Санкт-Петербургский государственный университет

Прикладная математика и информатика

Вычислительная стохастика и статистические модели

Коврыга Валерия Валерьевна

Задачи оценивания параметров модели IRT

Бакалаврская работа

Научный руководитель:

к.ф.-м.н., доцент А. И. Коробейников

Рецензент:

ассистент А. Ю. Шлемов

Санкт-Петербург

2016

Saint Petersburg State University
Applied Mathematics and Computer Science
Computational Stochastics and Statistical Models

Kovryga Veleriia

Problems of parameter estimation in IRT model

Bachelor's Thesis

Scientific Supervisor:
Associate Professor A. I. Korobeynikov

Reviewer:
Assistant Professor A. Yu. Shlemov

Saint Petersburg
2016

Оглавление

Введение

В психологии, образовании и других прикладных научных областях часто возникает необходимость измерить некоторые «скрытые» человеческие качества, то есть качества, не характеризующиеся специфическим видом деятельности. Например, в психологии и психиатрии может стоять задача определить предрасположенность человека к психическим заболеваниям, в образовании может потребоваться определить уровень знания иностранного языка или другого предмета и т.д. Существует общепринятый термин для обозначения таких качеств — *способность* (англ. ability).

Несмотря на то, что способности не поддаются непосредственному измерению, их можно оценивать. Существует специальный подход — Item Response Theory (IRT), одной из рассматриваемых задач которого является оценивание способностей людей. В рамках этого подхода для оценивания способностей люди отвечают на вопросы, из полученных ответов делается вывод о способностях отвечавших людей. Каждому вопросу сопоставляется понятие *сложность*. Существуют различные математические модели, которые позволяют оценить вероятность правильного ответа человека на вопрос в зависимости от способности человека и сложности вопроса.

Примером использования подхода IRT может послужить тестирование СПбГУ по английскому языку, целью которого является проверить, обладают ли студенты уровнем знания языка B2 или нет. Студентам предлагается ответить на вопросы разной сложности. Ответы на вопросы позволяют оценить способности студентов.

В данной работе рассматриваются некоторые из возможных решений задачи оценивания способностей людей и сложностей вопросов, проблемы, возникающие по ходу решения данной задачи, способы их устранения, а также некоторые естественные усложнения этой задачи.

В главе 1 рассматривается простейшая вероятностная модель (Rasch model, [1]), в которой способности человека и сложности вопроса сопоставляются числовые параметры. Вводится дискретная случайная величина, которая принимает значения $\{0, 1\}$ в зависимости от ответа человека на вопрос. В рамках данной модели ставится задача оценивания неизвестных параметров и обсуждается главная проблема, состоящая в том, что с увеличением объема выборки растет число неизвестных параметров.

Один из возможных способов решения проблемы состоит в том, чтобы считать параметры способностей людей мешающими и оценивать только сложности вопросов. Оказывается, для этих мешающих параметров существуют достаточные статистики, а условная функция правдоподобия не зависит от параметров способностей людей. Это позволяет избежать проблемы оценивания мешающих параметров, и оценки сложностей вопросов строятся как точка максимума функции условного правдоподобия. Данный метод построения оценок называется методом Conditional Maximum Likelihood (CML) [2]. Несмотря на то, что метод CML позволяет решить проблему оценивания мешающих параметров, численная реализация метода оказывается трудоемкой, и происходит накопление ошибок. Для того чтобы уменьшить трудоемкость и вычислительные ошибки, при реализации метода CML были использованы рекуррентные соотношения, предложенные в работе [2].

Известно [3], что CML-оценки асимптотически несмещенные. Однако, для состоятельности построенных оценок параметры сложностей и способностей должны удовлетворять некоторым ограничениям, которые описываются в работе [4]. Эти ограничения накладывают дополнительные условия на рассматриваемую модель.

Для того, чтобы продемонстрировать устойчивость реализованного алгоритма оценивания, были проверены некоторые свойства CML-оценок. Для фиксированных значений параметров способностей и сложностей были промоделированы результаты тестирования, по которым были построены выборки CML-оценок параметров сложностей вопросов, и было исследовано поведение оценок при увеличении объема выборки.

В главе 3 на примере тестирования СПбГУ по английскому языку рассматривается естественное расширение задачи оценивания сложностей вопросов, а именно ситуация, когда составляется несколько вариантов тестирования, и каждому студенту предлагается ответить на вопросы только одного теста. В случае, когда тесты независимы, и для каждого теста оценки сложностей вопросов строятся отдельно, построенные оценки будут несравнимы между собой. Один из возможных способов решения этой проблемы состоит в том, чтобы рассматривать тесты с «общими» вопросами, и строить для них оценки методом, аналогичным методу CML. В главе 3 исследуется поведение построенных оценок при изменяющемся числе «общих» вопросов, числе вопросов теста и увеличивающемся объеме выборки.

Глава 1

Описание модели. Постановка задачи

Для оценивания способностей людей и сложностей вопросов необходимо ввести математическую модель, которая позволит вычислять вероятность правильного ответа человека на вопрос, в зависимости от сложности вопроса и способности человека. Рассмотрим простейшую модель, которая удовлетворяет условиям:

1. ответы на вопросы тестирования могут быть только «Да» и «Нет» (для удобства будем считать, что на каждый вопрос существует только «правильный» и «неправильный» ответы);
2. способность человека — это число, которое не зависит от сложности вопроса;
3. вопросы не связаны друг с другом, т. е. правильный ответ на один вопрос не дает подсказки к правильному ответу на любой другой вопрос;
4. вероятность правильного ответа зависит только от способности человека и сложности вопроса;
5. чем выше способность, тем больше вероятность правильного ответа;
6. чем сложнее вопрос, тем меньше вероятность правильного ответа.

1.1. Rasch Model

В 1960 году Георгом Рашем была предложена вероятностная модель, описанная в статье [1], для которой выполняются условия 1 – 6. Способности человека и сложности вопроса сопоставляются числовые параметры θ и β соответственно, и рассматривается бернуллиевская случайная величина ξ, которая принимает значение 1, если человек со значением параметра способности θ правильно ответил на вопрос сложности β. Распределение случайной величины ξ задается функцией

$$\mathsf{P}(\xi = 1) = \frac{\exp(\theta - \beta)}{1 + \exp(\theta - \beta)}, \tag{1.1}$$

причем из работы [1] известно, что при введении определенных дополнительных ограничений функция (1.1) единственная, с точностью до перепараметризации, которая удовлетворяет всему набору условий 1 – 6.

1.2. Вероятностная модель. Оценивание неизвестных параметров

В случае, когда необходимо оценить способности n человек и сложности k вопросов, которым соответствуют параметры $\theta_1, \ldots, \theta_n$ и $\beta_1, \ldots, \beta_k$ соответственно, рассматривается набор совместно независимых бернуллиевских случайных величин

$$x_{ij}, \quad i = 1 \ldots n, \quad j = 1 \ldots k. \tag{1.2}$$

Случайная величина $x_{ij} = 1$, если человек со способностью θ_i правильно ответил на вопрос сложности β_j, и вероятность правильного ответа вычисляется по формуле

$$\mathsf{P}\left(x_{ij} = 1\right) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}.$$

Требуется оценить значения неизвестных параметров β_j и θ_i.

Проблема рассматриваемой задачи заключается в том, что при увеличении объема выборки n растет число неизвестных параметров θ_i, вследствие чего оценки, получаемые с помощью стандартных методов оценивания параметров, оказываются несостоятельными. Один из способов решения этой проблемы — оценить β_j, затем по оценкам $\widehat{\beta}_j$ и известной функции правдоподобия оценить θ_i. Таким образом, первичный интерес задачи представляет оценка β_j при мешающих параметрах θ_i, число которых растет с увеличением объема выборки n.

