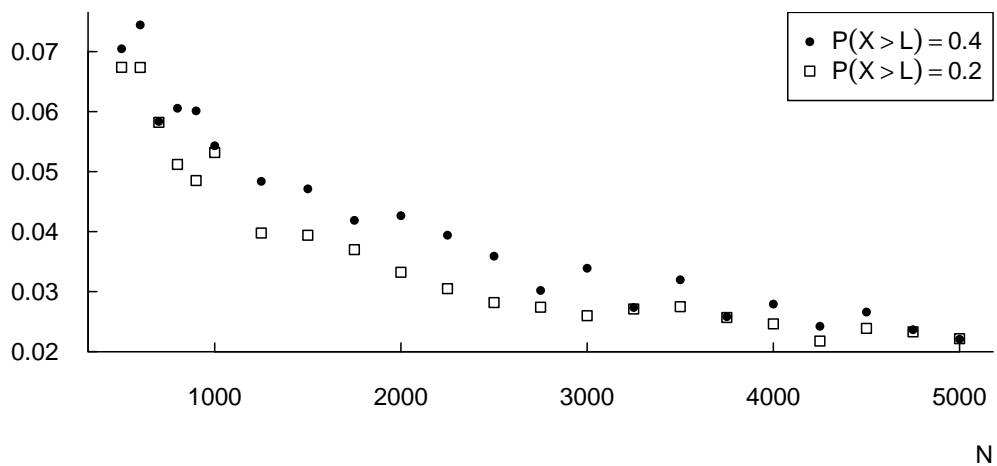
Результаты моделирования представлены на рис. 4.1, 4.2. Мы видим четкую зависимость стандартного отклонения оценок от уровня цензурирования: чем выше потеря информации о выборке, тем выше стандартное отклонение оценок. Кроме этого, моделирование подтверждает выводы теоремы 1.8 относительно скорости сходимости оценок параметров: здесь она $O_P(1/\sqrt{n})$.

4.2.2. Обобщенное гамма-распределение

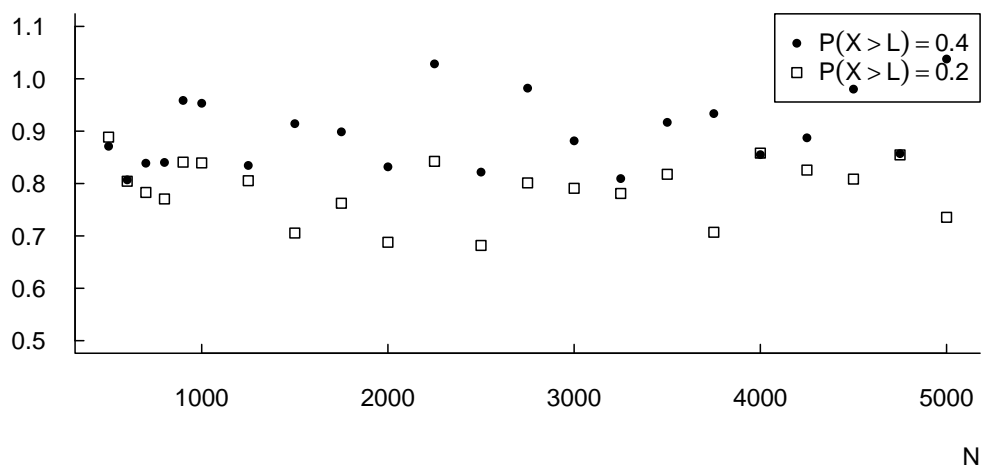
В исходной параметризации гамма-распределению соответствует значение параметра $b = 1$. Для моделирования мы использовали значения параметров, полученные в работах [2, 73] при анализе реальных медико-биологических данных типа «мышца»: $k = 10, a = 0.93$, или в параметризации из [51]: $\lambda = \sigma = 0.32, \mu = 2.23$. Результаты моделирования представлены в таблице 4.1.

4.2.3. Модель Гомперца-Макегама

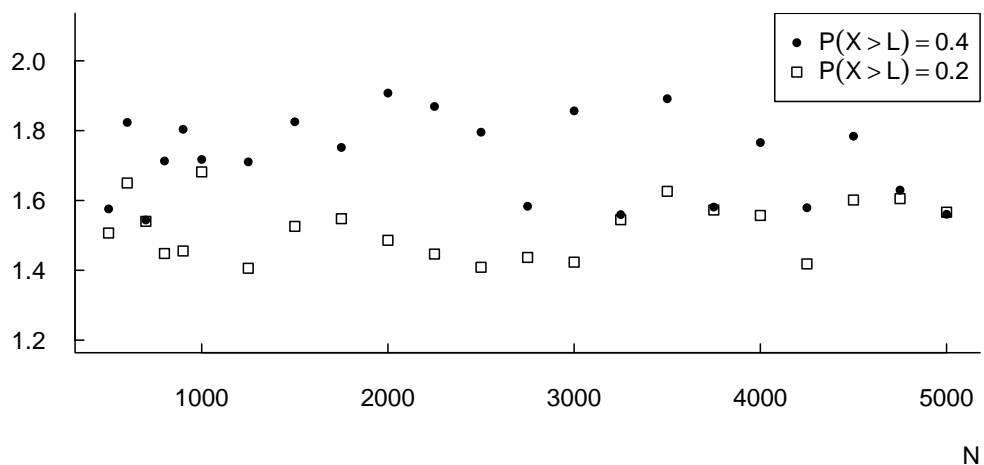
В качестве параметров использовались следующие значения, взятые из стандартной таблицы G82M, описывающей продолжительность жизни мужского населения Дании [4, 7]: $\alpha = 0.0005, \beta = 0.08749825, \gamma = 7.585776 \times 10^{-5}$. Результаты моделирования представлены в таблице 4.2. Из нее видно, что интересующие свойства оценок выполняются для данной модели. Стоит от-



(a) Зависимость $\text{std}(\hat{\beta}_n)$ от объема выборки.

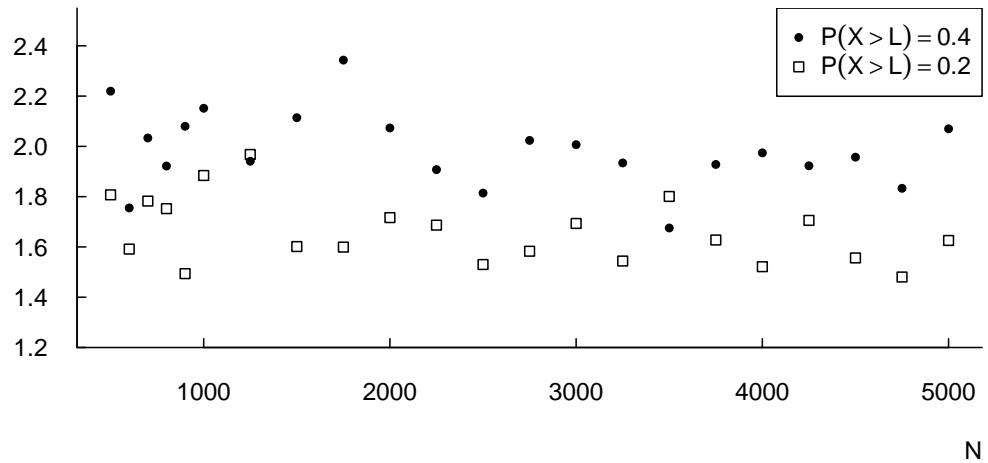


(b) Зависимость $\sqrt{n} \text{std}(\hat{\lambda}_n)$ от объема выборки.

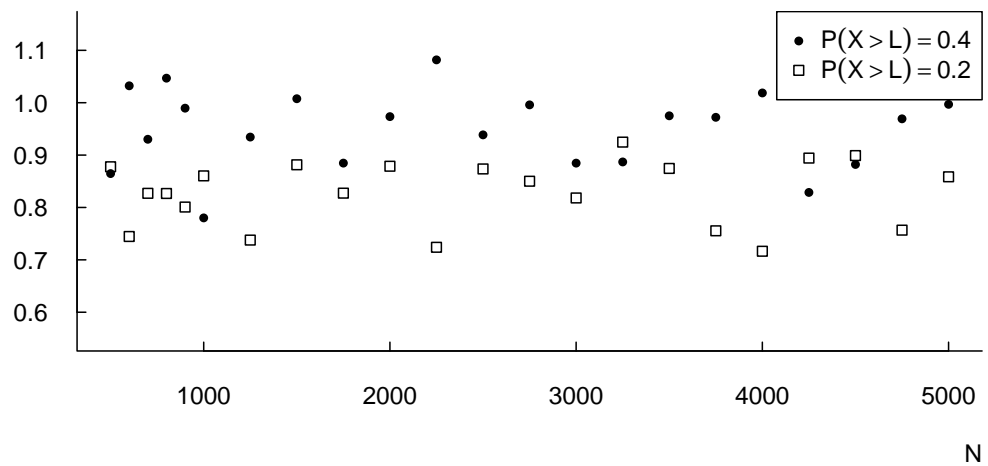


(c) Зависимость $\sqrt{n} \text{std}(\hat{\beta}_n)$ от объема выборки.

Рис. 4.1. Зависимость стандартного отклонения оценок $\hat{\beta}_n$ и $\hat{\lambda}_n$ от объема выборки. Распределение Вейбулла, $\beta_0 = 1.5$



(a) Зависимость $\sqrt{n} \text{std}(\hat{\lambda}_n)$ от объема выборки.



(b) Зависимость $\sqrt{n} \text{std}(\hat{\beta}_n)$ от объема выборки.

Рис. 4.2. Зависимость стандартного отклонения оценок $\hat{\beta}_n$ и $\hat{\lambda}_n$ от объема выборки. Распределение Вейбулла, $\beta_0 = 0.7$

n	100	1000	2500	5000	10000
$P(X > L) = 0.2$					
$\text{std}(\hat{\lambda}_n)$	5.27×10^{-5}	1.63×10^{-5}	1.19×10^{-5}	9.75×10^{-6}	5.53×10^{-6}
$\text{std}(\hat{\mu}_n)$	0.0251	0.0097	0.007	0.00429	0.00234
$\text{std}(\hat{\sigma}_n)$	2.18×10^{-4}	8.29×10^{-5}	3.58×10^{-5}	2.69×10^{-5}	2.63×10^{-5}
$P(X > L) = 0.4$					
$\text{std}(\hat{\lambda}_n)$	4.96×10^{-4}	1.44×10^{-4}	9.98×10^{-5}	5.79×10^{-5}	4.45×10^{-5}
$\text{std}(\hat{\mu}_n)$	0.12	0.041	0.02	0.02	0.013
$\text{std}(\hat{\sigma}_n)$	5.19×10^{-4}	1.72×10^{-4}	8.84×10^{-5}	6.74×10^{-5}	4.58×10^{-5}

Таблица 4.1. Стандартное отклонение оценок $\hat{\lambda}_n, \hat{\mu}_n, \hat{\sigma}_n$. Обобщенное гамма-распределение.

метить несколько выделяющийся характер результатов для объема выборки $n = 10000$. В данном случае функция правдоподобия имела несколько близко расположенных локальных экстремумов вблизи глобального, и, вследствие этого, процедура численной оптимизации не всегда могла его определить с требуемой точностью.

n	100	1000	2500	5000	10000
$P(X > L) = 0.2$					
$\text{std}(\hat{\alpha}_n)$	1.58×10^{-4}	3.61×10^{-5}	4.86×10^{-5}	8.35×10^{-6}	1.77×10^{-5}
$\text{std}(\hat{\beta}_n)$	0.03	0.0022	0.0035	0.00034	0.003
$\text{std}(\hat{\gamma}_n)$	1.44×10^{-5}	7.68×10^{-7}	5.17×10^{-6}	2.72×10^{-6}	7.47×10^{-7}
$P(X > L) = 0.4$					
$\text{std}(\hat{\alpha}_n)$	3.52×10^{-4}	5.50×10^{-5}	9.34×10^{-5}	9.45×10^{-6}	1.95×10^{-5}
$\text{std}(\hat{\beta}_n)$	0.045	0.0046	0.011	0.0032	0.0045
$\text{std}(\hat{\gamma}_n)$	1.60×10^{-5}	1.49×10^{-6}	8.76×10^{-6}	6.87×10^{-6}	1.10×10^{-6}

Таблица 4.2. Стандартное отклонение оценок $\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n$. Модель Гомперца-Макегама.

