Задачи оценивания параметров модели IRT

Коврыга Валерия Валерьевна, гр. 422

Санкт-Петербургский государственный университет
Математико-механический факультет
Кафедра статистического моделирования

Научный руководитель: к.ф.-м.н., доц. Коробейников А.И.
Рецензент: ассистент Шлемов А.Ю.

г. Санкт-Петербург
2016 г.

Введение

Способности (abilities) — «скрытые» качества:

- не описываются определенным видом деятельности,
- не поддаются непосредственному измерению.

Вопрос: как оценивать способности людей?

Пример: тестирование СПбГУ по английскому языку.
Проверяется знание студентов английского языка (уровень B2).

Подход **Item Response Theory (IRT)**:

- люди отвечают на вопросы,
- ответы на вопросы позволяют оценить способности.

Вероятностная модель (Rasch, 1960):

$\theta, \beta \in \mathbb{R} \quad \theta$ — способность человека, β — сложность вопроса.

Бернуллиевская с.в. $\xi \in \{0, 1\}$ — ответ человека на вопрос.

Вероятность правильного ответа:

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

n человек и k вопросов:

$$\begin{aligned} \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_n) \\ \boldsymbol{\beta} &= (\beta_1, \ldots, \beta_k) \end{aligned} \quad \Rightarrow$$

- независимая выборка $\mathrm{x}_{ij} \sim \mathsf{P}_{\theta_i,\beta_j}$,
- кол-во вопросов k фиксировано.

Задача оценивания параметров в рамках модели Rasch

Проблема: с ростом объема выборки по n увеличивается количество параметров θ_i.

$\Rightarrow$ Стандартные методы оценивания неприменимы.

Задача: оценить β_j при мешающих параметрах θ_i.

Метод оценивания: Conditional Max. Likelihood (Andersen, 1972).

$\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}$ строятся без вычисления оценок мешающих параметров θ_i.

$\measuredangle$ $\boldsymbol{n = 1}$.

Выборка: $X = (x_1, \ldots, x_k) \in \{0,1\}^k, \quad \mathrm{r} = \sum_{j=1}^k x_j$.

Функция правдоподобия:

$$\mathbf{L}(X, \theta, \boldsymbol{\beta}) = \underbrace{\frac{\exp\left(-\sum_{j=1}^k \beta_j x_j\right)}{\gamma_{\mathrm{r}_\mathbf{0}}(\boldsymbol{\beta})}}_{\mathbf{L}(X, \boldsymbol{\beta} \mid \mathrm{r}=\mathrm{r}_0)} \cdot \underbrace{\frac{\exp\left(\theta \sum_{j=1}^k x_j\right) \cdot \gamma_{\mathrm{r}_\mathbf{0}}(\boldsymbol{\beta})}{\prod_{j=1}^k \left(1 + \exp(\theta - \beta_j)\right)}}_{\mathsf{P}(\mathrm{r}=\mathrm{r}_0)},$$

$$\gamma_{\mathbf{r}_0}(\boldsymbol{\beta}) \coloneqq \sum_{y|\mathrm{r}_0} \exp\left(-\sum_{j=1}^k \beta_j y_j\right), \quad y \in \{0,1\}^k : \quad \sum_{j=1}^k y_j = \mathrm{r}_0.$$

CML-оценка:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}(X_1, \ldots, X_n) = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} \prod_{i=1}^n \frac{\exp\left(-\sum_{j=1}^k \beta_j x_{ij}\right)}{\gamma_{\mathrm{r}_\mathbf{i}}(\boldsymbol{\beta})}, \quad \mathrm{r}_i = \sum_{j=1}^k x_{ij}.$$

Свойства CML-оценок

Утверждение (J. Pfanzagl, 1993)

Для любого набора θ_i, *т.ч.* $\lim_n \sum_{i=1}^{n} \exp(-\theta_i) = \infty$, $n \to +\infty$, *оценки* $\widehat{\boldsymbol{\beta}}_{\mathrm{CML}}$ *состоятельны.*

Метод оценивания CML позволяет построить оценки, обладающие хорошими статистическими свойствами.

Недостатки метода CML:

- **проблема:** точка максимума не единственная, т. к.

$$\operatorname{argmax}_{\boldsymbol{\beta}} \mathbf{L}(X, \boldsymbol{\beta} - \boldsymbol{c} \mid \mathrm{r}) = \operatorname{argmax}_{\boldsymbol{\beta}} \mathbf{L}(X, \boldsymbol{\beta} \mid \mathrm{r}),$$
$$\boldsymbol{c} = (c, \ldots, c), \quad c = const.$$

  Достаточно ∡ линейное ограничение: $\sum_{j=1}^{k} \beta_j = 0.$

- трудоемкость реализации.