1.3. CML

Одним из методов оценивания, с помощью которых можно построить оценки сложности вопросов $\widehat{\beta}_j$ без вычисления оценок значений мешающих параметров θ_i, является метод Conditional Maximum Likelihood (CML), описанный в работе [2]. Данный метод использует тот факт, что для мешающих параметров θ_i существуют достаточные статистики. Это позволяет избежать проблемы оценивания значений параметров способностей людей θ_i.

Рассмотрим выборку x_{ij} из (1.2) как независимую выборку случайных векторов $X_i = (x_{i1}, \ldots, x_{ik})$ из $\mathrm{X} = \{0, 1\}^k$. Для одного наблюдения ($n = 1$, $\theta_1 = \theta$) и выборки $X = (x_1, \ldots, x_k)$ рассмотрим функцию $\mathrm{r} = \mathrm{r}(X) = \sum_{j=1}^{k} x_j$ — количество правильных ответов человека на вопросы. Заметим, что

$$\mathsf{P}\left(\mathrm{r}=\mathrm{r}_0\right)=\sum_{y|\mathrm{r}_0}\prod_{j=1}^{k}\frac{\exp\left((\theta-\beta_j)y_j\right)}{1+\exp(\theta-\beta_j)}=\frac{\exp(\theta\cdot\mathrm{r}_0)\sum_{y|\mathrm{r}_0}\prod_{j=1}^{k}\exp(-\beta_j y_j)}{\prod_{j=1}^{k}(1+\exp(\theta-\beta_j))},$$

где $y|\mathrm{r}_0$ означает суммирование по всем $y=(y_1,\dots,y_k)\in\{0,1\}^k$ таким, что $\sum_{j=1}^{k}y_j=\mathrm{r}_0$. Тогда по формуле полной вероятности функция правдоподобия

$$\mathbf{L}(X,\theta,\boldsymbol{\beta})=\prod_{j=1}^{k}\frac{\exp\left((\theta-\beta_j)x_j\right)}{1+\exp(\theta-\beta_j)}$$

раскладывается в произведение

$$\mathbf{L}(X,\theta,\boldsymbol{\beta})=\mathbf{L}(X,\theta,\boldsymbol{\beta}\mid\mathrm{r}=\mathrm{r}_0)\cdot\mathsf{P}(\mathrm{r}=\mathrm{r}_0),$$

то есть

$$\mathbf{L}(X,\theta,\boldsymbol{\beta})=\frac{\exp\left(-\sum_{j=1}^{k}\beta_j x_j\right)}{\sum_{y|\mathrm{r}}\exp\left(-\sum_{j=1}^{k}\beta_j y_j\right)}\cdot\frac{\exp\left(\theta\sum_{j=1}^{k}x_j\right)\cdot\sum_{y|\mathrm{r}}\exp\left(-\sum_{j=1}^{k}\beta_j y_j\right)}{\prod_{j=1}^{k}\left(1+\exp(\theta-\beta_j)\right)},$$

и

$$\mathbf{L}(X,\theta,\boldsymbol{\beta}\mid\mathrm{r}=\mathrm{r}_0)=\frac{\exp\left(-\sum_{j=1}^{k}\beta_j x_j\right)}{\sum_{y|\mathrm{r}_0}\exp\left(-\sum_{j=1}^{k}\beta_j y_j\right)}.\tag{1.3}$$

Видно, что функция (1.3) не зависит от параметра θ, что означает, что $\mathrm{r}(X)=\sum_{j=1}^{k}x_j$ является достаточной статистикой для параметра θ. Введем вектор оценок $\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}=(\widehat{\beta}_1,\dots,\widehat{\beta}_k)$ как точку максимума функции (1.3),

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}=\operatorname*{argmax}_{\boldsymbol{\beta}}\frac{\exp\left(-\sum_{j=1}^{k}\beta_j x_j\right)}{\sum_{y|\mathrm{r}_0}\exp\left(-\sum_{j=1}^{k}\beta_j y_j\right)}.$$

Тогда для выборки объема n оценки будут вычисляться по формуле

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}=\operatorname*{argmax}_{\boldsymbol{\beta}}\prod_{i=1}^{n}\frac{\exp\left(-\sum_{j=1}^{k}\beta_j x_{ij}\right)}{\sum_{y|\mathrm{r}_i}\exp\left(-\sum_{j=1}^{k}\beta_j y_j\right)},\tag{1.4}$$

где $\mathrm{r}_i=\sum_{j=1}^{k}x_{ij}$.

Введем обозначение

$$\gamma_{\mathrm{r}}(\boldsymbol{\beta})\coloneqq\sum_{y|\mathrm{r}}\exp\Big(-\sum_{j=1}^{k}\beta_j y_j\Big)$$

, где суммирование ведется по всем $y=(y_1,\dots,y_k)\in\{0,1\}^k$ таким, что $\sum_{j=1}^{k}y_j=\mathrm{r}$. При таком обозначении формула (1.4) примет вид

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}=\operatorname*{argmax}_{\boldsymbol{\beta}}\prod_{i=1}^{n}\frac{\exp\left(-\sum_{j=1}^{k}\beta_j x_{ij}\right)}{\gamma_{\mathrm{r}_i}(\boldsymbol{\beta})}.$$

Заметим, что в данной задаче поиска точки максимума (1.4) решение не единственное, так как при сдвиге $\beta_j + c$, где $c = const$, точка максимума функции (1.3) не изменится:

$$\underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, \mathbf{L}(X, \boldsymbol{\theta}, \boldsymbol{\beta} + \boldsymbol{c} \mid \mathbf{r}) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{\exp\big(-\sum_{j=1}^{k}(\beta_j + c)x_{ij}\big)}{\sum_{y|\mathrm{r_i}} \exp\big(-\sum_{j=1}^{k}(\beta_j + c)y_j\big)} =$$
$$= \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{\exp(-c \cdot \mathrm{r_i}) \exp\big(-\sum_{j=1}^{k} \beta_j x_{ij}\big)}{\sum_{y|\mathrm{r_i}} \exp(-c \cdot \mathrm{r}_i) \exp\big(-\sum_{j=1}^{k} \beta_j y_j\big)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, \mathbf{L}(X, \boldsymbol{\theta}, \boldsymbol{\beta} \mid \mathbf{r}).$$

Для того чтобы идентифицировать набор параметров сложности вопросов, достаточно ввести линейное ограничение на значения параметров β_j, например,

$$\sum_{j=1}^{k} \beta_j = 0. \tag{1.5}$$

Существует еще одно условие на параметры рассматриваемой вероятностной модели. В работе [4] были сформулированы условия состоятельности CML-оценок.

Утверждение 1. *Для любого набора $\theta_i \in \mathbb{R}$, $i = 1, \ldots, +\infty$ такого, что*

$$\sum_{i=1}^{+\infty} \exp(-\theta_i) = \infty \tag{1.6}$$

оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}$ существуют и состоятельны.

Далее вероятностная модель рассматривается с учетом введенных ограничений (1.5), (1.6) на значения параметров θ_i, β_j.

1.4. Сложности вычисления CML-оценок

При реализации метода CML возникает проблема вычисления функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$, значения которой находятся по формуле

$$\gamma_{\mathbf{r}}(\boldsymbol{\beta}) = \sum_{y|r} \prod_{j=1}^{k} \exp(-\beta_j y_j), \tag{1.7}$$

где суммирование ведется по всем возможным комбинациям $y_j \in \{0, 1\}$ таким, что $\sum_{j=1}^{k} y_j = \mathrm{r}$.

