Задачи оценивания параметров модели IRT

Черниговская Мария Александровна, 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.

Санкт-Петербург
2016г.

Введение

Ability (способность) — скрытое качество человека.

На практике ставится задача оценивать способности людей.
Подход IRT: анализ ответов на тестовые вопросы.

Модель Rasch (Rasch, 1961) — базовая модель:

- Способности людей и сложности вопросов — оцениваемые параметры.
- Постулируется модель для вероятности правильного ответа человека с конкретной способностью на вопрос конкретной сложности.

Проблема: число параметров растет с увеличением объема выборки.
$\Rightarrow$ Стандартные методы оценивания неприменимы.

Идея: отказаться от одновременного оценивания параметров сложности и способности и модифицировать модель.

Модель эксперимента (Bock, Lieberman 1970)

Предположение: существует популяционное распределение способностей людей $\mathcal{P}_\alpha$, известное с точностью до параметра α.

Параметры:

- k вопросов, n человек: $k \ll n$, k и n известны.
- $\mathbf{B} = \{\beta_j\}_{j=1}^{k}$, β_j — сложность j-го вопроса.
- α — параметр популяционного распределения.

$\{x_{ij}\}_{i=1 j=1}^{n k}$ — с.в., $x_{ij} \in \{0, 1\}$:

- $X_i = (x_{i1}, \ldots, x_{ik})$, $i \in \{1, \ldots, n\}$ независимы.
- $\{\vartheta_i\}_{i=1}^{n}$ — повторная независимая выборка из распределения $\mathcal{P}_\alpha$.
- $x_{i1}, \ldots, x_{ik}$ условно независимы при условии $\vartheta_i = \vartheta$, при этом

$$\mathrm{P}\big(x_{ij} = \delta \,\big|\, \vartheta_i = \vartheta\big) = \frac{\exp\big(\delta(\vartheta - \beta_j)\big)}{1 + \exp\left(\vartheta - \beta_j\right)}\ , \delta \in \{0, 1\}\ .$$

MML оценки (Bock, Lieberman 1970)

Задача: по выборке $\{x_{ij}\}_{i=1,j=1}^{n,k}$ оценить параметр сложности $\mathbf{B}$ и параметр популяционного распределения α.

Функция правдоподобия в рамках модели:

$$\mathbf{L}\big(\mathbf{B},\alpha\,|\,X_1,\ldots,X_n\big)=\prod_{i=1}^{n}\left(\int_{\mathbb{R}}\prod_{j=1}^{k}\frac{\exp\left(x_{ij}(\vartheta-\beta_j)\right)}{1+\exp\left(\vartheta-\beta_j\right)}\,\mathcal{P}_\alpha(d\vartheta)\right)\ .$$

MML оценка (ОМП в рамках модели):

$$(\hat{\mathbf{B}},\hat{\alpha})=\operatorname*{argmax}_{\mathbf{B},\alpha}\mathbf{L}\big(\mathbf{B},\alpha\,|\,X_1,\ldots,X_n\big)\ .$$

Проблема: сложная структура функции правдоподобия.

Дискретный случай (Hanson, 1998)

$$\mathcal{P}_\alpha : \begin{pmatrix} t_1 & t_2 & \dots & t_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}; \quad p_s > 0, \sum_{s=1}^{m} p_s = 1,$$

$\{t_s\}_{s=1}^m$ — известные градации, $\mathsf{P} = \{p_s\}_{s=1}^{m-1}$ — параметры.

$$\mathbf{L}\big(\mathbf{B}, \mathsf{P} \mid X_1, \dots, X_n\big) = \prod_{i=1}^{n} \left(\sum_{s=1}^{m} \prod_{j=1}^{k} \frac{\exp\left(x_{ij}(t_s - \beta_j)\right)}{1 + \exp\left(t_s - \beta_j\right)} p_s \right) .$$

$$(\hat{\mathbf{B}}, \hat{\mathsf{P}}) = \underset{\mathbf{B}, \mathsf{P}}{\operatorname{argmax}} \, \mathbf{L}\big(\mathbf{B}, \mathsf{P} \mid X_1, \dots, X_n\big) .$$

Решение: EM-алгоритм (Hanson, 1998).

Преимущество: оптимизация функции многих переменных $\mathbf{L}\big(\mathbf{B}, \mathsf{P} \mid X_1, \dots, X_n\big)$ сводится к оптимизациям нескольких функций одной переменной.

Непрерывный случай

Проблема: вычисление функции $\mathbf{L}\big(\mathbf{B}, \alpha \,|\, X_1, \ldots, X_n\big)$.