4.2.4. Модель ExpCos

Для моделирования значения параметров были выбраны совпадающими с оценками, полученными в работе [84] для реальных данных из кардиологии: $\eta = 0.125$, $\tau = 16$. Заметим, что носитель распределения зависит от параметра τ и поэтому ОМП этого параметр по выборке без цензурирования обладает свойством сверхэффективности [78]: скорость сходимости составляет $O_P(1/n)$ против ожидаемого $O_P(1/\sqrt{n})$ (см. зависимость на рисунке 4.3).

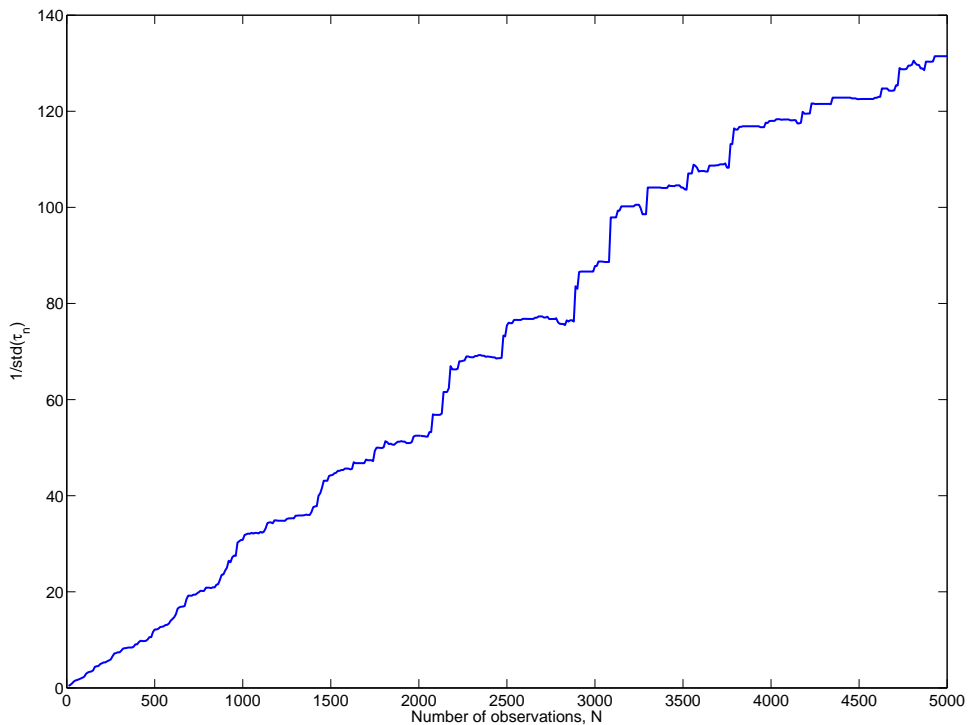


Рис. 4.3. $1/\text{std}(\hat{\tau}_n)$ при оценивании по выборке без цензурирования

Ситуация кардинально меняется при оценивании в условиях цензурирования: выполняются условия теоремы 1.8, вследствие этого $\hat{\tau}_n$ асимптотически нормальна (в частности, поэтому имеет скорость сходимости $O_P(1/\sqrt{n})$). Данный результат подтверждается моделированием, результаты которого показаны на рисунке 4.4

Качественное поведение оценки $\hat{\eta}_n$ не меняется, как показано в [78] она асимптотически нормальна во всех случаях.

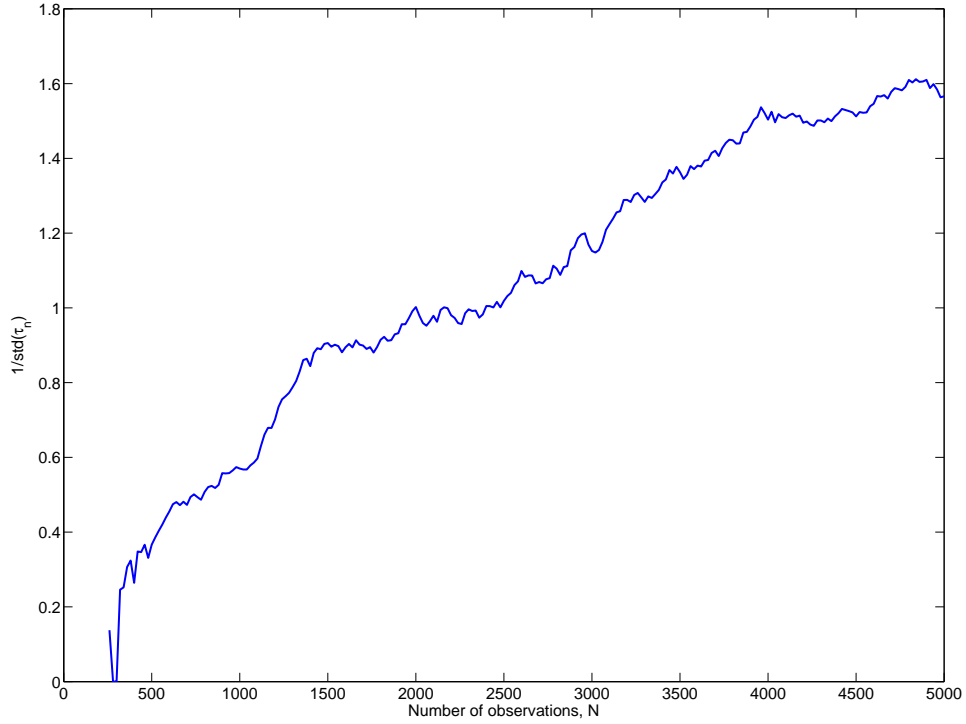


Рис. 4.4. $1/\text{std}(\hat{\tau}_n)$ при оценивании по выборке в условиях интервального цензурирования

4.3. Асимптотические свойства ОМКЛ

Как было отмечено в разделе 2.5, ОМП \hat{F}_n представляет из себя класс эквивалентности функций распределения, для которых известен набор интервалов $\mathcal{H} = \{H_i\}$, на которых они доставляют массу $p_i = \mathbf{P}_F(H_i)$. Вне набора интервалов \mathcal{H} ОМП \hat{F}_n постоянна.

Таким образом, для вычисления ОМКЛ $\tilde{\theta}_n$ необходимо вычислять функцию

$$l(\theta) = \int p_\theta d\hat{F}_n = \sum_{H_i \in \mathcal{H}} \int_{H_i} \log p_\theta d\hat{F}_n,$$

которая, вообще говоря, корректно не определена, так как зависит от поведения функции \hat{F}_n внутри каждого интервала H_i .

Однако, если интеграл вида $\int_{H_i} \log p_\theta d\hat{F}_n$ аппроксимировать при помощи квадратурной суммы, то полученное выражение зависит только от значения величины $\mathbf{P}_F(H_i)$, и, следовательно, такая сумма определена корректно. В

связи с этим рассмотрим следующие аппроксимации для $l(\theta)$:

$$\begin{aligned}
l_n^{(1)}(\theta) &= \sum_{H_i \in \mathcal{H}} p_i \log p_\theta(a_i), \\
l_n^{(2)}(\theta) &= \sum_{H_i \in \mathcal{H}} p_i \log p_\theta(b_i), \\
l_n^{(3)}(\theta) &= \frac{1}{2} \sum_{H_i \in \mathcal{H}} p_i (\log p_\theta(a_i) + \log p_\theta(b_i)), \\
l_n^{(4)}(\theta) &= \frac{1}{6} \sum_{H_i \in \mathcal{H}} p_i [\log p_\theta(a_i) + 4 \log p_\theta((a_i + b_i)/2) + \log p_\theta(b_i)],
\end{aligned}$$

где $H_i = [a_i, b_i)$. Кроме этого, известно, что интеграл $\int_{H_i} \log p_\theta d\hat{F}_n$ является смещенной оценкой величины $\int_{H_i} \log p_\theta dF$. Скорректируем это смещение при помощи процедуры jack-knife:

$$l_n^{(jk)}(\theta) = n l_n^{(4)}(\theta) - \frac{n-1}{n} \sum_{i=1}^n l_{n-1,i}^{(4)}(\theta),$$

где $l_{n-1,i}^{(4)}(\theta)$ — аппроксимация величины $\int \log p_\theta d\hat{F}_n^{(i)}$ при помощи функции $l_n^{(4)}$, а $\hat{F}_n^{(i)}$ — ОМП для F , построенная по $Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$.

После этого определим оценки $\tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)}, \tilde{\theta}_n^{(3)}, \tilde{\theta}_n^{(4)}, \tilde{\theta}_n^{(jk)}$ как точки максимумов соответствующих функций $l_n^{(1)}(\theta), l_n^{(2)}(\theta), l_n^{(3)}(\theta), l_n^{(4)}(\theta), l_n^{(jk)}(\theta)$.

4.3.1. Сравнение оценок $\tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)}, \tilde{\theta}_n^{(3)}, \tilde{\theta}_n^{(4)}, \tilde{\theta}_n^{(jk)}$

Для начала рассмотрим асимптотические свойства различных аппроксимаций оценки $\tilde{\theta}_n$, введенных выше. Для этого сравним дисперсии и среднеквадратическое отклонение этих оценок в модели распределения Вейбулла ($\beta_0 = 0.7, \lambda_0 = 1$) и интервального цензурирования первого типа. Объем выборки n был фиксирован на величине 10000. Уровень цензурирования (т.е. величина $\mathbf{P}(X > L)$) был равен 0.2. Результаты моделирования представлены в таблице 4.3.

	$\tilde{\theta}_n^{(1)}$	$\tilde{\theta}_n^{(2)}$	$\tilde{\theta}_n^{(3)}$	$\tilde{\theta}_n^{(4)}$	$\tilde{\theta}_n^{(jk)}$
$\text{std}(\tilde{\beta}_n)$	2.26×10^{-2}	2.78×10^{-2}	2.01×10^{-2}	1.69×10^{-2}	1.73×10^{-2}
$\text{MSE}(\tilde{\beta}_n)$	3.73×10^{-2}	4.24×10^{-2}	3.65×10^{-2}	3.21×10^{-2}	2.97×10^{-2}
$\text{std}(\tilde{\lambda}_n)$	3.82×10^{-2}	5.73×10^{-2}	3.35×10^{-2}	2.78×10^{-2}	3.13×10^{-2}
$\text{MSE}(\tilde{\lambda}_n)$	5.10×10^{-2}	6.32×10^{-2}	4.97×10^{-2}	4.14×10^{-2}	4.01×10^{-2}

Таблица 4.3. Стандартное отклонение и среднеквадратичное отклонение для различных аппроксимаций ОМКЛ. Распределение Вейбулла, $\lambda_0 = 1$, $\beta_0 = 0.7$.

Как видно, оценка $\tilde{\theta}_n^{(2)}$ обладает наибольшей дисперсией и среднеквадратичным отклонением. По-видимому, это связано с тем, что величина $\int f_\theta d\hat{F}_n$ как оценка интеграла $\int f_\theta dF_0$ имеет отрицательное смещение. Перенос массы распределения \hat{F}_n на правую границу интервала только увеличивает это смещение.

Оценки $\tilde{\theta}_n^{(4)}$ и $\tilde{\theta}_n^{(jk)}$ являются наилучшими по сравнению с остальными оценками, при этом первая обладает наименьшей дисперсией, а вторая — среднеквадратичным отклонением. Таким образом, в приложениях имеет смысл останавливаться на этих двух оценках.

Для простоты мы не будем демонстрировать таблицы, содержащие непосредственно значения дисперсий оценок параметром. Выделим общие качественные выводы, которые были сделаны в результате моделирования.