Так при увеличении числа параметров β_j растет число слагаемых в формуле (1.7), которое равно C_n^k. Более того, при большом числе параметров β_j значение $\prod_{j=1}^{k} \exp(-\beta_j y_j)$ может быть очень мало, что приводит к потери точности. Например, если

$$\beta_2 = \ldots = \beta_k = -1, \;\; \beta_1 = k - 1 \;\; \text{и} \;\; y_1 = 0,$$

то при больших k возможно

$$\prod_{j=1}^{k} \exp(-\beta_j y_j) = \prod_{j=2}^{k} \exp(-y_j) \ll 1.$$

Помимо этого, многие методы численного нахождения точки максимума функции требуют вычисления значений $\mathrm{grad}\gamma_{\mathbf{r}}(\boldsymbol{\beta})$.

Таким образом, при вычислении γ_{r} по формуле (1.7) алгоритм имеет большую трудоемкость вычислений и происходит накопление ошибок.

1.5. Вычисление функции γ

В работе [2] были предложены рекуррентные соотношения, позволяющие вычислять значения функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$ и ее частных производных с меньшей трудоемкостью и большей точностью.

Обозначим $\exp(\beta_i)$ через δ_i, а значение функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$, вычисленной по вектору

$$\boldsymbol{\delta}^{(\boldsymbol{p})} = (\delta_1, \ldots, \delta_{p-1}, \delta_{p+1}, \ldots, \delta_k),$$

обозначим как $\gamma_{\mathbf{r}}^{(-p)}$. Тогда формула (1.7) примет вид

$$\gamma_{\mathbf{r}}(\delta_1, \ldots, \delta_k) = \sum_{y|\mathbf{r}} \prod_{j=1}^{k} \delta_j^{y_j}. \tag{1.8}$$

Согласно [2], имеет место следующее рекуррентное соотношение:

$$\gamma_{\mathbf{r}}(\delta_1, \ldots, \delta_k) = \gamma_{\mathbf{r}}^{(-p)}(\delta_1, \ldots, \delta_{p-1}, \delta_{p+1}, \ldots, \delta_k) + \delta_p \gamma_{\mathbf{r-1}}^{(-p)}(\delta_1, \ldots, \delta_k), \tag{1.9}$$

где $p = 1, \ldots, k$.

При вычислении $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$ по рекуррентным соотношениям (1.9) при $\mathrm{r} = k$ трудоемкость вычислений равна $O(k)$, при $0 < \mathrm{r} < k$ трудоемкость равна $O(\mathrm{r} \cdot k)$, и при $\mathrm{r} = 0$, соответственно, $O(1)$, тогда как вычисление значений функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$ по формуле (1.8) имеет экспоненциальную трудоемкость, так как при суммировании осуществляется перебор C_n^k комбинаций $y_j \in \{0, 1\}$. Следовательно, вычисление функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$ по рекуррентным соотношениям (1.9) уменьшает трудоемкость вычислений.

Глава 2

Проверка свойств CML-оценок.

Вычислительные процедуры, подобные процедуре метода CML, описанной в работе [2], зачастую корректны в теоретическом изложении, однако, по причине сложности алгоритма и используемых рекурсивных соотношений, они могут быть не применимы на практике из-за возможного накопления ошибок и потери точности. Для того, чтобы проверить устойчивость предложенного в статье [2] алгоритма и убедиться в правильности его реализации, потому как в оригинальной статье изложение осуществлено конспективно и часть переходов опущена, были проверены свойства CML-оценок, а именно состоятельность и асимптотическая несмещенность.

Для фиксированных значений параметров θ_i и β_j, удовлетворяющих условию (1.5), были промоделированы данные, по которым были построены выборки оценок параметров сложности вопросов $\widehat{\boldsymbol{\beta}}^{(d)} = (\widehat{\beta}_1^{(d)}, \dots, \widehat{\beta}_k^{(d)})$, $d \in \{1, \dots, 100\}$. Были построены оценки среднего $\mathbb{E}\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}$ и среднеквадратического отклонения $\mathbb{E}\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\mathrm{CML}}\|^2$ при количестве вопросов $k = 2$ и $k = 5$ и увеличивающемся объеме выборки n.

2.1. Доверительные интервалы

Было произведено моделирование для числа вопросов $k = 5$ и фиксированных значений $\beta_1 = -0.4$, $\beta_2 = -0.2$, $\beta_3 = 0.00$, $\beta_4 = 0.2$, $\beta_5 = 0.4$.

На рисунке 2.1 изображены 95% доверительные интервалы для истинных значений параметров сложности вопросов β_j при объеме выборки $n = 500, 1000, 1500, \dots, 10000$. Из рисунка видно, что с ростом объема выборки n истинные значения параметров β_j попадают в построенные 95% доверительные интервалы и близки к их середине.

2.2. Состоятельность и асимптотическая несмещенность

Для проверки свойств состоятельности и асимптотической несмещенности для фиксированного числа параметров сложности $k = 5$ и фиксированных значений параметров $\beta_1 = -0.4$, $\beta_2 = -0.2$, $\beta_3 = 0.00$, $\beta_4 = 0.2$, $\beta_5 = 0.4$ была построена выборка CML-оценок и вычислены оценки среднеквадратического отклонения для увеличивающегося объема