Трудности реализации метода CML

Проблема: трудоемкость вычислений значений $\gamma_{\mathbf{r}}(\boldsymbol{\beta})$.

$$\gamma_{\mathbf{r}}(\boldsymbol{\beta}) := \sum_{y|\mathrm{r}} \exp\Big(-\sum_{j=1}^{k} \beta_j y_j\Big), \quad y \in \{0,1\}^k : \sum_{j=1}^{k} y_j = \mathrm{r}.$$

- большое число слагаемых: C_k^r,
- $\prod_{j=1}^{k} \exp(-\beta_j y_j) \ll 1$.

При численной реализации:

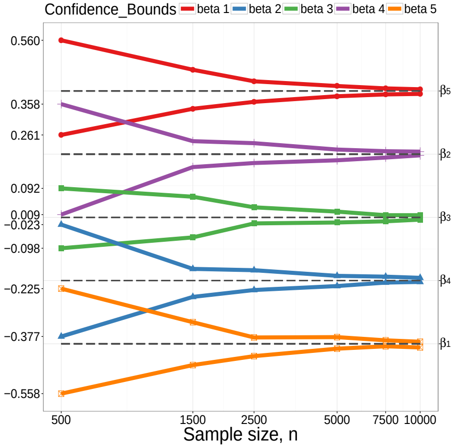
- накопление ошибок,
- экспоненциальная трудоемкость.

Решение: рекуррентные соотношения (Andersen, 1972):

- уменьшение трудоемкости: $C \asymp O(k \cdot \mathrm{r})$ при $1 < \mathrm{r} < k$,
- уменьшение количества вычислений $\Rightarrow$ меньшая потеря точности.

Результаты: алгоритм Андерсена

Задача: реализация метода CML с использованием рекуррентных соотношений (Andersen, 1972) + проверка свойств CML-оценок.



Зависимость ширины 95% доверительного интервала для β_j от кол-ва человек n.

Условия эксперимента:

- фиксированное число вопросов $k = 5$,
- фиксированные значения параметров сложности $\beta_1, \ldots, \beta_5$,
- $\theta_i \sim N(0, 1)$.

Промоделирована повторная выборка $\widehat{\beta}_j$ объема 100 при увеличивающемся кол-ве человек n.

Построены 95% доверительные интервалы для значений β_j.

Расширение задачи

Пример: тестирование СПбГУ по английскому языку:

- несколько вариантов тестирования,
- каждый студент пишет только один вариант.

$\Rightarrow$ Разные группы студентов пишут разные варианты.

Вопрос: как сравнивать сложности разных вариантов тестирования?

Тривиальный подход: построить оценки отдельно для каждого варианта.

Расширение задачи

Проблема: оценки для разных вариантов несравнимы между собой.

$\widehat{\boldsymbol{\beta}}_{\text{CML}}$ определялись относительно, с точностью до сдвига на $const$.
$\Rightarrow$ Д+ля однозначности $\widehat{\boldsymbol{\beta}}_{\text{CML}}$ вводилось условие: $\sum_{j=1}^{k} \beta_j = 0$.

Пусть кол-во вопросов k увеличивается:

- $\measuredangle$ вопросы сложности: $\beta_1, \ldots, \beta_k$.
  Для CML-оценок верно: $\widehat{\beta}_1 + \ldots + \widehat{\beta}_k = 0$.
- $\measuredangle$ $k+1$ вопрос $\beta_1, \ldots, \beta_k, \beta_{k+1} \quad \Rightarrow \quad \widehat{\beta}_1 + \ldots \widehat{\beta}_k + \widehat{\beta}_{k+1} = 0$.

$\Rightarrow$ Новые оценки не сравнить с предыдущими.

Решение: тесты с «общими» вопросами:

- тесты рассматриваются как один общий тест,
- «общие» вопросы — вопросы, на которые отвечали все группы студентов.

Оценивание парамеров

Метод оценивания: метод CML в случае неполных данных

Выборка: $X = (x_1, \ldots, x_k)$, $\mathrm{r}(X) = \sum_{j=1}^{k} b_j x_j$, $n = 1$ наблюдение.

Индикатор: $b_j = \begin{cases} 1, & \text{если человек отвечал на } j\text{-й вопрос,} \\ 0, & \text{иначе.} \end{cases}$

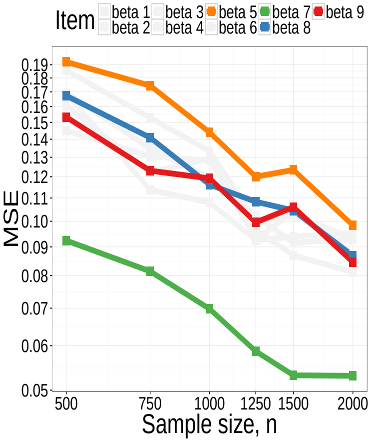
Совместное правдоподобие:

$$\mathbf{L}(X, \theta, \boldsymbol{\beta}) = \underbrace{\frac{\exp\big(-\sum_{j=1}^{k} \beta_j x_j b_j\big)}{\sum_{y|\mathrm{r}_0} \prod_{j=1}^{k} \exp(-b_j \beta_j y_j)}}_{\mathbf{L}(X, \boldsymbol{\beta} \,|\, \mathrm{r}=\mathrm{r}_0) \text{ не зависит от } \theta} \cdot \mathsf{P}(\mathrm{r} = \mathrm{r}_0), \quad \begin{array}{c} \sum_{j=1}^{k} y_j b_j = \mathrm{r}_0, \\ y_j \in \{0, 1\}. \end{array}$$