1. Оценки $\tilde{\theta}_n$ обладают скоростью сходимости порядка $O_P(1/\sqrt{n})$.
2. С увеличением уровня общего цензурирования и, как следствие, с увеличением потери информации о выборке растет дисперсия оценок.
3. Как правило, дисперсия оценок $\tilde{\theta}_n$ больше, нежели дисперсия оценок $\hat{\theta}_n$, построенных в идентичных условиях. Несколько подробнее этот вопрос рассмотрен ниже.

4.4. Сравнение ОМП $\hat{\theta}_n$ и ОМКЛ $\tilde{\theta}_n$

Как показано выше ОМП $\hat{\theta}_n$ обладают в большинстве случаев меньшей дисперсией по сравнению с ОМКЛ $\tilde{\theta}_n$. Сравним дополнительно среднеквадратичное отклонение оценок при оценивании в случае точной и неточной модели.

Параметры моделирования (распределение, механизм цензурирования) совпадали с использованными в разделе 4.3.1 для сравнения оценок $\tilde{\theta}_n^{(1)}$, $\tilde{\theta}_n^{(2)}$, $\tilde{\theta}_n^{(3)}$, $\tilde{\theta}_n^{(4)}$, $\tilde{\theta}_n^{(jk)}$. В качестве неточной модели использовалось экспоненциальное распределение с одним параметром масштаба. Использование таких простых моделей как для распределения интересующей случайной величины, так и для механизма цензурирования позволило вычислить в явном виде предельные точки θ^* для оценок $\tilde{\theta}_n$ и $\hat{\theta}_n$ и тем самым оценить среднеквадратичное отклонение для этих оценок в случае, когда предполагаемая модель данных не точна.

Результаты моделирования представлены в таблицах 4.4, 4.5.

	$P(X > L) = 0.2$	$P(X > L) = 0.4$
$\text{std}(\hat{\lambda}_n)$	1.72×10^{-2}	2.16×10^{-2}
$\text{std}(\hat{\beta}_n)$	0.78×10^{-2}	1.03×10^{-2}
$\text{MSE}(\hat{\lambda}_n)$	2.01×10^{-2}	2.37×10^{-2}
$\text{MSE}(\hat{\beta}_n)$	1.15×10^{-2}	1.21×10^{-2}
$\text{std}(\tilde{\lambda}_n)$	3.13×10^{-2}	3.35×10^{-2}
$\text{std}(\tilde{\beta}_n)$	1.73×10^{-2}	1.92×10^{-2}
$\text{MSE}(\tilde{\lambda}_n)$	4.01×10^{-2}	4.27×10^{-2}
$\text{MSE}(\tilde{\beta}_n)$	2.97×10^{-2}	3.16×10^{-2}

Таблица 4.4. Стандартное отклонение и среднеквадратичное отклонение для ОМП и ОМКЛ. Распределение Вейбулла, $\lambda_0 = 1$, $\beta_0 = 0.7$.

	$P(X > L) = 0.2$	$P(X > L) = 0.4$
$\text{std}(\hat{\lambda}_n)$	5.83×10^{-2}	6.92×10^{-2}
$\text{MSE}(\hat{\lambda}_n)$	7.32×10^{-2}	7.56×10^{-2}
$\text{std}(\tilde{\lambda}_n)$	3.78×10^{-2}	3.97×10^{-2}
$\text{MSE}(\tilde{\lambda}_n)$	4.21×10^{-2}	4.81×10^{-2}

Таблица 4.5. Стандартное отклонение и среднеквадратичное отклонение для ОМП и ОМКЛ при оценивании вне модели.

Из них мы можем заключить, что в том случае, когда предполагаемая параметрическая модель точна, оценка $\hat{\theta}_n$ имеет меньшую дисперсию и среднеквадратичное отклонение. Ситуация совершенно противоположная при оценивании «вне модели»: тут оценка $\hat{\theta}_n$ имеет как большее смещение, так и дисперсию. Более того, величина отклонения возрастает с ростом потери информации о выборке.

Анализ реальных данных

В данном разделе будет рассмотрено несколько примеров обработки реальных данных из медицины. В большинстве из них в качестве модели кривой дожития используется модель А.Г. Барта [76]. Ее параметры (η и τ) оказались удобными для интерпретации экспериментаторами: η информативен с точки зрения интенсивности наблюдаемого процесса, а τ — его максимальной возможной продолжительности.

5.1. Пример из стоматологии

Сравнивались два способа лечения хронического генерализованного пародонтита у больных пожилого и старческого возраста: консервативное лечение и лечение с применением пептидного биорегулятора «Вилон», способствующего оптимизации процессов регенерации тканей пародонта. Клинический этап исследования осуществлялся в Тарховском военном санатории МО РФ в 2002-2005 гг. [86]

Сравнение качественной динамики лечения хронического генерализованного пародонтита определялось по изменению во времени основного признака тяжести течения заболевания — средней глубины пародонтальных карманов. С точки зрения протекания процесса излечения нормальным считается глубина кармана в 2.5 мм и меньше. В связи с этим в качестве интересующей случайной величины использовалось время до наступления уменьшения глубины всех карманов до величины 2.5 мм

Объем выборки в основной группе (получавшие лечение «Вилоном») составил 38 индивидов, а контрольной (получавшими консервативное лече-

ние) — 56 индивидов. Наблюдение за состоянием пациентов прекращались на 10 сутки, кроме этого имело место интервальное цензурирование. Оценивание параметров модели с игнорированием факта наличия интервального цензурирования (учитывалось только прекращение наблюдений в виде цензурирования справа) проводилось в работе [86]. Результаты представлены в таблице 5.1.

Группа	$\hat{\eta}_n$	$\hat{\tau}_n$
Основная	2.29	12.4
Контрольная	0.001	15.8

Таблица 5.1. Оценки параметров модели из [86]

В работе [85] мы учли наличие интервального цензурирования и произвели оценивание при помощи ОМКЛ. Результаты представлены в таблице 5.2.

Группа	$\tilde{\eta}_n$	$\tilde{\tau}_n$
Основная	1.73	13.4
Контрольная	0.02	21.8

Таблица 5.2. Оценки параметров модели с учетом цензурирования

Как видно, учет цензурирования привел к существенно иным значениям оценок. Тем не менее, очевидно, что с точки зрения протекания процессов саногенеза-патогенеза группы качественно отличаются. Наблюдаются различия как в интенсивности процесса восстановления тканей (параметр η) так и в сроках окончания процесса (параметр τ).

Более того, следуя общему подходу к модели с точки зрения системы «орган-организм» [75, 76], можно сделать вывод, что при лечении с применением «Вилона» преобладают изменения показателей состояния тканей пародонта

в сторону выздоровления на уровня организма, чего нельзя сказать в случае применения только консервативного метода лечения у больных старческого и пожилого возраста.

5.2. Пример из кардиологии

Рассматривалась группа больных гипертонической болезнью. Исследовалась зависимость частоты наступления комбинированной конечной точки (инфаркт миокарда, инсульт, смерть от сердечно-сосудистых причин) от типа гипертрофии миокарда левого желудочка.

В исследование [74, 84] вошло 734 больных гипертонической болезнью (преимущественно пожилого возраста) с гипертонией I-II стадии и 1-3 степени. Наблюдения проводились в течение 3-8 лет. Имело место интервальное цензурирование смешанного типа.

Выборка была разбита на две группы: больные с концентрической гипертрофией (тяжелое течение заболевания) и с другими типами гипертрофии (обычное течение заболевания). Объем выборки в группе с тяжелым течением заболевания составил 232 индивида, а в группе с обычным течением заболевания — 502 индивида.

На рисунках 5.1(a) и 5.1(b) представлены графики эмпирической функции распределения, оцененной теоретической функции распределения и доверительные границы для теоретической функции распределения. Оценки параметров приведены в таблице 5.3.

Группа	$\tilde{\eta}_n$	$\hat{\tau}_n$
Тяжелое течение	1.9	25.4
Обычное течение	0.9	36.0

Таблица 5.3. Оценки параметров модели

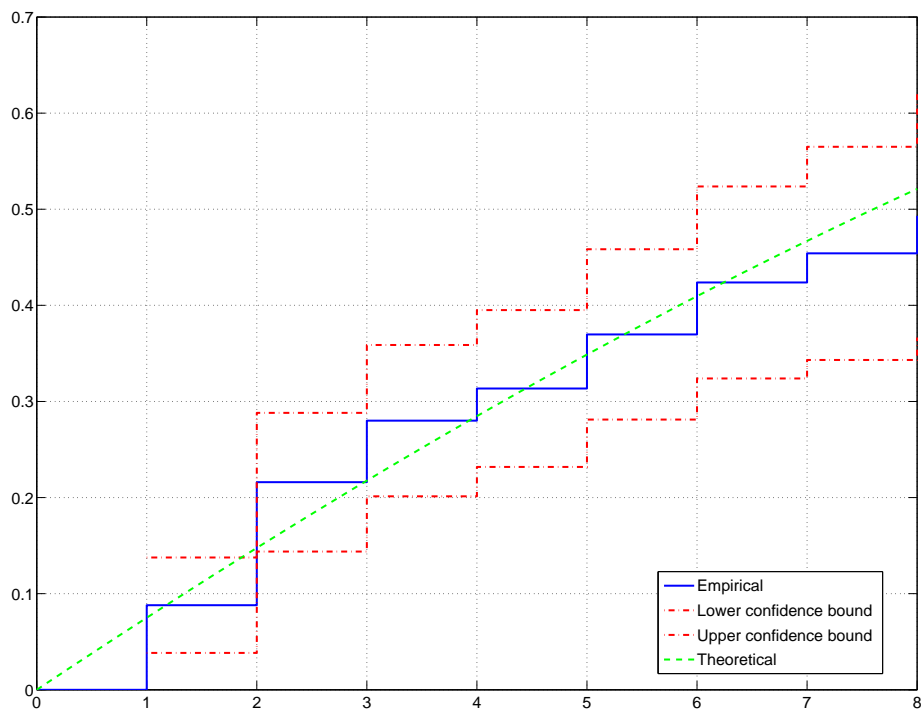
Отличия в значениях параметра η вполне ожидаемы: параметр интерпретируется как «интенсивность» процесса. Более интересны различия в значениях параметра τ : разница в 10-11 лет хорошо согласуется с известными фактами.

5.3. Пример из фармакологии

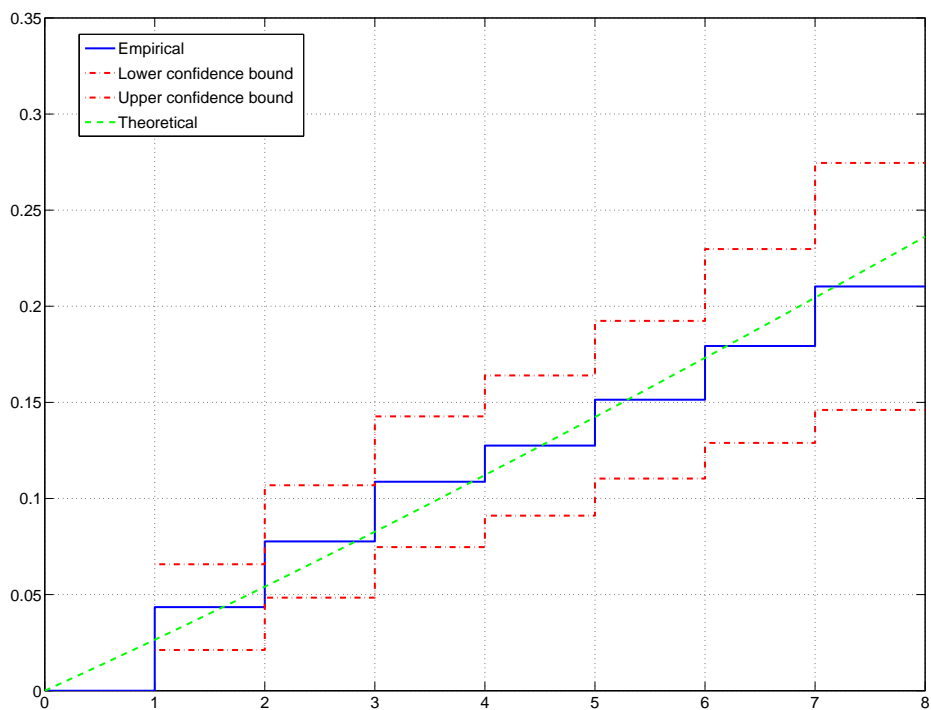
В работе [39] приведены результаты эксперимента по исследованию медикаментозного снижения наркотической зависимости, в котором приняло участие 331 человек с героиновой зависимостью. Все участники получали либо препарат налтрексон, являющийся антагонистом опиатных рецепторов и блокирующий действие наркотиков [56], либо плацебо. Нас будет интересовать задача сравнения этих двух групп с точки зрения времени выбывания пациента из программы. Общая продолжительность наблюдения составила 6 месяцев, на протяжении которых каждый пациент наблюдался в среднем раз в две недели, вследствие этого в процессе сбора данных имело место интервальное цензурирование.