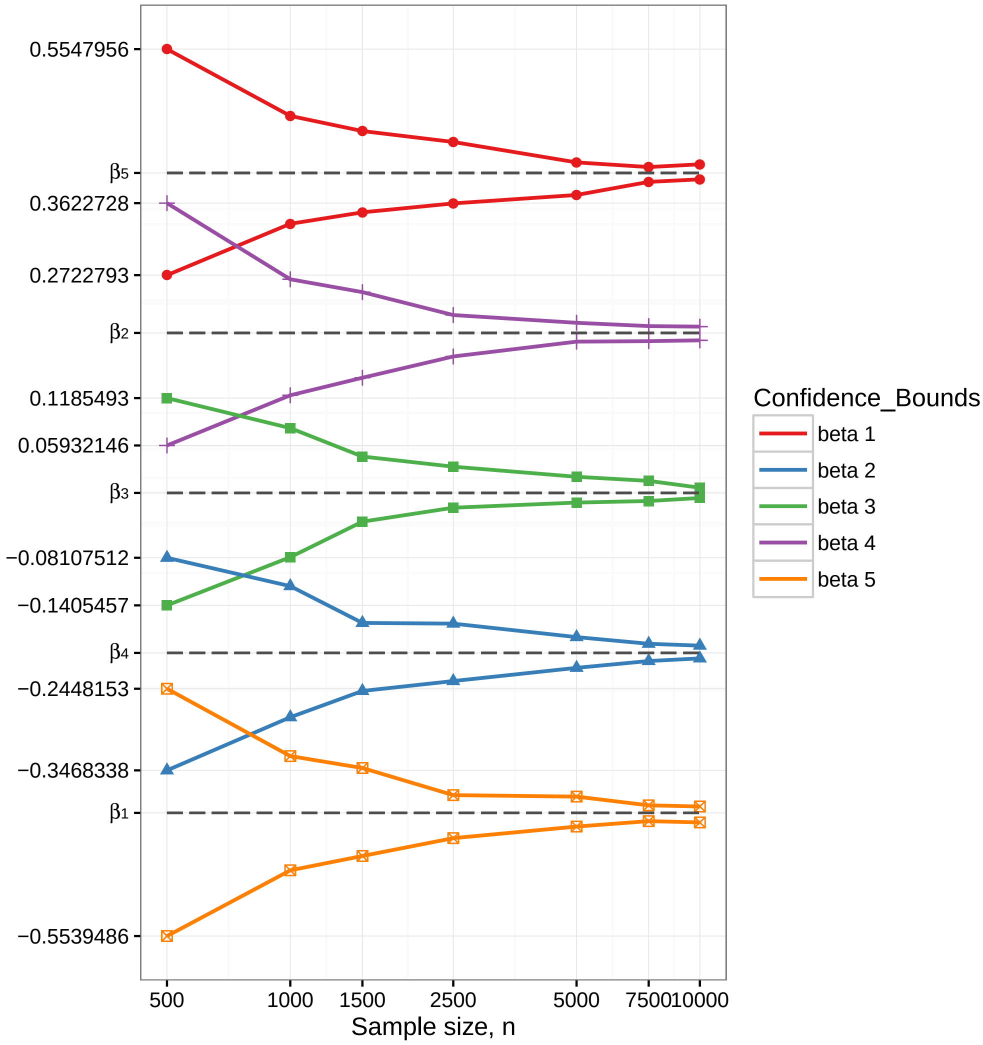

Рис. 2.1. 95% доверительные интервалы для значений $\beta_1, \ldots, \beta_5$.

выборки. По рисунку 2.2, на котором изображена зависимость оценок среднеквадратичного отклонения от объема выборки, можно видеть, что с ростом объема выборки n среднеквадратическое отклонение убывает к 0.

2.3. Сравнение CML-оценок и оценок, полученных одним из стандартных методов оценивания

Ранее, в главе 1, упоминалось, что проблема рассматриваемой задачи заключается в том, что число мешающих параметров увеличивается с ростом объема выборки. По этой причине оценки, построенные с помощью стандартных методов оценивания, оказываются несостоятельными. Для того чтобы проиллюстрировать недостатки стандартных методов и преимущества оценивания методом CML был реализован один из

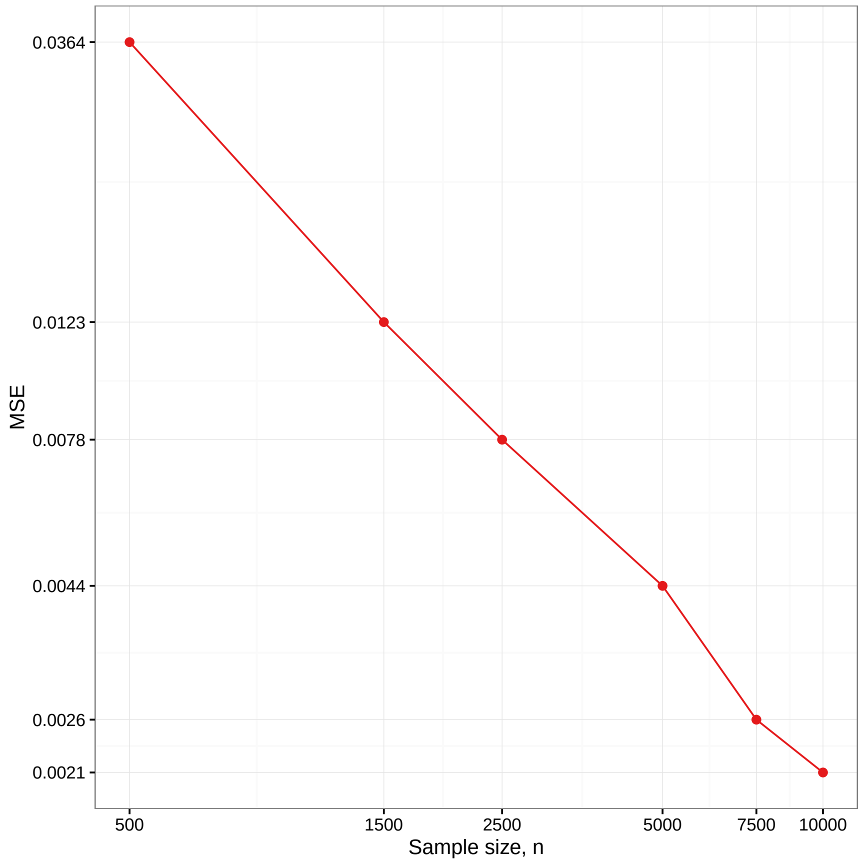

Рис. 2.2. Зависимость оценки $\mathbb{E}\|\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{cml}\|^2$ от объема выборки n.

стандартных методов оценивания, а именно метод максимального правдоподобия (Joint Maximum Likelihood, JML), [1]. Идея метода JML заключается в том, что рассматривается функция правдоподобия $\mathrm{L}(x_{ij}, \theta, \beta)$ для выборки (1.2), записываемая как

$$\mathbf{L}(x_{ij}, \theta, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{k} \frac{\exp((\theta_i + \beta_j)x_{ij})}{1 + \exp(\theta_i + \beta_j)}, \tag{2.1}$$

и оценки $\widehat{\boldsymbol{\theta}}_{\text{JML}}$, $\widehat{\boldsymbol{\beta}}_{\text{JML}}$ вычисляются как точка максимума функции правдоподобия

$$(\widehat{\boldsymbol{\theta}}_{\text{JML}}, \widehat{\boldsymbol{\beta}}_{\text{JML}}) = \underset{\boldsymbol{\theta},\boldsymbol{\beta}}{\operatorname{argmax}}\, \mathbf{L}(x_{ij}, \boldsymbol{\theta}, \boldsymbol{\beta}). \tag{2.2}$$

Потому как функция правдоподобия (2.1) инвариантна относительно сдвига параметров θ_i, β_j на константу, то есть

$$\mathrm{L}(x_{ij}, \boldsymbol{\theta} + \boldsymbol{c}, \boldsymbol{\beta} - \boldsymbol{c}) = \prod_{i=1}^{n} \prod_{j=1}^{k} \frac{(\exp(\theta_i + c + \beta_j - c)x_{ij})}{1 + \exp(\theta_i + c + \beta_j - c)} = \mathrm{L}(x_{ij}, \boldsymbol{\theta}, \boldsymbol{\beta}),$$

решение задачи поиска точки максимума (2.2) не единственное. Для того чтобы из множества возможных решений выбрать одно, вводится линейное ограничение на значения параметров β_j, а именно

$$\sum_{j=1}^{k} \beta_j = 0.$$

Таким образом, для фиксированного числа k и фиксированных значений параметров сложности вопросов β_j были промоделированные данные, по которым были построены выборки JML- и CML-оценок и было осуществлено сравнение поведения построенных оценок при увеличивающемся объеме выборки.

На рисунке 2.3 изображена зависимость 95% доверительных интервалов истинных значений параметров β_j, построенных по выборкам CML- и JML-оценок, от объема выборки n. Из рисунка видно, что значения параметров β_j не попадают в доверительные интервалы, построенные по JML-оценкам, и, наоборот, доверительные интервалы, построенные по выборкам CML-оценок, содержат истинные значения параметров сложности вопросов.