CML-оценка:

$$\hat{\boldsymbol{\beta}}_{\mathrm{CML}} = \operatorname*{argmax}_{\boldsymbol{\beta}} \prod_{i=1}^{n} \frac{\exp\big(-\sum_{j=1}^{k} \beta_j x_{ij} b_{ij}\big)}{\sum_{y|\mathrm{r}_i} \prod_{j=1}^{k} \exp(-b_{ij} \beta_j y_j)}.$$

Задача: экспериментально исследовать поведение $\hat{\boldsymbol{\beta}}_{\mathrm{CML}}$ при разных n, k и числе «общих» вопросов.



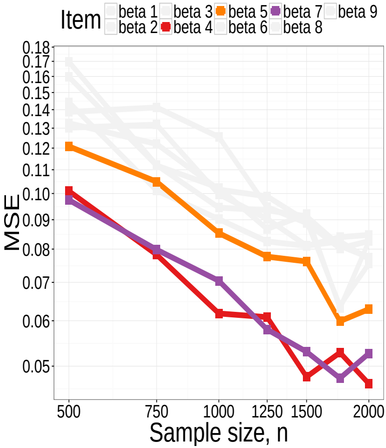
Зависимость среднеквадратичного отклонения β_j от кол-ва человек n.

Описание условий эксперимента:

- два варианта тестирования по $k_1 = k_2 = 5$ вопросов,
- кол-во «общих» вопросов $k_3 = 1$,
- кол-во тестируемых $n_1 = n_2$,

  $\Rightarrow \measuredangle$ общий тест из $k = 9$ вопросов,
- общее кол-во тестируемых $n = n_1 + n_2$,
- $\theta_i \sim N(0, 1)$,
- зафиксированы значения параметров сложности $\beta_1, \ldots, \beta_9$,
- параметр сложности «общего» вопроса: β_7,
- повторная выборка $\widehat{\beta_j}$ объема 100.

Исследование свойств CML-оценок. 3 общих вопроса из 9

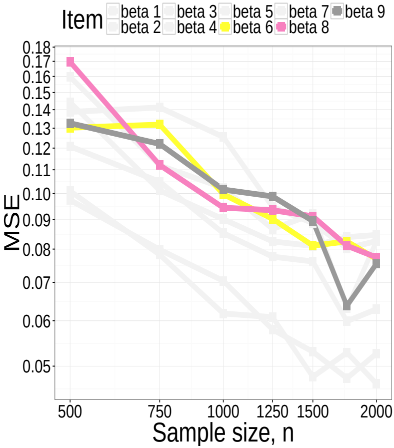


Зависимость среднеквадратичного отклонения «общих» вопросов от кол-ва человек n.

Описание условий эксперимента:

- два варианта тестирования по $k_1 = k_2 = 6$ вопросов,
- кол-во «общих» вопросов $k_3 = 3$,
- кол-во тестируемых $n_1 = n_2$,

  $\Rightarrow \measuredangle$ общий тест из $k = 9$ вопросов,
- общее кол-во тестируемых $n = n_1 + n_2$,
- $\theta_i \sim N(0, 1)$,
- зафиксированы значения параметров сложности $\beta_1, \ldots, \beta_9$,
- параметры сложностей «общих» вопросов: β_4, β_5, β_7,
- повторная выборка $\widehat{\beta_j}$ объема 100.



Зависимость среднеквадратичного отклонения не «общих» вопросов от кол-ва человек n.

Описание условий эксперимента:

- два варианта тестирования по $k_1 = k_2 = 6$ вопросов,
- кол-во «общих» вопросов $k_3 = 3$,
- кол-во тестируемых $n_1 = n_2$,

  $\Rightarrow \measuredangle$ общий тест из $k = 9$ вопросов,
- общее кол-во тестируемых $n = n_1 + n_2$,
- $\theta_i \sim N(0, 1)$,
- зафиксированы значения параметров сложности $\beta_1, \ldots, \beta_9$,
- параметры сложностей «общих» вопросов: β_4, β_5, β_7,
- повторная выборка $\widehat{\beta_j}$ объема 100.

Результаты

- Реализован метод оценивания CML с вычислением функции $\gamma_{\mathrm{r}}(\boldsymbol{\beta})$ по рекуррентным соотношениям (Andersen, 1972).
- Было произведено моделирование CML-оценок и проверены их свойства.
- Реализован метод оценивания CML для случая неполных данных.
- На модельных данных исследовано поведение построенных оценок для некоторых частных случаев.