Оценивание параметров для модели будем производить при помощи ОМКЛ в два этапа:

1. Выбор наиболее подходящей параметрической модели при помощи информационного критерия типа Акайке, описанного в главе 3.



(a) Группа с тяжелым течением заболевания



(b) Группа с обычным течением заболевания

Рис. 5.1. Пример из кардиологии. Эмпирическая, теоретическая функции распределения, доверительные границы для теоретической функции распределения. По оси x — годы.

2. Оценивание параметров и построение 95% доверительных интервалов методом профилей правдоподобия, описанным в [68].

В качестве возможной параметрической модели будем рассматривать:

1. Распределение Вейбулла.
2. Гамма-распределение.
3. Степенное гамма-распределение.

В таблице 5.4 приведены значения информационного критерия для двух рассматриваемых групп индивидов. Мы видим, что для группы с плацебо более подходящей является модель с распределением Вейбулла, а для группы, получавшей лечение налтрексоном — степенное гамма-распределение. Конечно, различие между распределением Вейбулла и степенным гамма-распределением незначительны (и, скорее всего, случайны), однако в таком случае имеет смысл предпочесть более простое распределение сложному, что и обуславливает выбор распределения Вейбулла.

	Placebo	Naltrexon
Weibull	-1731.28	-2309.5
Gamma	-1925.71	-1908.61
PowGamma	-1734.31	-1362.84

Таблица 5.4. Значение информационного критерия

Оценки параметров и доверительные интервалы для двух групп представлены в таблице 5.5. В случае степенного гамма-распределения оценки параметров были пересчитаны в естественную параметризацию, и уже в ней строились доверительные интервалы.

Из таблицы 5.5 мы видим, что $\tilde{\beta}_n > 1$, и, следовательно, для группы с плацебо функция риска возрастает со временем. Несложно видеть, что

	Значение оценки	Левая граница	Правая граница
	Плассебо		
$\tilde{\beta}_n$	1.73	1.26	2.2
$\tilde{\lambda}_n$	56.8	42.3	71.3
	Naltrexon		
\tilde{b}_n	1.2	0.8	1.6
\tilde{a}_n	63.8	46.7	81.9
\tilde{k}_n	3.2	2.7	3.7

Таблица 5.5. Оценки параметров и доверительные интервалы для двух групп

0.7-квантиль распределения составляет 60 дней, что достаточно хорошо согласуется с данными из [39]. Очевидно, что группа с налтрексоном существенно отличается от группы с плацебо. Интерпретация параметров в этом случае — задача для экспериментаторов.

Заключение

В диссертационной работе рассмотрена задача построения оценок для специальных (параметрических) моделей кривых дожития в условиях неполных данных. Задачи оценивания такого вида возникают не только в медико-биологических задачах, но и во многих других областях, где требуется изучение «времени до наступления некоторого события», например, при анализе демографических, экономических, финансовых, эпидемиологических показателей, в теории надежности и социологии.

Кратко перечислим основные результаты диссертации.

1. Предложены два класса оценок для параметрических моделей в условиях интервального цензурирования смешанного типа: оценки типа максимального правдоподобия и робастные оценки по минимуму расстояния Кульбака-Лейблера.
2. Получены достаточные условия строгой состоятельности (Теорема 1.3) и асимптотической нормальности (Теорема 1.8) оценок типа максимального правдоподобия для выборки с интервальным цензурированием смешанного типа.
3. Получены достаточные условия строгой состоятельности (Теорема 2.4) и асимптотической нормальности (Теорема 2.8) оценок по минимуму расстояния Кульбака-Лейблера в случае интервального цензурирования смешанного типа.
4. Информационные критерии типа Акайке для сравнения различных параметрических моделей обобщены на случай интервального цензурирования смешанного типа (Глава 3).

5. Разработан программный комплекс, реализующий предложенные способы оценивания параметров.
6. Свойства оценок параметров были проверены при помощи статистического моделирования (Глава 4). Оценки применены для изучения ряда примеров с реальными данными (Глава 5).

Литература

1. *Akaike H.* Information theory and an extension of the maximum likelihood principle // Second International Symposium on Information Theory / Ed. by B. Petrov, B. Csaki. Academiai Kiado: Budapest, 1973. Pp. 267–281.
2. *Alexeyeva N., A. A.* Synonymy of power gamma- distributions in the statistical model of “muscles” // Proc. of 6th 5t. Petersburg Workshop on Simulation / Ed. by S. M. Ermakov, V. B. Melas, A. N. Pepelyshev. 2005. Pp. 39–43.
3. *Andersen P. K., Borgan Ø., Gill R. D., Keiding N.* Statistical Models Based on Counting Processes. Springer, 1993. 784 pp.
4. *Andreev K.* Evolution of the Danish Population from 1835 to 2000. Odense: Odense University Press, 2002.
5. *Asmussen S., Møller J. R.* Risk comparisons of premium rules: optimality and a life insurance study // *Insurance: Mathematics and Economics*. 2003. Vol. 32, no. 3. Pp. 331–344.
6. *Bart A. G., Bart V. A., Steland A., Zaslavskiy M. L.* Modeling disease dynamics and survivor functions by sanogenesis curves // *Journal of Statistical Planning and Inference*. 2005. Vol. 32. Pp. 33–51.
7. Befolkningens bevægelser 1992. Copenhagen: Danmarks Statistik, 1994.
8. *Berk R. H.* Limiting Behavior of Posterior Distributions when the Model is Incorrect // *The Annals of Mathematical Statistics*. 1966. Vol. 37, no. 1. Pp. 51–58.
9. *Bickel P. J., Klaassen C. A. J., Ritov Y., Wellner J. A.* Efficient and Adaptive Estimation for Semiparametric Models. Springer, 1998. 588 pp.

10. *Byrd R. H., Lu P., Nocedal J., Zhu C.* A limited memory algorithm for bound constrained optimization // *SIAM J. Sci. Comput.* 1995. Vol. 16, no. 5. Pp. 1190–1208.
11. *Dempster A., Laird N., Rubin D.* Maximum likelihood data from incomplete data via the EM algorithm // *J. R. Stat. Soc. Ser. B.* 1977. no. 39. Pp. 1–38.
12. *Efron B.* Bootstrap Methods: Another Look at the Jackknife // *The Annals of Statistics.* 1979. Vol. 7, no. 1. Pp. 1–26.
13. *Efron B.* Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods // *Biometrika.* 1981. Vol. 68, no. 3. Pp. 589–599.
14. *Finkelstein D., Wolfe R. A.* Isotonic regression for interval-censored survival data using an E-M algorithm // *Comm. Statist.: Theory & Methods.* 1986. Vol. 15. Pp. 2493–2505.
15. *Fleming T. R., Harrington D. P.* Counting Processes and Survival Analysis. Wiley-Blackwell, 2005. 448 pp.
16. *Gentleman R., Geyer C.* Maximum likelihood for interval censored data: Consistency and computation // *Biometrika.* 1994. Vol. 81. Pp. 618–623.
17. *Gentleman R., Vandal A.* Computational algorithms for censored data problems using intersection graphs // *J. Comput. & Graph. Stat.* 2001. Vol. 10. Pp. 403–421.
18. *Geskus R. B., Groeneboom P.* Asymptotically optimal estimation of smooth functionals for interval censoring, case 2 // *The Annals of Statistics.* 1999. Vol. 27, no. 2. Pp. 627–674.

19. *Geskus R. B., Groeneboorn P.* Asymptotically optimal estimation of smooth functionals for interval censoring, part 1 // *Statistica Neerlandica*. 1996. Vol. 50. Pp. 69–88.
20. *Geskus R. B., Groeneboorn P.* Asymptotically optimal estimation of smooth functionals for interval censoring, part 2 // *Statistica Neerlandica*. 1997. Vol. 51. Pp. 201–219.
21. *Gompertz B.* On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies // *Philosophical Transactions of the Royal Society of London*. 1825. Vol. 115. Pp. 513–583.
22. *Grenander U.* On the theory of mortality measurement. II. // *Skandinavisk Aktuarietidskrift*. 1957. Vol. 1956. Pp. 125–153.
23. *Groeneboom P., Wellner J. A.* Information bounds and nonparametric maximum likelihood estimation. Basel, Switzerland: Birkhäuser Verlag, 1992.
24. *Hjort N. L.* On Inference in Parametric Survival Data Models // *International Statistical Review / Revue Internationale de Statistique*. 1992. Vol. 60, no. 3. Pp. 355–387.
25. *Hoffmann-Jørgensen J.* Probability with a View toward Statistics. New York: Chapman and Hall, 1994. Vol. 2.
26. *Huang J., Wellner J. A.* Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1 // *Statistica Neerlandica*. 1995. Vol. 49. Pp. 153–163.
27. *Huber C., Solev V., Vonta F.* Estimation Of Density For Arbitrarily Censored And Truncated Data // *Probability, Statistics and Modelling in Public*

- Health / Ed. by M. Nikulin, D. Commenges, C. Huber. Springer US, 2006. Pp. 246–265.
28. *Huber C., Solev V., Vonta F.* Interval censored and truncated data: Rate of convergence of NPMLE of the density // *Journal of Statistical Planning and Inference*. 2009. Vol. 139, no. 5. Pp. 1734 – 1749.
 29. *Huber C., Vonta F.* A semiparametric model for interval censored and truncated data // *Probability and Statistics*. 14–1. POMI, Saint Petersburg, 2009. Vol. 363 of *Zapiski Nauchnyh Seminarov POMI im. V.A. Steklova RAS*. Pp. 139–150.
 30. *Huber P. J.* The behavior of Maximum Likelihood Estimates under nonstandard conditions // *Proc. 5th Berkeley Sympos. Math. Statist. Probab.*, Univ. Calif. 1967. Pp. 221–233.
 31. *Jongbloed G.* The iterative convex minorant algorithm for nonparametric estimation // *J. Comput. & Graph. Stat.* 1998. Vol. 7. Pp. 301–321.
 32. *Jordan C. W.* Textbook on Life Contingencies. Chicago: Society of Actuaries, 1967.
 33. *Kalbfleisch J. D., Prentice R. L.* The Statistical Analysis of Failure Time Data. Wiley-InterScience, 2002. 462 pp.
 34. *Kaplan E. L., Meier P.* Nonparametric Estimation from Incomplete Observations // *Journal of the American Statistical Association*. 1958. Vol. 53, no. 282. Pp. 457–481.
 35. *Konishi S., Kitagawa G.* Generalised information criteria in model selection // *Biometrika*. 1996. Vol. 83, no. 4. Pp. 875–890.