Известные решения:

- Замена интегралов на квадратурные формулы. Применение алгоритма Ньютона–Рапсона. (Bock, Lieberman — 1970)
- MCMC (Patz, Junker — 1999)
- Сведение задачи к дискретной, с помощью аппроксимации $\mathcal{P}_\alpha$ дискретным распределением с известными градациями. Применение алгоритма Hanson. (Bock, Aitkin — 1981)

Проблема: выбор градаций должен зависеть от неизвестного параметра популяционного распределения α.

$\Rightarrow$ Нужно использовать переменные градации.

$\Rightarrow$ В бакалаврской работе была предложена модификация алгоритма Hanson.

Модификация алгоритма Hanson

ℓ-я итерация алгоритма:

1. Вход: $\mathbf{B}^\ell$, α^ℓ ($\mathbf{B}^0$, α^0 — начальные данные)
2. $H^\ell :=$ аппроксимация $\mathcal{P}_{\alpha^\ell}$ дискретным распределением
3. К дискретному распределению H^ℓ применяется одна итерация алгоритма Hanson $\Rightarrow \mathbf{B}^{\ell+1}$
4. $\alpha^{\ell+1} := \operatorname*{argmax}_\alpha \mathbf{L}\big(\alpha; \mathbf{B}^{\ell+1} \,|\, X_1, \ldots, X_n\big)$

Проблема: как выбрать аппроксимацию $\mathcal{P}_{\alpha^\ell}$?

В бакалаврской работе были предложены следующие способы дискретизации:

- с помощью ОМП способностей
- разбивание по квантилям
- с помощью корней многочлена Эрмита
  (для нормального популяционного распределения)

Проверка метода оценивания на модельных данных

5 вопросов: $\mathbf{B} = (-1,\ -0.5,\ 0,\ 0.5,\ 1)$.
Число человек n: $64, 128, \ldots, 8192$.

Популяционное распределение:

- $\mathcal{P}_\alpha = \mathcal{N}(\alpha, 1)$ — параметр сдвига, $\alpha_0 = 0.5$.
- $\mathcal{P}_\alpha = \mathcal{N}(0, \alpha^2)$— параметр масштаба, $\alpha_0 = 0.7$.

Вопрос: Как выбрать параметры дискретизации?

- Максимальное правдоподобие
  - градации := MLE оценки способностей
  - градации := MLE оценки способностей с коррекцией смещения (Hoijtink, Boomsma 1995)
- Квантили
  - фиксированное число градаций
  - $\lceil \log_2(n) \rceil$ градаций
  - $\lceil \sqrt{n} \rceil$ градаций
  - $\lceil n/4 \rceil$ градаций
- Корни многочлена Эрмита
  Разное число фиксированных и «плавающих» градаций.

Способ ОМП

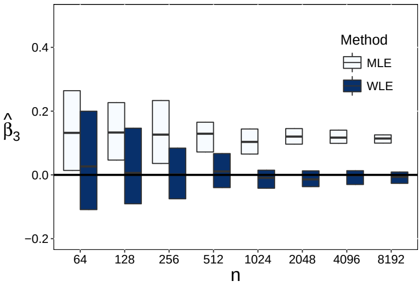

Рис.: Параметр сдвига, $\beta_3 = 0$

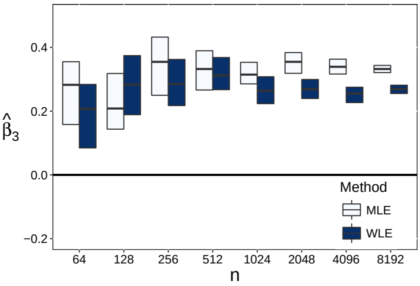

Рис.: Параметр масштаба, $\beta_3 = 0$

Вывод:

- MLE оценки с коррекцией смещения подходят в случае параметра сдвига.
- При оценивании популяционного распределения с помощью ОМП способностей в случае параметра масштаба оценки сложностей вопросов $\mathbf{B}$ несостоятельны.

MLE с коррекцией смещения: другие способы назначить вероятности

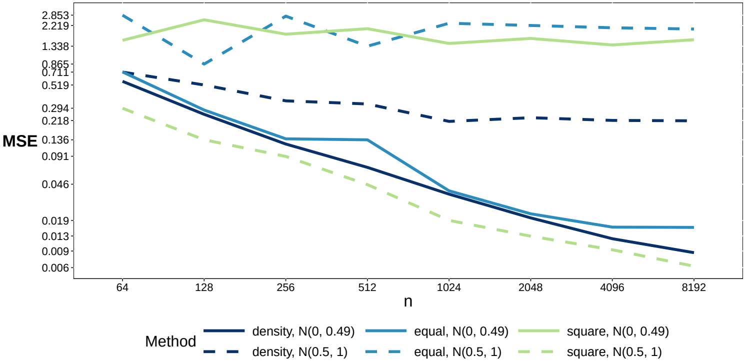

Рис.: Среднеквадратическое отклонение оценок.