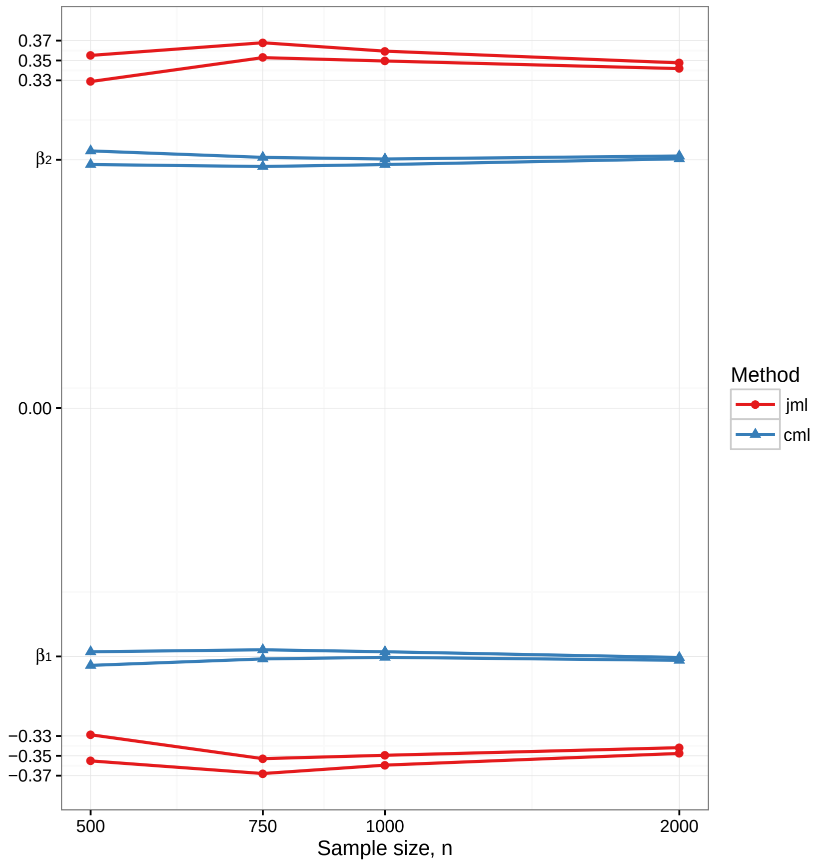

Рис. 2.3. Доверительные интервалы с уровнем доверия 95% для параметров β_1, β_2, построенные по JML- и CML-оценкам.

Из рисунка 2.4 видно, что оценка среднеквадратического отклонения для CML-оценок расположена к 0 ближе, чем оценка среднеквадратичного отклонения для JML-оценок.

Таким образом, видно, что для данной модели метод оценивания значений параметров сложности вопросов CML является более подходящим чем метод максимального правдоподобия.

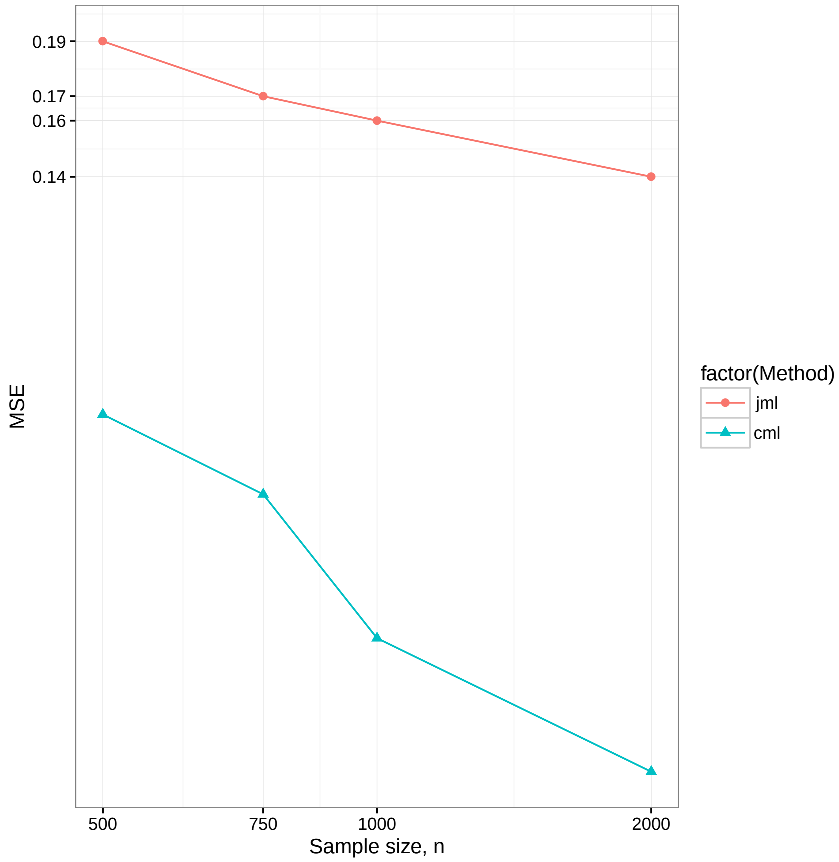

Рис. 2.4. Сравнение оценок среднеквадратического отклонения для JML- и CML-оценок параметров β_1, β_2.

Глава 3

Случай неполного дизайна

3.1. Постановка задачи

Вернемся к примеру тестирования СПбГУ по английскому языку. Для определения уровня владения языком студентам предлагается ответить на вопросы разной сложности. Чтобы исключить списывание составляется несколько вариантов тестирования. Каждый студент отвечает на вопросы только одного варианта. Иными словами, непересекающиеся группы студентов пишут независимые тесты. В таком случае возникает необходимость уметь сравнивать результаты тестирований среди студентов всех групп, и, что более важно, необходимо проверить одинаковую сложность различных вариантов тестирования, то есть убедиться в отсутствии вариантов, более простых или более сложных чем все остальные.

Несмотря на то, что логичным представляется оценивать способности студентов и сложности вопросов для каждого теста по отдельности, такие оценки параметров оказываются несравнимыми между собой. Это происходит, потому что функция условного правдоподобия из формулы (1.3) инвариантна относительно сдвига β_j на $const$, и, следовательно, решение задачи поиска точки максимума (1.4) не единственное. Для того чтобы выделить одно из решений, вводилось линейное ограничение на значения параметров β_j (1.5). Однако, значения параметров сложности вопросов для реальных данных могут не удовлетворять этому условию, и, значит, сравнение оценок, построенных для независимых тестов по отдельности, некорректно.

Один из способов получения сравнимых оценок заключается в том, чтобы рассматривать тесты с некоторым количеством «общих» вопросов. Рассмотрим простейший случай: пусть есть два теста, в которых, соответственно, по k_1 и k_2 уникальных вопросов, то есть вопросов, которые присутствуют только в одном тесте, и k_3 совпадающих вопросов. Рассмотрим эти тесты как один тест с $k = k_1 + k_2 + k_3$ независимыми вопросами, в котором половина студентов (для каждого теста одинаковое количество студентов) ответила на вопросы только первого теста, другая половина — на вопросы только второго теста. «Общими» вопросами назовем пересекающиеся вопросы, число этих вопросов было обозначено как k_3, то есть те вопросы, которые присутствуют в обоих тестах, и на которые отвечали все студенты.

В работе [3] был предложен метод построения оценок значений параметров сложности вопросов для случая неполного дизайна, аналогичный методу CML. Однако, неизвестно, как ведут себя такие оценки при изменяющемся числе вопросов k и разном числе «общих вопросов» k_3. Требуется исследовать поведение CML-оценок для случая неполного дизайна при увеличивающемся числе параметров способностей θ_i и сложностей β_j и меняющемся числе «общих» вопросов.

3.2. Оценивание неизвестных параметров, CML

В качестве метода оценивания параметров сложности вопросов для случая неполного дизайна был выбран метод, предложенный в статье [3], аналогичный методу CML, описанному в главе 1.