36. *Konishi S., Kitagawa G.* Information Criteria and Statistical Modeling. Springer, 2008. 276 pp.
37. *Korobeynikov A.* On the Consistency of ML-estimates for the Special Model of Survival Curves with Incomplete Data // Proc. of 6th St. Petersburg Workshop on Simulation / Ed. by S. M. Ermakov, V. B. Melas, A. N. Pepelyshev. 2009. Pp. 1039–1045.
38. *Kress R.* Linear Integral Equations. Springer, 1999. 388 pp.
39. *Krupitsky E. M., Verbitskaya E. V., Zvartau E. E. et al.* Naltrexone for heroin dependence treatment in St. Petersburg, Russia // *Journal of Substance Abuse Treatment*. 2004. Vol. 26, no. 4. Pp. 285 – 294.
40. *Kullback S., Leibler R. A.* On Information and Sufficiency // *The Annals of Mathematical Statistics*. 1951. Vol. 22, no. 1. Pp. 79–86.
41. *Lawless J. F.* Statistical Models and Methods for Lifetime Data. Wiley-Interscience, 2003. 664 pp.
42. *Le Cam L.* On some asymptotic properties of maximum likelihood estimates and related Bayes estimates // *University of California publications in statistics*. 1953. Vol. 1, no. 11. Pp. 277–239.
43. *Maathuis M.* Reduction algorithm for the NPMLE for the distribution of bivariate interval-censored data // *J. Comput. & Graph. Stat.* 2005. Vol. 14. Pp. 352–362.
44. *Marshall A. W., Olkin I.* Gompertz and Gompertz-Makeham Distributions // Life Distributions. Springer New York, 2007. Pp. 363–398.
45. *McLachlan G., Krishnan T.* The EM Algorithm and Extensions. Wiley, 1997. 274 pp.

46. *Neuhaus G.* On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter // *The Annals of Mathematical Statistics.* 1971. Vol. 42, no. 4. Pp. 1285–1295.
47. *Oakes D.* An Approximate Likelihood Procedure for Censored Data // *Biometrics.* 1986. Vol. 42, no. 1. Pp. 177–182.
48. *Perlman M. D.* On the strong consistency of approximate maximum likelihood estimators // Proc. 6th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. 1972. Pp. 263–281.
49. *Pfanzagl J.* On the measurability and consistency of minimum contrast estimates // *Metrika.* 1969. Vol. 14, no. 1. Pp. 249–272.
50. *Pfanzagl J.* Contributions to a general asymptotic statistical theory (Lecture Notes in statistics). Springer-Verlag, 1982. 315 pp.
51. *Prentice R. L.* A Log Gamma Model and Its Maximum Likelihood Estimation // *Biometrika.* 1974. Vol. 61, no. 3. Pp. 539–544.
52. *Quenouille M. H.* Notes on Bias in Estimation // *Biometrika.* 1956. Vol. 43, no. 3–4. Pp. 353–360.
53. *R Development Core Team.* R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL: <http://www.R-project.org> (дата обращения: 01.06.2010).
54. *Rockafellar R.* Convex analysis. Princeton University Press, 1997. 451 pp.
55. *Schick A., Yu Q.* Consistency Of The GMLE With Mixed Case Interval-Censored Data // *Scand. J. Statist.* 1998. Vol. 27. Pp. 45–55.
56. *Shader R. I.* Antagonists, Inverse Agonists, and Protagonists // *Journal of Clinical Psychopharmacology.* 2003. Vol. 32, no. 4. Pp. 321–322.

57. *Song S.* Estimation with Univariate “Mixed Case” Interval Censored Data // *Statistica Sinica*. 2004. Vol. 14, no. 1. Pp. 269–282.
58. *Stacy E. W.* A Generalization of the Gamma Distribution // *The Annals of Mathematical Statistics*. 1962. Vol. 33, no. 3. Pp. 1187–1192.
59. *Stute W.* Strong consistency of the MLE under random censoring // *Metrika*. 1992. Vol. 39, no. 1. Pp. 257–267.
60. *Stute W., Wang J.-L.* The Strong Law under Random Censorship // *The Annals of Statistics*. 1993. Vol. 21, no. 3. Pp. 1591–1607.
61. *Sun J.* The Statistical Analysis of Interval-censored Failure Time Data (Statistics for Biology and Health). Springer, 2006. 406 pp.
62. *Tierney L., Rossini A. J., Li N., Sevcikova H.* snow: Simple Network of Workstations, 2010. R package version 0.3-3. URL: <http://CRAN.R-project.org/package=snow> (дата обращения: 01.06.2010).
63. *Tukey J.* Bias and confidence in not quite large samples // *The Annals of Mathematical Statistics*. 1958. P. 614.
64. *Turnbull B.* The empirical distribution function with arbitrarily grouped, censored and truncated data // *J. R. Stat.. Soc. Ser. B*. 1976. Vol. 38. Pp. 290–295.
65. *van der Laan M. J.* Efficient and inefficient estimation in semiparametric models. CWI Tracts. 114. Amsterdam: CWI. vi, 219 p. Dfl. 50.00 , 1995.
66. *van der Vaart A. W.* On Differentiable Functionals // *The Annals of Statistics*. 1991. Vol. 19, no. 1. Pp. 178–204.
67. *van der Vaart A. W., Wellner J. A.* Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, 1996.

68. *Venzon D. J., Moolgavkar S. H.* A Method for Computing Profile-Likelihood-Based Confidence Intervals // *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. 1988. Vol. 37, no. 1. Pp. 87–94.
69. *Wald A.* Note on the Consistency of the Maximum Likelihood Estimate // *The Annals of Mathematical Statistics*. 1949. Vol. 20. Pp. 595–601.
70. *White H.* Maximum Likelihood Estimation of Misspecified Models // *Econometrica*. 1982. Vol. 50, no. 1. Pp. 1–25.
71. *Yu H.* Rmpi: Interface (Wrapper) to MPI (Message-Passing Interface), 2010. R package version 0.5-8. URL: <http://CRAN.R-project.org/package=Rmpi> (дата обращения: 01.06.2010).
72. *Zhu C., Byrd R. H., Lu P., Nocedal J.* Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization // *ACM Trans. Math. Softw.* 1997. Vol. 23, no. 4. Pp. 550–560.
73. *Алексеева Н. П., Алексеев А. О., Вербицкая Е. В., Крупицкий Е. М.* Логический принцип проективности и конечные геометрии в статистическом анализе медицинских данных // *Математические модели. Теория и приложения* / Под ред. М. К. Чиркова. 2007. Т. 8. С. 3–14.
74. *Алексеева Н. П., Бондаренко Б. Б., Конради А. О.* Симптомный анализ в исследовании долгосрочного клинического прогноза // *Артериальная Гипертензия*. 2008. Т. 14, № 1. С. 38–43.
75. *Барт А. Г.* Анализ медико-биологических систем. Метод частично-обратных функций. СПб.: Изд-во С.-Петербур. ун-та., 2003. 276 с.
76. *Барт А. Г., Бондаренко Б. Б., Бойко В. И.* Математический анализ течения ХГН // *Гломерулонефрит*. М.: Наука, 1980. С. 213–215.

77. *Барт А. Г., Клочкова (Алексеева) Н. П.* Критические периоды в кривых дожития // Статистические методы в клинических испытаниях / Под ред. А. А. Жиглявского, В. В. Некруткина. СПб.: Изд-во С.-Петерб. ун-та., 1999.
78. *Барт А. Г., Коробейников А. И.* Об оценке параметров специальной модели кривой дожития // Математические модели. Теория и приложения / Под ред. М. К. Чиркова. 2007. Т. 8. С. 15–25.
79. *Ибрагимов И., Хасъминский Р.* Асимптотическая теория оценивания. Москва: Наука, 1979. 527 с.
80. *Калинин О. М.* О единых математических трактовках в биологической систематике и динамике популяций и о связи диффузии с нелинейными уравнениями // *Проблемы кибернетики*. 1972. Т. 25. С. 107–117.
81. *Кокс Д. Р., Оукс Д.* Анализ данных типа времени жизни. Москва: Финансы и статистика, 1988. 192 с.
82. *Коробейников А. И.* Методы и программное обеспечение задач оценивания параметров в специальном случае модели кривых дожития // Математические модели. Теория и приложения / Под ред. М. К. Чиркова. 2009. Т. 10. С. 28–42.
83. *Коробейников А. И.* Сравнение оценок параметров специальной модели кривой дожития для выборки с интервальным цензурированием // *Вестник С.-Петербургского университета, сер. 10*. 2009. Т. 2. С. 36–47.
84. *Коробейников А. И., Бондаренко Б. Б., Алексеева Н. П.* Оценка критических периодов в кривых дожития // *Тез. Всерос. науч.-практ. конференции «Высокотехнологичные методы диагностики и лечения заболеваний сердца и крови»*. Артериальная гипертензия. 2008. Т. 14, № 1. С. 26.

85. *Коробейников А. И., Евсеев С. М., Боярова С. К.* Особенности пространственного поражения тканей пародонта при хроническом генерализованном пародонтите у людей пожилого и старческого возраста // Тез. международного научного конгресса «Современная курортология: проблемы, решения, перспективы». СПб.: Бастион, 2008. С. 40–41.
86. *Коробейников А. И., Мадай Д. Ю., Барт А. Г. и др.* Репаративная эффективность вилона при лечении больных пожилого и старческого возраста с хроническим генерализованным пародонтитом и сопутствующими возрастными соматическими заболеваниями: Тех. доклад. Великий Новгород: Изд-во НовГУ им. Ярослава Мудрого, 2006.
87. *Крамер Г.* Математические методы статистики. Москва: Наука, 1975. 678 с.

Приложение А

Разрешимость информационного уравнения в случае интервального цензурирования смешанного вида

Как было отмечено в главе 2, достаточные условия разрешимости интегрального уравнения вида (2.20) с ядром (2.21) были рассмотрены в работе [19]. В данном разделе мы приведем обобщение теорем из этой работы на случай интервального цензурирования смешанного типа.

Напомним, что рассматривается вопрос разрешимости интегрального уравнения

$$\varphi(x) - \int_{\tau_0}^{\tau_\infty} K_F(x, t) \varphi(t) dt = a(x)$$

с ядром

$$K_F(x, t) = \frac{\bar{d}_F(x) \bar{D}_F(x, t)}{1 + \bar{d}_F(x) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(x, t) dt}, \quad (\text{A.1})$$

где

$$\bar{D}_F(x, t) = \begin{cases} \frac{\bar{h}(t, x)}{F(x) - F(t)}, & t \leq x, \\ \frac{\bar{h}(x, t)}{F(t) - F(x)}, & t \geq x, \end{cases}$$

и правой частью $a(x)$ в виде

$$a(x) = \frac{\tilde{\kappa}'(x) \bar{d}_F(x)}{1 + \bar{d}_F(x) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(x, t) dt}.$$

Теорема А.1. Пусть выполняются условия S1-S5. Тогда интегральный оператор с ядром (2.21) компактен на банаховом пространстве $D([\tau_0, \tau_\infty], \|\cdot\|_\infty)$ càdlàg функций на интервале $[\tau_0, \tau_\infty]$.

Более того, однородное интегральное уравнение

$$\varphi(x) - \int_{\tau_0}^{\tau_\infty} K_F(x, t)\varphi(t) dt = 0$$

имеет только тривиальное решение, и, следовательно, решение уравнение (2.20) существует и притом единственно.

Доказательство. Из условий S1-S5 несложно видеть, что ядро интегрального оператора $K_F(x, t)$ таково, что $K_F(\cdot, t)$ и $K_F(x, \cdot)$ являются ограниченными càdlàg функциями на интервале $[\tau_0, \tau_\infty]$. При этом, множество E скачков конечно (возможно, пусто) и не зависит от x . Поэтому, применяя теорему 2.22 из [38], заключаем, что интегральный оператор с ядром K_F является компактным на пространстве $D([\tau_0, \tau_\infty], \|\cdot\|_\infty)$.