Хорошее качество оценок либо для параметра сдвига, либо для параметра масштаба. $\Rightarrow$ Способ ОМП даёт неудовлетворительные оценки сложностей вопросов $\mathbf{B}$.

Способ квантилей

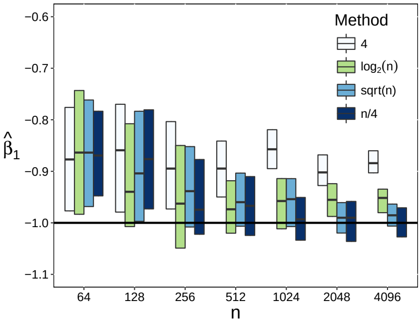

Рис.: Параметр сдвига, $\beta_1 = -1$

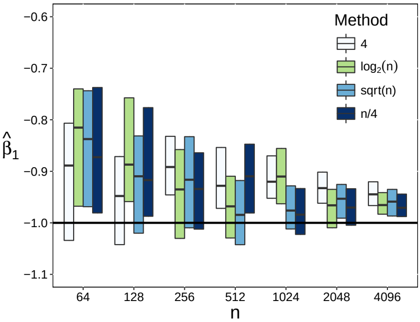

Рис.: Параметр масштаба, $\beta_1 = -1$

Вывод:

- Число градаций должно увеличиваться с ростом n.
- Параметр сдвига: лучшая оценка с помощью $n/4$ градаций.
- Параметр масштаба: нельзя сделать вывод, о наилучшем числе градаций.

Среднекв-е отклонение для вопросов $\beta_1 = -1$, $\beta_2 = -0.5$, $\beta_3 = 0$.

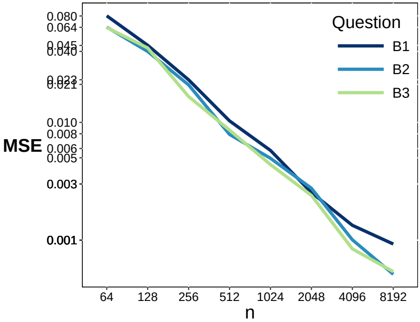

Рис.: Параметр сдвига

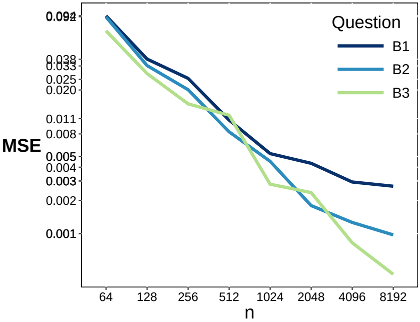

Рис.: Параметр масштаба

Чем меньше модуль сложности вопроса, тем лучше качество оценки.

Способ корней многочлена Эрмита

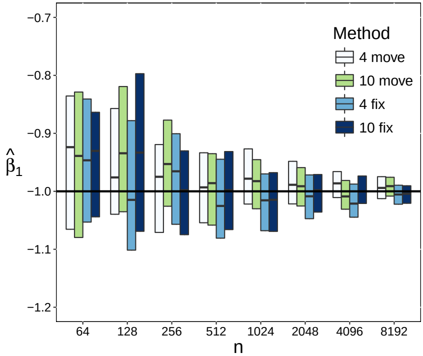

Рис.: Параметр сдвига, $\beta_1 = -1$

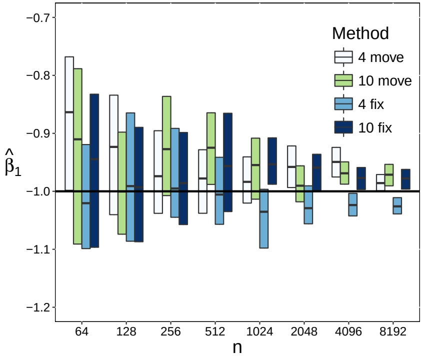

Рис.: Параметр масштаба, $\beta_1 = -1$

Вывод:

- Нельзя сделать вывод, что увеличение числа градаций улучшает оценки сложностей $\mathbf{B}$.
- В случае параметра масштаба нельзя сказать, что «плавающие» градации улучшают оценки сложностей.

Заключение

Результаты:

- В работе предложен новый способ вычисления MML-оценок, основанный на модификации EM-алгоритма для дискретного распределения.
- Применимость нового способа проверена на модельных данных.
- Показано, что способ ОМП для аппроксимации распределения способностей в текущем виде неприменим.
- Если в качестве градаций используются квантили, то необходимо увеличивать число градаций с увеличением числа человек. Показано, что качество оценок ухудшается с увеличением модуля сложности вопросов.
- В случае нормального популяционного распределения способностей предлагается использовать градации, построенные с помощью «плавающих» корней многочлена Эрмита.

Открытые вопросы:

- Сравнение оценок с оценками, полученными известными методами.
- Свойства оценок для реальных данных.
- ...