Как и прежде, для случая одного наблюдения ($n = 1$, $\theta_1 = \theta$) и выборки $X = (x_1, \ldots, x_k)$ рассматривается статистика $\mathrm{r}(X) = \sum_{j=1}^k b_j x_j$, для которой

$$\mathsf{P}\left(\mathrm{r} = \mathrm{r}_0\right) = \sum_{y|\mathrm{r}_0} \prod_{j=1}^k \frac{\exp\left((\theta - \beta_j) b_j y_j\right)}{(1 + \exp(\theta - \beta_j))^{b_j}} = \frac{\exp(\theta \cdot \mathrm{r}_0) \sum_{y|\mathrm{r}_0} \prod_{j=1}^k \exp(-b_j \beta_j y_j)}{\prod_{j=1}^k (1 + \exp(\theta - \beta_j))^{b_j}},$$

где b_j является индикатором ответа студентом на j-й вопрос, то есть

$$b_j \coloneqq \begin{cases} 1, & \text{если есть ответ на } j\text{-й вопрос}, \\ 0, & \text{если ответа нет}, \end{cases} \tag{3.1}$$

и суммирование ведется по всем $\mathbf{y} = (y_1, \ldots, y_k) \in \{0, 1\}^k$ таким, что $\sum_{j=1}^k y_j b_j = \mathrm{r}_0$. Тогда по формуле полной вероятности совместное правдоподобие для выборки $X = (x_1, \ldots, x_k)$

$$\mathbf{L}(X, \theta, \boldsymbol{\beta}) = \frac{\exp\left(-\sum_{j=1}^k \beta_j x_j b_j\right) \exp\left(\theta \sum_{j=1}^k x_j b_j\right)}{\prod_{j=1}^k \left(1 + \exp(\theta - \beta_j)\right)^{b_j}},$$

запишется в виде

$$\mathbf{L}(X, \theta, \boldsymbol{\beta}) = \frac{\exp\left(-\sum_{j=1}^k \beta_j x_j b_j\right)}{\sum_{y|\mathrm{r}_0} \prod_{j=1}^k \exp(-b_j \beta_j y_j)} \cdot \frac{\exp\left(\theta \sum_{j=1}^k x_j b_j\right) \cdot \sum_{y|\mathrm{r}_0} \prod_{j=1}^k \exp(-b_j \beta_j y_j)}{\prod_{j=1}^k \left(1 + \exp(\theta - \beta_j)\right)^{b_j}},$$

и

$$\mathbf{L}(X, \theta, \boldsymbol{\beta} | \mathrm{r} = \mathrm{r}_0) = \frac{\exp\left(-\sum_{j=1}^k \beta_j x_j b_j\right)}{\gamma_{\mathrm{r}_0}(\boldsymbol{\beta})}, \tag{3.2}$$

где

$$\gamma_{\mathrm{r}_0}(\boldsymbol{\beta}) \coloneqq \sum_{y|\mathrm{r}_0} \prod_{j=1}^k \exp(-b_j \beta_j y_j),$$

и суммирование ведется по всем $\mathbf{y} = (y_1, \dots, y_k) \in \{0,1\}^k$ таким, что $\sum_{j=1}^{k} y_j b_j = \mathrm{r}_0$.

Заметим, что функция $\mathbf{L}(X, \theta, \boldsymbol{\beta} | \mathrm{r} = \mathrm{r}_0)$ не зависит от θ_i. Тогда CML-оценка находится как точка максимума функции (3.2), то есть

$$\hat{\boldsymbol{\beta}}_{\mathrm{CML}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \frac{\exp\left(-\sum_{j=1}^{k} \beta_j x_j b_j\right)}{\gamma_{\mathrm{r}_0}(\boldsymbol{\beta})}.$$

Соответственно, для объема выборки $n \geqslant 1$ CML-оценки параметров сложностей вопросов вычисляются по формуле

$$\hat{\boldsymbol{\beta}}_{\mathrm{CML}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{\exp\left(-\sum_{j=1}^{k} \beta_j x_{ij} b_{ij}\right)}{\gamma_{\mathrm{r}_i}(\boldsymbol{\beta})}.$$

Таким образом, отличие реализации метода оценивания CML для случая неполного дизайна от метода CML, описанного в главе 1, состоит в том, что если $b_{ij} = 0$, соответствующий множитель функции правдоподобия (3.2), записанной для $n \geqslant 1$ наблюдений, обращается в 1, то есть

$$\frac{\exp\left(-\beta_j x_{ij} b_{ij}\right) \exp\left(\theta_i x_{ij} b_{ij}\right)}{\left(1 + \exp(\theta_i - \beta_j)\right)^{b_{ij}}} = \frac{1}{\left(1 + \exp(\theta_i - \beta_j)\right)^{0}} = 1.$$

При этом оценки значений параметров сложностей вопросов вычисляются, как и в случае полного дизайна, как точка максимума функции условного правдоподобия для неполного дизайна (3.2).

3.3. Свойства CML-оценок для случая неполного дизайна

Потому как реализация метода оценивания CML для случая неполного дизайна аналогична методу CML для полного дизайна, логично предположить, что получаемые оценки будут обладать теми же свойствами, то есть будут асимптотически несмещенными и состоятельными. Для проверки этих свойств был рассмотрен случай двух вариантов тестирования с совпадающим числом вопросов в обоих тестах, при условии, что каждый вариант тестирования проходит одинаковое число студентов, при этом каждый студент отвечает на вопросы только одного теста. Были промоделированы выборки CML-оценок для разного числа вопросов k, «общих» вопросов r и числа параметров способностей n и исследовано поведение построенных оценок при увеличивающемся объеме выборки n.

3.3.1. Моделирование. Число вопросов $k = 9$

Была рассмотрена ситуация, когда каждый из тестов состоит из $k_i = 5$, $i = 1, 2$ вопросов, и число «общих» вопросов $r = 1$. Согласно рассматриваемому методу решения задачи оценивания параметров сложности вопросов для случая неполного дизайна, вместо двух рассматривался один общий тест из $k = 9$ вопросов. Параметры сложности этих вопросов обозначим как $\beta_1, \dots, \beta_9$.

Были зафиксированы значения параметров β_j, $j = 1, \dots, 9$, указанные в таблице 3.1, и «общий» вопрос с параметром сложности β_7. Вопросы, со значениями па-

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
-1.05	-0.85	-0.50	0.60	1.80	-0.05	0.00	0.01	0.04

Таблица 3.1. Значения параметров сложности вопросов β_j, $k = 9$.

раметров сложности $\beta_1, \dots, \beta_4, \beta_7$, составили список вопросов первого теста, вопросы сложности $\beta_5, \beta_6, \dots, \beta_9$ — второго.

Для объемов выборки $n = 500, \dots, 2000$ были промоделированы результаты тестирования x_{ij}, по которым были построены выборки CML-оценок. Из рисунка 3.1, на котором изображена зависимость среднеквадратичного отклонения от объема выборки n, можно видеть, что среднеквадратичное отклонение убывает к нулю. Можно приблизительно оценить порядок сходимости как $\sqrt{n}$. Из рисунка видно, что среднеквадратичное отклонение для параметра «общего» вопроса β_7 меньше, чем для других параметров, что закономерно, так как объем выборки для параметра β_7 равен n, тогда как для параметров сложности других вопросов он равен $n/2$, потому что «общий» вопрос встречается в обоих тестах. При моделировании также было проверено, что смещение оценок мало и не превышает 0.01.