Рассмотрим однородное уравнение

$$\varphi(x) - \int_{\tau_0}^{\tau_\infty} K_F(x, t)\varphi(t) dt = 0.$$

Используя выражение для ядра (A.1), приводим его к виду

$$\varphi(x) \left(1 + 1 + \bar{d}_F(x) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(x, t) dt \right) = \bar{d}_F(x) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(x, t)\varphi(t) dt.$$

Пусть $\varphi(s) = \sup_{x \in [\tau_0, \tau_\infty]} \varphi(x) \geq 0$. Тогда имеем

$$\begin{aligned} \varphi(s) \left(1 + \bar{d}_F(s) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(s, t) dt \right) &= \\ &= \bar{d}_F(s) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(s, t)\varphi(t) dt \leq \varphi(s) \bar{d}_F(s) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(s, t) dt, \end{aligned}$$

откуда имеем $\varphi(s) \leq 0$, и, следовательно, $\varphi(s) = 0$. Аналогично, пусть $\varphi(m) = \inf_{x \in [\tau_0, \tau_\infty]} \varphi(x) \leq 0$. Тогда

$$\begin{aligned} \varphi(m) \left(1 + \bar{d}_F(m) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(m, t) dt \right) &= \\ &= \bar{d}_F(m) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(m, t) \varphi(t) dt \geq \varphi(m) \bar{d}_F(m) \int_{\tau_0}^{\tau_\infty} \bar{D}_F(m, t) dt, \end{aligned}$$

откуда $\varphi(m) = 0$. Таким образом, если функция φ достигает на отрезке $[\tau_0, \tau_\infty]$ своего максимального и минимального значения, то с необходимостью $\varphi = 0$.

Ясно, что функция φ может иметь скачки только в тех же точках, что и ядро $K_F(x, t)$. Между скачками φ непрерывна. Поэтому осталось рассмотреть случай, когда φ достигает экстремума сразу перед точкой скачка, то есть что $\varphi(s) < \varphi(s-) = \sup_{x \in [\tau_0, \tau_\infty]} \varphi(x)$. Предположим, что $\varphi(s-) \geq 0$. Тогда для всякого достаточно малого $\delta > 0$ имеем:

$$\bar{d}_F(s - \delta) \int_{\varphi(t) > \varphi(s - \delta)} \bar{D}_F(s - \delta, t) \varphi(t) dt \leq \frac{1}{2} \varphi(s - \delta).$$

Тогда, аналогично вышеприведенному случаю, когда достигается супремум, получаем, что $\varphi(s - \delta) \leq 0$ для всякого достаточно малого $\delta > 0$. И, по непрерывности φ , имеем $\varphi(s-) = 0$.

Случай $\varphi(m) > \varphi(m-) = \inf_{x \in [\tau_0, \tau_\infty]} \varphi(x)$ рассматривается аналогично. ■

Теорема А.2. Пусть выполняются условия S1-S5 и L1-L2. Кроме этого, пусть плотность f отделена от 0: $f > C > 0$. Тогда информационное уравнение

$$\tilde{\kappa}_F(x) = \left[(A^{(1)})^* A^{(1)} \right] g(x) \quad F - n.в..$$

с операторами $(A^{(1)})^*$ и $A^{(1)}$, определенными в (2.12), (2.13), имеет и притом единственное решение.

Доказательство. Так как плотность f отделена от нуля, то для доказательства теоремы достаточно показать, что функция $\varphi(x)$ является Липшицевой.

Запишем уравнение (2.19):

$$\varphi(x) + \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \bar{h}(t, x) dt - \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) = \tilde{\kappa}'(x) \bar{d}_F(x). \quad (\text{A.2})$$

Для начала заметим, что в условиях теоремы $\bar{d}_F(x) > 0$, и, следовательно,

$$\left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \bar{h}(t, x) dt - \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) = \tilde{\kappa}'(x) - \frac{\varphi(x)}{\bar{d}_F(x)}. \quad (\text{A.3})$$

Продифференцируем уравнение (A.2), переходя, если нужно, к левым и правым производным:

$$\begin{aligned} & \varphi'(x) + \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \bar{h}(t, x) dt - \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) + \\ & + \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi'(x)}{F(x) - F(t)} \bar{h}(t, x) dt + \int_x^{\tau_\infty} \frac{\varphi'(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) - \\ & - \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{(F(x) - F(t))^2} f(x) \bar{h}(t, x) dt + \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{(F(t) - F(x))^2} f(x) \bar{h}(x, t) dt \right) + \\ & + \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \frac{\partial}{\partial x} \bar{h}(t, x) dt - \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \frac{\partial}{\partial x} \bar{h}(x, t) dt \right) = \\ & = \tilde{\kappa}''(x) \bar{d}_F(x) + \tilde{\kappa}'(x) \bar{d}_F'(x). \end{aligned}$$

Перепишем данное выражение относительно $\varphi'(x)$, используя формулу (A.3):

$$\begin{aligned} \varphi'(x) & \left(1 + \int_{\tau_0}^x \frac{\bar{d}_F(x)}{F(x) - F(t)} \bar{h}(t, x) dt + \int_x^{\tau_\infty} \frac{\bar{d}_F(x)}{F(t) - F(x)} \bar{h}(x, t) dt \right) = \\ & = \tilde{\kappa}''(x) \bar{d}_F(x) + \bar{d}'_F(x) \varphi(x) / \bar{d}_F(x) + \\ & + \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{(F(x) - F(t))^2} f(x) \bar{h}(t, x) dt + \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{(F(t) - F(x))^2} f(x) \bar{h}(x, t) dt \right) - \\ & - \bar{d}_F(x) \left(\int_{\tau_0}^x \frac{\varphi(x) - \varphi(t)}{F(x) - F(t)} \frac{\partial}{\partial x} \bar{h}(t, x) dt - \int_x^{\tau_\infty} \frac{\varphi(t) - \varphi(x)}{F(t) - F(x)} \frac{\partial}{\partial x} \bar{h}(x, t) dt \right). \end{aligned}$$

Правая часть полученного равенства в силу условий S1-S5 и L1-L2 равномерно ограничена по x . Для выражения в скобках, очевидно, имеем

$$1 + \int_{\tau_0}^x \frac{\bar{d}_F(x)}{F(x) - F(t)} \bar{h}(t, x) dt + \int_x^{\tau_\infty} \frac{\bar{d}_F(x)}{F(t) - F(x)} \bar{h}(x, t) dt > 0, \quad \forall x \in [\tau_0, \tau_\infty].$$

Следовательно, для производной $\varphi'(x)$ выполняется: $\varphi'(x) < K$, $x \in [\tau_0, \tau_\infty]$.

Отсюда выводим, что

$$|\varphi(x) - \varphi(y)| < C' |y - x| < C'/C |F(x) - F(y)|,$$

и поэтому $d\varphi/dF \in L^2_0(F)$. ■

Приложение Б

Доказательство теоремы 2.3

Теорема 2.3 является ключевой в установлении асимптотических свойств ОМКЛ $\tilde{\theta}_n$. Фактически, она позволяет не только установить асимптотическую эффективность ОМП \hat{F}_n для оценивания функционалов вида $\int c dF$, но и применить теорию эмпирических процессов для исследования величин, являющихся производными от таких функционалов.

Для ее доказательства нам потребуются некоторые вспомогательные обозначения и результаты. Само же доказательство является обобщением аналогичной теоремы из [20] со случая интервального цензурирования второго типа на случай интервального цензурирования смешанного типа.

Б.1. Используемые результаты и обозначения

Скорость сходимости ОМП \hat{F}_n задается следующим утверждением, доказанным в работе [57].

Теорема Е7. *Введем расстояние Хеллинджера $h(q_{F_1,H}, q_{F_2,H})$ между плотностями $q_{F_1,H}$ и $q_{F_2,H}$, заданными формулой (2.11), посредством следующего выражения*

$$h(q_{F_1,H}, q_{F_2,H}) = \left[\frac{1}{2} \int \left(\sqrt{q_{F_1,H}} - \sqrt{q_{F_2,H}} \right)^2 d\nu' \right]^{1/2}.$$

Обозначим через $H_{i,j}$ функцию распределения случайной величины $T_{i,j}$. Тогда в условиях теоремы 2.1 для ОМП \hat{F}_n имеем

$$h(q_{\hat{F}_n,H}, q_{F_0,H}) = O_p \left(n^{-1/3} \right), \quad n \rightarrow \infty,$$

и, как следствие этого,

$$\left\| \hat{F}_n - F_0 \right\|_{H_{i,j}} = O_p \left(n^{-1/3} \right), \quad n \rightarrow \infty.$$

Введем процесс $t \mapsto W_F(t)$ для $t \in [t_0, t_\infty]$ как $W_F(t) = \sum_{i=1}^k p_k W_F^{(k)}(t)$, где

$$\begin{aligned} W_F^{(k)}(t) = & \delta_{k,1} \int_{t_{k,1}=\tau_0}^t \frac{dQ_n}{F(t_{k,1})} + \\ & + \sum_{i=2}^k \delta_{k,i} \left(\int_{t_{k,i}=\tau_0}^t \frac{dQ_n}{F(t_{k,i}) - F(t_{k,i-1})} - \int_{t_{k,i-1}=\tau_0}^t \frac{dQ_n}{F(t_{k,i}) - F(t_{k,i-1})} \right) - \\ & - \delta_{k,k+1} \int_{t_{k,k}=\tau_0}^t \frac{dQ_n}{1 - F(t_{k,k})}. \end{aligned}$$

Критерием ОМП \hat{F}_n служит следующая теорема, являющаяся частным случаем теоремы 28.3 из [54].

Теорема Е8. *Любая функция h , интервалы постоянства которой совпадают с интервалами постоянства J_1, \dots, J_m функции \hat{F}_n , удовлетворяет соотношению*

$$\int_{J_i} h(t) dW_{\hat{F}_n}(t) = 0, \quad i = 1, \dots, m.$$

Нам также понадобятся некоторые результаты из теории эмпирических процессов.

Определение 13. *Класс функций $\mathcal{F} = \{f : B \rightarrow \mathbb{R}\}$, где B — некоторое банахово пространство, называется классом Донскера, если для эмпирического процесса как элемента $L^\infty(\mathcal{F})$ имеет место*

$$\mathbb{G}_n = \sqrt{n} (Q_n - Q) \xrightarrow{d} G, \quad n \rightarrow \infty,$$

где G – плотный измеримый элемент $L^\infty(\mathcal{F})$.

Другими словами, класс \mathcal{F} является донскеровским, если имеет место центральная предельная теорема равномерно на \mathcal{F} .

Введем на \mathcal{F} полуметрику следующим образом:

$$\rho_Q(f, g) = \left(\int \left(f - g - \int (f - g) dQ \right)^2 dQ \right)^{1/2}$$

Определение 14. Эмпирический процесс \mathbb{G}_n называется асимптотически равностепенно непрерывным на \mathcal{F} относительно ρ_Q , если для всякого $\varepsilon > 0$

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbf{P} \left(\sup_{\rho_Q(f, g) < \delta} \left| \int (f - g) d\mathbb{G}_n \right| > \varepsilon \right) = 0.$$

Условие асимптотической равностепенной непрерывности и донскеровости являются почти идентичными, как следует из теоремы 1.5.7 из [67]:

Теорема E9 (van der Vaart-Wellner). Класс функций \mathcal{F} является классом Донскера тогда и только тогда, когда выполняются следующие условия:

1. Эмпирический процесс \mathbb{G}_n является асимптотически равностепенно непрерывным на \mathcal{F} относительно ρ_Q .

2. Класс \mathcal{F} является ограниченным подмножеством $L^2(Q)$.

Определение 15. Секционной вариационной нормой [46] функции $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ назовем следующую величину

$$\|f\|_v^* = \max \left\{ \|f\|_\infty, \|f\|_v, \sup_x \|f(x, \cdot)\|_v, \sup_x \|f(\cdot, x)\|_v \right\},$$

где $\|\cdot\|_v$ означает обычную вариационную норму.

Замечание. Секционная вариационная норма для функции $f : \mathbb{R}^l \rightarrow \mathbb{R}$ вводится аналогично.

Нам потребуется следующая теорема из [65], связывающая донскеровость и равномерную ограниченность по вариации

Теорема E10. *Рассмотрим класс функций*

$$\mathcal{F}_M = \{f \in \mathcal{D} : \|f\|_V^* \leq M\},$$

где \mathcal{D} — множество всех càdlàg функций из \mathbb{R}^2 в \mathbb{R} .