При увеличении числа «общих» вопросов $r = 3$, вопросы, соответствующие параметрам $\beta_7, \beta_4, \beta_5$ были зафиксированы как «общие». По рисунку 3.2, на котором изображена зависимость среднеквадратичного отклонения от объемов выборки $n = 500, 750, \dots, 2000$, видно, что среднеквадратичное отклонение CML-оценок уменьшилось по сравнению со случаем $r = 1$, и убывает к 0.

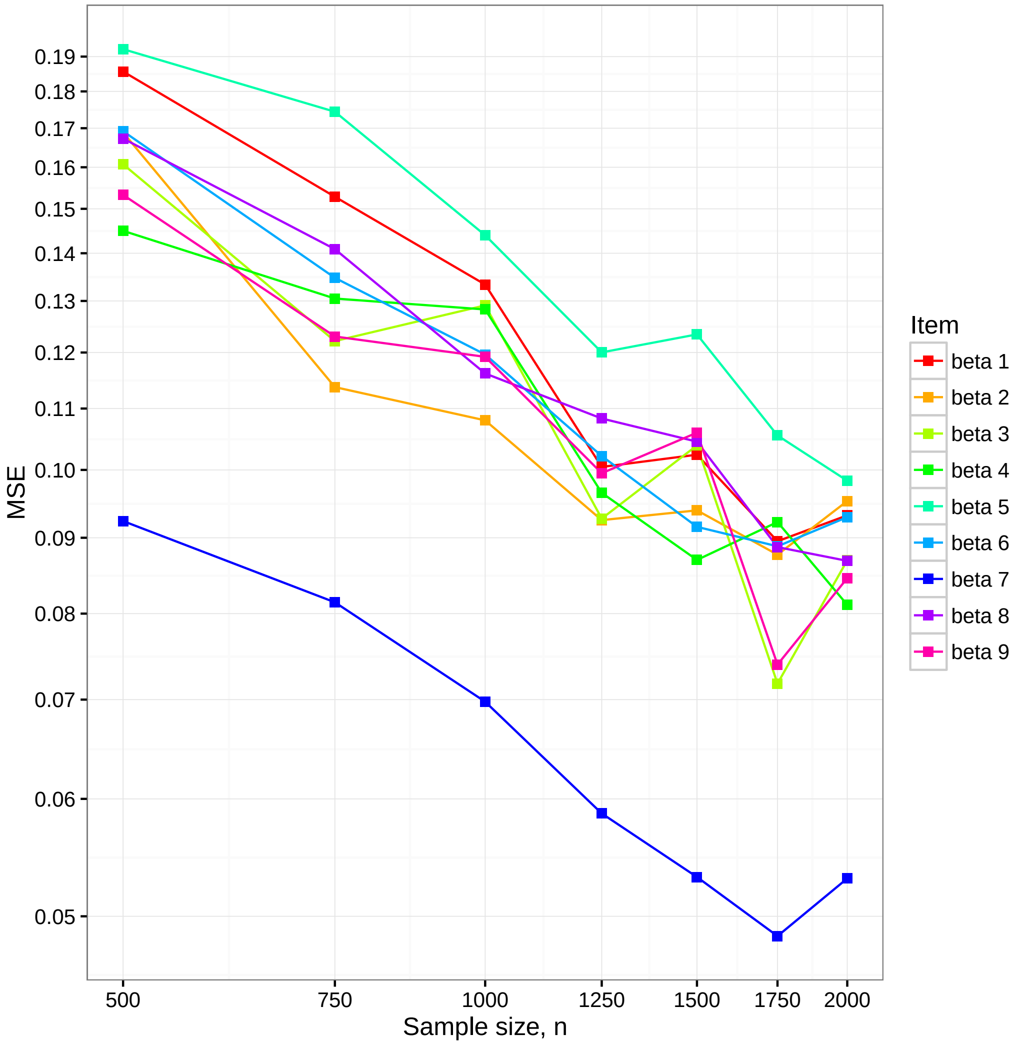

Рис. 3.1. Зависимость среднеквадратичного отклонения от объема выборки n, $k = 9$, $r = 1$. β_7 — «общий» вопрос.

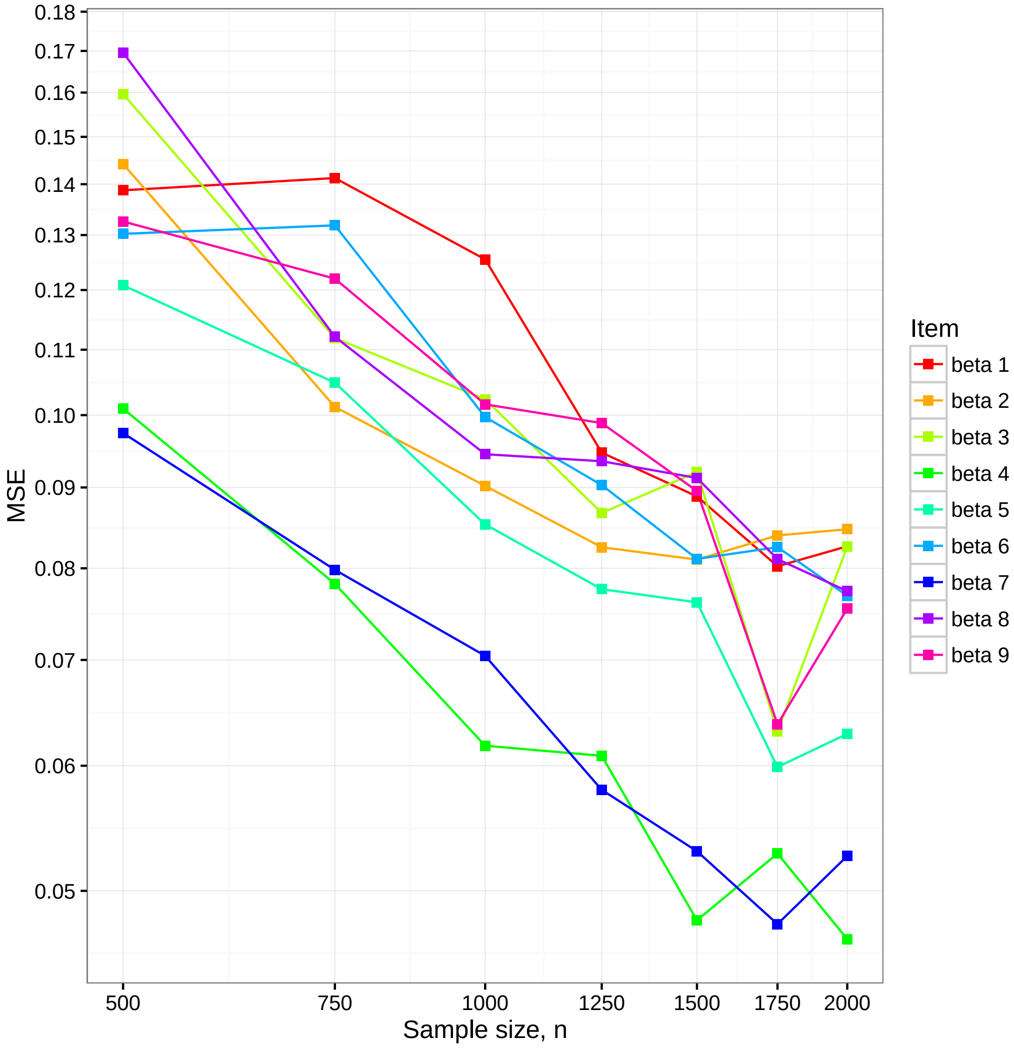

Рис. 3.2. Зависимость оценки среднеквадратичного отклонения от объема выборки n, $k = 9$, $r = 3$. $\beta_4, \beta_5, \beta_7$ — «общие» вопросы.