Тогда \mathcal{F}_M является классом Донскера равномерно относительно всех распределений на \mathbb{R}^2 .

Б.2. Ход доказательства

Обозначим через \mathcal{F} класс всех функций распределения, удовлетворяющих условию S3.

В лемме Б.1 будем показано, что для всякой $F \in \mathcal{F}$ выполняется

$$\int \tilde{\kappa}_{F_0} d(F - F_0) = - \int \tilde{\xi}_F dQ_{F_0}.$$

Обозначим

$$\psi_F(x) = \frac{\varphi_F(x)}{F(x)(1 - F(x))}.$$

Для функции $\psi_F(x)$ можно выписать интегральное уравнение, аналогичное уравнению (2.19). При этом, в условиях теоремы 2.1 это уравнение будет иметь единственное решение. Аналог теоремы 2.2 для функции $\psi_F(x)$ доказывается так же, как и теорема 2.2 (см. приложение А).

Для $x \in J_i$, $i = 1, \dots, m$ определим кусочно-постоянную функцию $\bar{\psi}_{\hat{F}_n}$ следующим образом:

1. $\bar{\psi}_{\hat{F}_n}(x) = \psi_{\hat{F}_n}(x_0)$, если найдется $x_0 \in J_i$, что $\hat{F}_n(x_0) = F_0(x_0)$;
2. $\bar{\psi}_{\hat{F}_n}(x) = \psi_{\hat{F}_n}(\tau_i)$, если $F_0(x) > \hat{F}_n(\tau_i)$, $\forall x \in J_i$;

3. $\bar{\psi}_{\hat{F}_n}(x) = \psi_{\hat{F}_n}(\tau_{i+1} - 0)$, если $F_0(x) < \hat{F}_n(\tau_i)$, $\forall x \in J_i$.

Положим $\bar{\varphi}_{\hat{F}_n}(x) = \hat{F}_n(x) \left(1 - \hat{F}_n(x)\right) \bar{\psi}_{\hat{F}_n}$ и определим $\bar{\xi}_{\hat{F}_n}(x)$ аналогично формуле (2.26):

$$\begin{aligned} \bar{\xi}_{\hat{F}_n}(\delta_k, t_k, k) &= \\ &= -\frac{\delta_{k,1} \bar{\varphi}_{\hat{F}_n}(t_{k,1})}{\hat{F}_n(t_{k,1})} - \sum_{i=2}^k \frac{\delta_{k,i} \left(\bar{\varphi}_{\hat{F}_n}(t_{k,i}) - \bar{\varphi}_{\hat{F}_n}(t_{k,i-1})\right)}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} + \frac{\delta_{k,k+1} \bar{\varphi}_{\hat{F}_n}(t_{k,k})}{1 - \hat{F}_n(t_{k,k})}. \end{aligned} \quad (\text{B.1})$$

Тогда, из теоремы Е8 следует, что

$$\int \bar{\xi}_{\hat{F}_n} dQ_n = 0,$$

откуда выводим

$$-\sqrt{n} \int \tilde{\xi}_{\hat{F}_n} dQ_{F_0} = \sqrt{n} \int \bar{\xi}_{\hat{F}_n} d(Q_n - Q_{F_0}) + \sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n}\right) dQ_{F_0}. \quad (\text{B.2})$$

В лемме Б.2 доказывається, что $\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n}\right) dQ_{F_0} = o_p(1)$, $n \rightarrow \infty$.

Далее, для первого слагаемого в (B.2) имеем

$$\sqrt{n} \int \bar{\xi}_{\hat{F}_n} d(Q_n - Q_{F_0}) = \sqrt{n} \int \tilde{\xi}_{\hat{F}_n} d(Q_n - Q_{F_0}) + \sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n}\right) d(Q_n - Q_{F_0}).$$

Наконец, из леммы Б.3 следует, что

$$\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n}\right) d(Q_n - Q_{F_0}) = o_p(1), \quad n \rightarrow \infty.$$

Б.3. Шаги доказательства

Лемма Б.1. Для всякой $F \in \mathcal{F}$ имеет место соотношение

$$\int \tilde{\kappa}_{F_0} d(F - F_0) = - \int \tilde{\xi}_F dQ_{F_0}.$$

Доказательство. Для $F \in \mathcal{F}$ рассмотрим оператор $A_F^{(1)} : L^2(F) \rightarrow L^2(Q_F)$, действующий по правилу:

$$\left(A_F^{(1)} g \right) (\delta_k, t_k, k) = \sum_{i=1}^{k+1} \frac{\delta_{k,i} \int_{t_{k,i-1}}^{t_{k,i}} g(x) dF(x)}{F(t_{k,i}) - F(t_{k,i-1})}$$

Рассмотрим константную функцию $1(x) = 1 \in L^2(F)$. Несложно видеть, что оператор $A_F^{(1)}$ переводит эту функцию в единицу на пространстве $L^2(Q_F)$.

Далее заметим, что из доказательства теоремы 2.2 фактически следует, что $\tilde{\xi}_F \in L^2(Q_{F_0})$. В таком случае имеем

$$\int \tilde{\xi}_F dQ_{F_0} = \int \tilde{\xi}_F \cdot 1 dQ_{F_0} = \int \tilde{\xi}_F \cdot A_{F_0}^{(1)}(1) dQ_{F_0} = \int \left(A_{F_0}^{(1)} \right)^* \tilde{\xi}_F dF_0. \quad (\text{Б.3})$$

Рассмотрим уравнение (2.18) с заменой $\tilde{\xi}_{F_0}$ на $\tilde{\xi}_F$. Проинтегрируем левую и правую часть уравнение. Тогда получаем:

$$\left(A^{(1)} \right)^* \tilde{\xi}_F(x) = \tilde{\kappa}_{F_0}(x) + C, \quad x \in [\tau_0, \tau_\infty].$$

Вычислим константу C :

$$C = \int C dF = \int \left(A^{(1)} \right)^* \tilde{\xi}_F dF - \int \tilde{\kappa}_{F_0} dF.$$

Несложно видеть, что $\tilde{\xi}_F \in L_0^2(Q_F)$. Тогда очевидно, что

$$\int \left(A^{(1)} \right)^* \tilde{\xi}_F dF = \int \left(A^{(1)} \right)^* \tilde{\xi}_F \cdot 1 dF = \int \tilde{\xi}_F \cdot A^{(1)} 1 dQ_F = \int \tilde{\xi}_F \cdot 1, dQ_F = 0.$$

Таким образом, получаем

$$\left(A_{F_0}^{(1)} \right)^* \tilde{\xi}_F = \tilde{\kappa}_{F_0} - \int \tilde{\kappa}_{F_0} dF.$$

Подставляя это выражение в формулу (Б.3) выводим

$$\int \tilde{\kappa}_{F_0} d(F - F_0) = - \int \tilde{\xi}_F dQ_{F_0}.$$

■

Лемма Б.2.

$$\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right) dQ_{F_0} = o_p(1), \quad n \rightarrow \infty.$$

Доказательство. Рассмотрим функцию $\vartheta(t_k, k)$, определенную следующим образом:

$$\begin{aligned} \vartheta(t_k, k) = & - \left[\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right] (\delta_k^{(1)}, t_k, k) F_0(t_{k,1}) - \\ & - \sum_{i=2}^k \left[\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right] (\delta_k^{(i)}, t_k, k) (F_0(t_{k,i}) - F_0(t_{k,i-1})) + \\ & + \left[\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right] (\delta_k^{(k+1)}, t_k, k) (1 - F_0(t_{k,k})), \quad (\text{Б.4}) \end{aligned}$$

где $\delta_{k,j}^{(i)} = 1$, если $i = j$ и 0 иначе.

Подставим $\bar{\varphi}$, определенную формулой (Б.1), в (Б.4)

$$\begin{aligned} \vartheta(t_k, k) = & \frac{\bar{\varphi}_{\hat{F}_n}(t_{k,1}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,1})}{\hat{F}_n(t_{k,1})} F_0(t_{k,1}) + \\ & + \sum_{i=2}^k \frac{\left(\bar{\varphi}_{\hat{F}_n}(t_{k,i}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,i}) \right)}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} (F_0(t_{k,i}) - F_0(t_{k,i-1})) - \\ & - \sum_{i=2}^k \frac{\left(\bar{\varphi}_{\hat{F}_n}(t_{k,i-1}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,i-1}) \right)}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} (F_0(t_{k,i}) - F_0(t_{k,i-1})) + \\ & + \frac{\bar{\varphi}_{\hat{F}_n}(t_{k,k}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,k})}{1 - \hat{F}_n(t_{k,k})} (1 - F_0(t_{k,k})). \end{aligned}$$

Воспользуемся представлением $\bar{\varphi}_F = F(1-F)\bar{\psi}_F$, отщепим по одному слагаемому от сумм посредине и перегруппируем слагаемые. Получаем:

$$\begin{aligned}
\vartheta(t_k, k) &= \left(\bar{\psi}_{\hat{F}_n}(t_{k,1}) - \tilde{\psi}_{\hat{F}_n}(t_{k,1}) \right) \left(1 - \hat{F}_n(t_{k,1}) \right) \times \\
&\quad \times \left[F_0(t_{k,1}) - \hat{F}_n(t_{k,1}) \frac{F_0(t_{k,2}) - F_0(t_{k,1})}{\hat{F}_n(t_{k,2}) - \hat{F}_n(t_{k,1})} \right] + \\
&\quad + \sum_{i=2}^{k-1} \frac{\left(\bar{\varphi}_{\hat{F}_n}(t_{k,i}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,i}) \right)}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} (F_0(t_{k,i}) - F_0(t_{k,i-1})) - \\
&\quad - \sum_{i=3}^k \frac{\left(\bar{\varphi}_{\hat{F}_n}(t_{k,i-1}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,i-1}) \right)}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} (F_0(t_{k,i}) - F_0(t_{k,i-1})) + \\
&\quad + \left(\bar{\psi}_{\hat{F}_n}(t_{k,k}) - \tilde{\psi}_{\hat{F}_n}(t_{k,k}) \right) \hat{F}_n(t_{k,k}) \times \\
&\quad \times \left[1 - F_0(t_{k,k}) + (1 - \hat{F}_n(t_{k,k})) \frac{F_0(t_{k,k}) - F_0(t_{k,k-1})}{\hat{F}_n(t_{k,k}) - \hat{F}_n(t_{k,k-1})} \right] = \\
&\quad = \left(\bar{\psi}_{\hat{F}_n}(t_{k,1}) - \tilde{\psi}_{\hat{F}_n}(t_{k,1}) \right) \left(1 - \hat{F}_n(t_{k,1}) \right) \times \\
&\quad \times \left[F_0(t_{k,1}) - \hat{F}_n(t_{k,1}) \frac{F_0(t_{k,2}) - F_0(t_{k,1})}{\hat{F}_n(t_{k,2}) - \hat{F}_n(t_{k,1})} \right] + \\
&\quad + \sum_{i=2}^{k-1} \left(\bar{\varphi}_{\hat{F}_n}(t_{k,i}) - \tilde{\varphi}_{\hat{F}_n}(t_{k,i}) \right) \left[\frac{F_0(t_{k,i}) - F_0(t_{k,i-1})}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} - \frac{F_0(t_{k,i+1}) - F_0(t_{k,i})}{\hat{F}_n(t_{k,i+1}) - \hat{F}_n(t_{k,i})} \right] + \\
&\quad + \left(\bar{\psi}_{\hat{F}_n}(t_{k,k}) - \tilde{\psi}_{\hat{F}_n}(t_{k,k}) \right) \hat{F}_n(t_{k,k}) \times \\
&\quad \times \left[1 - F_0(t_{k,k}) + (1 - \hat{F}_n(t_{k,k})) \frac{F_0(t_{k,k}) - F_0(t_{k,k-1})}{\hat{F}_n(t_{k,k}) - \hat{F}_n(t_{k,k-1})} \right].
\end{aligned}$$

Преобразуем выражения в квадратных скобках:

$$\begin{aligned}
F_0(t_{k,1}) - \hat{F}_n(t_{k,1}) \frac{F_0(t_{k,2}) - F_0(t_{k,1})}{\hat{F}_n(t_{k,2}) - \hat{F}_n(t_{k,1})} &= \\
&= \frac{F_0(t_{k,1}) \left(\hat{F}_n(t_{k,2}) - F_0(t_{k,2}) \right) + F_0(t_{k,2}) \left(F_0(t_{k,1}) - \hat{F}_n(t_{k,1}) \right)}{\hat{F}_n(t_{k,2}) - \hat{F}_n(t_{k,1})}.
\end{aligned}$$

$$\begin{aligned}
& 1 - F_0(t_{k,k}) + (1 - \hat{F}_n(t_{k,k})) \frac{F_0(t_{k,k}) - F_0(t_{k,k-1})}{\hat{F}_n(t_{k,k}) - \hat{F}_n(t_{k,k-1})} = \\
& = \frac{1}{\hat{F}_n(t_{k,k}) - \hat{F}_n(t_{k,k-1})} \left((1 - F_0(t_{k,k-1})) \left(\hat{F}_n(t_{k,k}) - F_0(t_{k,k}) \right) + \right. \\
& \quad \left. + (1 - F_0(t_{k,k})) \left(F_0(t_{k,k-1}) - \hat{F}_n(t_{k,k-1}) \right) \right).
\end{aligned}$$

$$\begin{aligned}
& \frac{F_0(t_{k,i}) - F_0(t_{k,i-1})}{\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1})} - \frac{F_0(t_{k,i+1}) - F_0(t_{k,i})}{\hat{F}_n(t_{k,i+1}) - \hat{F}_n(t_{k,i})} = \\
& = \frac{1}{\left(\hat{F}_n(t_{k,i}) - \hat{F}_n(t_{k,i-1}) \right) \left(\hat{F}_n(t_{k,i+1}) - \hat{F}_n(t_{k,i}) \right)} \times \\
& \quad \times \left(F_0(t_{k,i}) \left(\hat{F}_n(t_{k,i+1}) - F_0(t_{k,i+1}) \right) + F_0(t_{k,i+1}) \left(F_0(t_{k,i}) - \hat{F}_n(t_{k,i}) \right) \right) + \\
& \quad + F_0(t_{k,i-1}) \left(\hat{F}_n(t_{k,i}) - F_0(t_{k,i}) \right) + F_0(t_{k,i}) \left(F_0(t_{k,i-1}) - \hat{F}_n(t_{k,i-1}) \right) + \\
& \quad + F_0(t_{k,i+1}) \left(\hat{F}_n(t_{k,i-1}) - F_0(t_{k,i-1}) \right) + F_0(t_{k,i-1}) \left(F_0(t_{k,i+1}) - \hat{F}_n(t_{k,i+1}) \right) \Big)
\end{aligned}$$

Применяя неравенство Коши-Буняковского, получаем

$$\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right) dQ_{F_0} \leq \sqrt{n} C \sum_{k=1}^{\infty} p_k \zeta_k, \quad (\text{Б.5})$$

где

$$\begin{aligned}
\zeta_k = & \left\| \bar{\psi}_{\hat{F}_n} - \tilde{\psi}_{\hat{F}_n} \right\|_{H_{k,1}} \left(\left\| \hat{F}_n - F_0 \right\|_{H_{k,1}} + \left\| \hat{F}_n - F_0 \right\|_{H_{k,2}} \right) + \\
& + \left\| \bar{\psi}_{\hat{F}_n} - \tilde{\psi}_{\hat{F}_n} \right\|_{H_{k,k}} \left(\left\| \hat{F}_n - F_0 \right\|_{H_{k,k-1}} + \left\| \hat{F}_n - F_0 \right\|_{H_{k,k}} \right) + \\
& + \sum_{i=2}^{k-1} \left\| \bar{\psi}_{\hat{F}_n} - \tilde{\psi}_{\hat{F}_n} \right\|_{H_{k,i}} \left(\left\| \hat{F}_n - F_0 \right\|_{H_{k,i-1}} + \left\| \hat{F}_n - F_0 \right\|_{H_{k,i}} + \left\| \hat{F}_n - F_0 \right\|_{H_{k,i+1}} \right).
\end{aligned}$$

Рассмотрим интервал постоянства J_i функции \hat{F}_n . Рассмотрим три случая:

1. Найдется $s \in J_i$: $\hat{F}_n(s) = F_0(s)$. Тогда, применяя теорему 2.2, для $x \in J_i$ имеем:

$$\begin{aligned} \left| \bar{\psi}_{\hat{F}_n}(x) - \tilde{\psi}_{\hat{F}_n}(x) \right| &= \left| \tilde{\psi}_{\hat{F}_n}(s) - \tilde{\psi}_{\hat{F}_n}(x) \right| \leq \\ &\leq C_1 |s - x| \leq (C_1/C_2) |F_0(s) - F_0(x)| = \\ &= (C_1/C_2) \left| \hat{F}_n(s) - F_0(x) \right| = (C_1/C_2) \left| \hat{F}_n(x) - F_0(x) \right|. \end{aligned}$$

2. Пусть $F_0(x) > \hat{F}_n(\tau_i)$, $\forall x \in J_i$. Тогда

$$\begin{aligned} \left| \bar{\psi}_{\hat{F}_n}(x) - \tilde{\psi}_{\hat{F}_n}(x) \right| &= \left| \tilde{\psi}_{\hat{F}_n}(\tau_i) - \tilde{\psi}_{\hat{F}_n}(x) \right| \leq \\ &\leq C_1 |\tau_i - x| \leq (C_1/C_2) |F_0(\tau_i) - F_0(x)| = (C_1/C_2) (F_0(x) - F_0(\tau_i)) < \\ &< (C_1/C_2) (F_0(x) - \hat{F}_n(\tau_i)) = (C_1/C_2) \left| \hat{F}_n(\tau_i) - F_0(x) \right| = \\ &= (C_1/C_2) \left| \hat{F}_n(x) - F_0(x) \right|. \end{aligned}$$

3. Наконец, пусть $F_0(x) < \hat{F}_n(\tau_i)$, $\forall x \in J_i$. Аналогично получаем

$$\begin{aligned} \left| \bar{\psi}_{\hat{F}_n}(x) - \tilde{\psi}_{\hat{F}_n}(x) \right| &= \left| \tilde{\psi}_{\hat{F}_n}(\tau_{i+1} - 0) - \tilde{\psi}_{\hat{F}_n}(x) \right| \leq \\ &\leq C_1 |\tau_{i+1} - 0 - x| \leq (C_1/C_2) |F_0(\tau_{i+1} - 0) - F_0(x)| < \\ &< (C_1/C_2) \left| \hat{F}_n(\tau_{i+1} - 0) - F_0(x) \right| = (C_1/C_2) \left| \hat{F}_n(x) - F_0(x) \right|. \end{aligned}$$

Так как $\mathbf{EK} < \infty$, то применение теоремы E7 к формуле (Б.5) завершает доказательство. ■

Лемма Б.3.

$$\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right) d(\mathbb{Q}_n - Q_{F_0}) = o_p(1), \quad n \rightarrow \infty.$$

Доказательство. Рассмотрим класс функций

$$\mathcal{K} = \{ \bar{\xi}_F, F \in \mathcal{F} \} \cup \{ \tilde{\xi}_{F_0} \}.$$

Фактически лемма заключается в доказательстве равностепенной непрерывности процесса \mathbb{G}_n на классе \mathcal{K} . Для этого в силу теоремы Е9 достаточно показать, что \mathcal{K} является классом Донскера. Мы покажем, что \mathcal{K} ограничен по секционной вариационной норме, и воспользуемся теоремой Е10.

Заметим, что вследствие вида функции $\bar{\xi}_F$ (см. формулу (Б.1)) достаточно ограничиться двумерной секционной вариационной нормой (для каждого из слагаемых, входящих в $\bar{\xi}_F$). Более того, слагаемые с $\delta_{k,1} = 1$ и $\delta_{k,k+1} = 1$ по сути являются одномерными и поэтому, например, достаточно рассмотреть вариацию функции

$$t \mapsto -(1 - F(t))\bar{\psi}_F(t).$$

Из теоремы 2.1 для функции $\bar{\psi}_F$ имеем

$$|\bar{\psi}_F(y) - \bar{\psi}_F(x)| \leq C(|y - x| + |F(y) - F(x)|),$$

откуда следует ограниченность по вариации $\bar{\psi}_F$ равномерно по $F \in \mathcal{F}$.

Для $\delta_{k,j} = 1, 1 < j \leq k$ рассмотрим

$$f(x, y) = \frac{\bar{\xi}_F(y) - \bar{\xi}_F(x)}{F(y) - F(x)}.$$

Заметим, что функция f постоянна на прямоугольниках вида $[\tau_i, \tau_{i+1}) \times [\tau_i, \tau_{i+1})$. Обозначим $R_{ij} = (s, x] \times (t, y]$. Тогда $f(R_{ij})$ может быть представлена в виде

$$f(R_{ij}) = \frac{A_1 + A_2 - A_3}{D},$$

где

$$A_1 = (F(t) - F(x))(F(t) - F(s))(F(x) - F(s))(\bar{\xi}_F(y) - \bar{\xi}_F(t)),$$

$$A_2 = (F(y) - F(x))(F(t) - F(x))(F(y) - F(t))(\bar{\xi}_F(x) - \bar{\xi}_F(s)),$$

$$A_3 = (F(y) + F(t) - F(x) - F(s))(\bar{\xi}_F(t) - \bar{\xi}_F(x)) \times \\ \times (F(y) - F(t))(F(x) - F(s)),$$

$$D = (F(y) - F(x))(F(y) - F(s))(F(t) - F(x))(F(t) - F(s)).$$

Несложно видеть, что знаменатель D отделен от нуля. Снова применяя теорему 2.1, получаем

$$|\bar{\xi}_F(y) - \bar{\xi}_F(x)| \leq C(|y - x| + |F(y) - F(x)|),$$

откуда следует ограниченность $\|f\|_{V_2} = \sup_{R_{ij}} \sum_{ij} f(R_{ij})$.

Для одномерной вариации при $x < t < y$ имеем

$$\begin{aligned} f(x, y) - f(x, t) &= \\ &= \frac{(F(y) - F(x)) (\bar{\xi}_F(y) - \bar{\xi}_F(t)) - (F(y) - F(t)) (\bar{\xi}_F(y) - \bar{\xi}_F(x))}{(F(y) - F(x)) (F(t) - F(x))}, \end{aligned}$$

откуда следует ограниченность $\sup_x \|f(x, \cdot)\|_v$. Аналогичным образом следует ограниченность по вариации по второй переменной. Ограниченность по вариации функции $\tilde{\xi}_{F_0}$ показывается таким же образом.

Применяя теоремы E9 и E10, выводим свойство Донскера класса \mathcal{K} и, как следствие этого, равностепенную непрерывность процесса \mathbb{Q}_n на \mathcal{K} .

Далее, имеем

$$\left| \bar{\psi}_{\hat{F}_n}(t) - \tilde{\psi}_{F_0} \right| \leq C \left| \hat{F}_n(t) - F_0(t) \right|$$

и

$$\left\| \tilde{\psi}_{\hat{F}_n} - \tilde{\psi}_{F_0} \right\|_{L^2(Q)} = o_P(1), \quad n \rightarrow \infty.$$

Откуда следует, что

$$\left\| \bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{F_0} \right\|_{Q_{F_0}} = o_P(1), \quad n \rightarrow \infty,$$

и, наконец, ввиду равностепенной непрерывности процесса \mathbb{Q}_n на \mathcal{K}

$$\sqrt{n} \int \left(\bar{\xi}_{\hat{F}_n} - \tilde{\xi}_{\hat{F}_n} \right) d(\mathbb{Q}_n - Q_{F_0}) = o_p(1), \quad n \rightarrow \infty.$$

■