3.3.2. Моделирование. Число вопросов $k = 15$

Далее был рассмотрен случай двух тестов, содержащих по $k_i = 8$, $i = 1, 2$ вопросов, при числе «общих» вопросов $r = 1$. Был рассмотрен общий тест, состоящий из $k = 15$ вопросов с фиксированными параметрами сложности, указанными в Таблице 3.2. В качестве «общего» вопроса был зафиксирован вопрос с параметром сложности β_8. В список вопросов первого теста были включены вопросы, соответствующие параметрам $\beta_1, \ldots, \beta_8$, в список вопросов второго теста — вопросы со значениями параметров $\beta_8, \ldots, \beta_{15}$.

Также при моделировании было проверено, что смещение построенных оценок мало и не превышает 0.02.

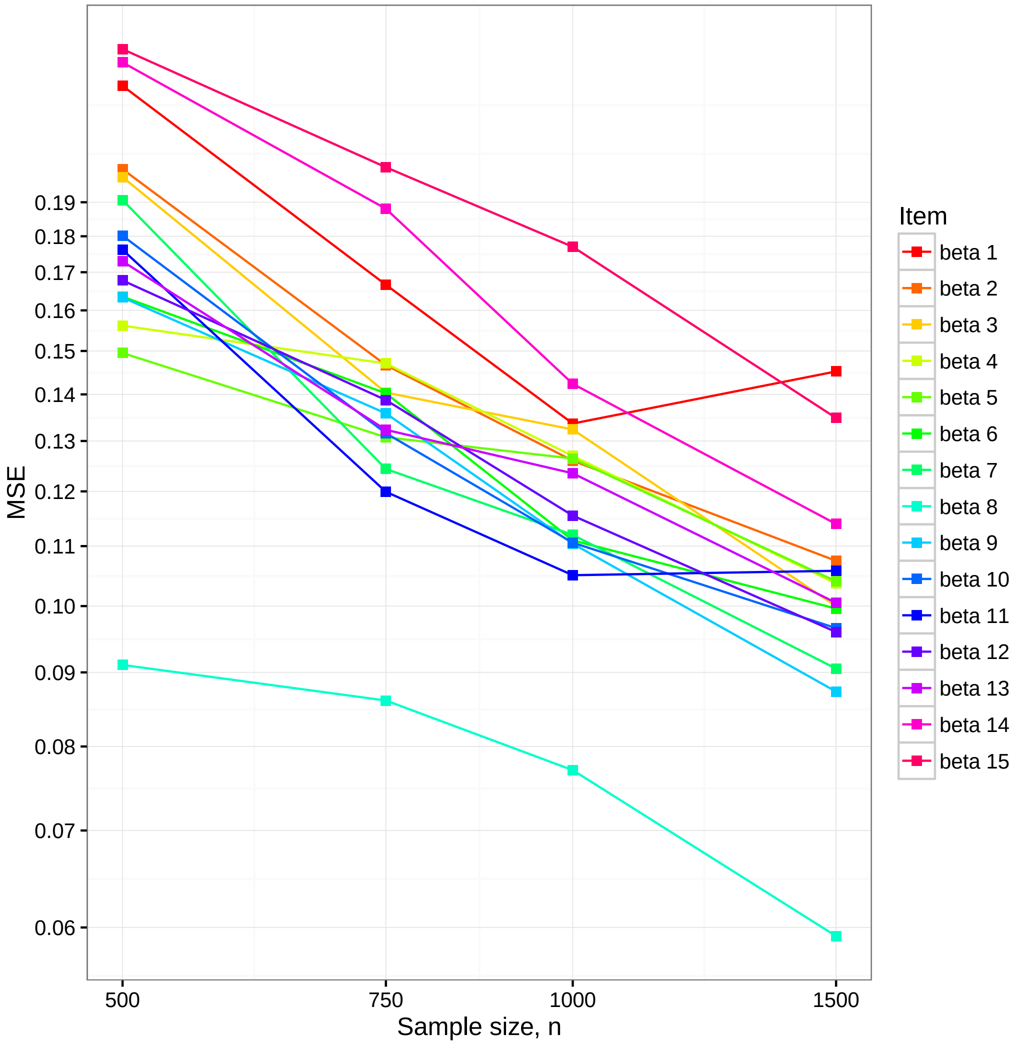

Рис. 3.3. Зависимость среднеквадратичного отклонения от объема выборки n, $k = 9$, $r = 1$. «Общий» вопрос β_8.

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
-2.00	-1.30	-1.05	-0.85	-0.50	-0.30	-0.05	0.00
β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	
0.01	0.04	0.40	0.60	0.80	1.80	2.40	

Таблица 3.2. Значения параметров сложности вопросов β_j, $k = 9$.

На рисунке 3.3 изображена зависимость среднеквадратичного отклонения от объема выборки n. Видно, что с увеличением объема выборки среднеквадратичное отклонение убывает к 0, скорость сходимости можно оценить как $\sqrt{n}$.

В результате моделирования для числа параметров сложностей вопросов $k = 9, k = 15$ и различного числа общих вопросов r было проверено, что CML-оценки в случае неполного дизайна асимтотически несмещенные, состоятельные, при этом скорость сходимости среднеквадратичного отклонения к нулю можно оценить как $\sqrt{n}$.

Заключение

В рамках данной работы в математической среде R был реализован метод оценивания значений параметров сложностей вопросов CML, где значения функции γ были вычислены по определению и по рекуррентным соотношениям. Для того, чтобы проиллюстрировать устойчивость реализованного алгоритма было выполнено моделирование данных, построены CML-оценки и проверены их свойства.

Для иллюстрации преимуществ метода CML перед стандартными методами оценивания, в среде R был реализован один из стандартных методов, метод JML, и для сравнения проиллюстрировано поведение построенных JML- и CML-оценок при растущем объеме выборки.

Также было рассмотрено естественное расширение задачи оценивания способностей людей и сложностей вопросов, а именно случай неполного дизайна. В среде R был реализован метод CML для случая неполного дизайна, построены оценки значений параметров сложностей вопросов и исследовано поведение построенных оценок при увеличении числа неизвестных параметров, разном числе «общих» вопросов и увеличивающемся объеме выборки. Для разного числа параметров сложности вопросов было показано, что смещение CML-оценок для случая неполного дизайна пренебрежимо мало, среднеквадратическое отклонение построенных оценок убывает к нулю, скорость сходимости была оценена как $\sqrt{n}$.

Список литературы

1. Molenaar I. Some Background for Item Response Theory and the Rasch Model // Rasch Models. Foundations, Recent Developments, and Applications / Ed. by G. H. Fischer, I.W. Molenaar. — Springer-Verlag New York, Inc., 1995. — P. 3–15.
2. Andersen E. B. The numerical solution of a set of conditional estimation equations // Journal of the Royal Statistical Society. Series B (Methodological). — 1972. — Vol. 34, no. 1. — P. 42–54.
3. Molenaar I. Estimation of Item Parameters // Rasch Models. Foundations, Recent Developments, and Applications / Ed. by G. H. Fischer, I.W. Molenaar. — Springer-Verlag New York, Inc., 1995. — P. 39–53.
4. Pfanzagl J. On the consistency of conditional maximum likelihood estimators // Annals of the Institute of Statistical Mathematics. — 1993. — Vol. 45. — P. 703–719.