Барицентрическая параметризация обобщенных обратных матриц в ANOVA

Белоусов Юрий

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

Научный руководитель: к.ф.-м.н., д. Алексеева Н. П.
Рецензент: к.т.н., д. Белякова Л. А.

Санкт-Петербург
2016г.

Общая линейная модель и постановка задачи

Общая линейная моедль: $Y = \mathbf{X}\beta + \varepsilon$.
$Y \in \mathbb{R}^n$ — наблюдения, $\beta \in \mathbb{R}^r$ — вектор параметров, ε — ошибки.

Задача: по наблюдениям оценить вектор параметров

Для оценки параметров используется МНК: $||Y - \mathbf{X}\hat{\beta}|| \to \min$.
Оценка получается как решение системы $\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\beta} = \mathbf{X}^{\mathrm{T}}Y$.

Если матрица $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ необратима — решение не единственно.

Классический подход к оценке параметров

Наложение дополнительных линейных ограничений, т.е. требуется $\mathbf{H}^{\mathrm{T}}\hat{\beta} = Y_2$, $\mathbf{H}^{\mathrm{T}} \in \mathbb{R}^{r \times p}, Y_2 \in \mathbb{R}^p$.

Теорема (Шеффе, 1980)

Если матрица $\begin{pmatrix} \mathbf{X}^{\mathrm{T}} \\ \mathbf{H}^{\mathrm{T}} \end{pmatrix}$ *имеет ранг* r*, и кроме нулевой строки никакая линейная комбинация строк* $\mathbf{H}^{\mathrm{T}}$ *не представляется в виде линейной комбинации строк* $\mathbf{X}^{\mathrm{T}}$*, то* $\hat{\beta}$ *единственна.*

Примеры линейных ограничений

- $\sum\limits_{i=0} \beta_i$, т.е. сумма эффектов равна 0.
- $\sum\limits_{i=0} w_i \beta_i = 0$, где $w_i \geq 0$. Иначе говоря, взвешенная сумма эффектов равна 0.
- $\beta_r = 0$. Функция aov в R реализована так.

Обобщенно обратные матрицы

Определение

Пусть $\mathbf{A} \in \mathbb{R}^{r\times n}$. $\mathbf{A}^- \in \mathbb{R}^{n\times r}$ *называется* $\{(i),(j),\ldots,(k)\}$*-обратной к* $\mathbf{A}$*, если она удовлетворяет соотношениям* $(i),(j),\ldots,(k)$ *из* $O1 - O4$*, где*

$$\begin{array}{rccc} O1: & \mathbf{A}\mathbf{A}^-\mathbf{A} & = & \mathbf{A}, \\ O2: & \mathbf{A}^-\mathbf{A}\mathbf{A}^- & = & \mathbf{A}^-, \\ O3: & (\mathbf{A}^-\mathbf{A})^{\mathrm{T}} & = & \mathbf{A}^-\mathbf{A}, \\ O4: & (\mathbf{A}\mathbf{A}^-)^{\mathrm{T}} & = & \mathbf{A}\mathbf{A}^-. \end{array}$$

$\mathbf{A}^{(i,j,\ldots,k)}$ — обозначение для $\{(i),(j),\ldots,(k)\}$-обратной к $\mathbf{A}$.
$\mathbf{A}\{i,j,\ldots,k\}$ — множество всех $\mathbf{A}^{(i,j,\ldots,k)}$.
$\mathbf{A}^{(1,2,3,4)}$ — псевдообратная.

Оценка параметров через обобщенные обратные

Теорема

Оценка $\hat{\beta}$ является МНК оценкой для $||Y - \mathbf{X}\hat{\beta}|| \to \min$, тогда и только тогда, когда $\hat{\beta} = \mathbf{X}^{(1,3)}Y$, где $\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}$.

Неединственность обобщенных обратных матриц используются для исследования свойств оценок.

Пусть, F — некоторая функция. Оценка получается как

$$\begin{cases} ||\mathbf{X}\beta - Y|| \to \min \\ F(\beta) \to \min . \end{cases}$$

Иначе, можно записать так: $\hat{\beta} = \mathbf{X}^{(1,3)}Y$ при

$$F(\mathbf{X}^{(1,3)}Y) \to \min_{\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}} .$$

Построение обобщенных обратных матриц

Пусть для матрицы $\mathbf{A} \in \mathbb{R}^{n\times m}$ ранга r берутся $\nu = \nu(r|n) = \{(\nu_1, \nu_2, \ldots, \nu_r) \mid 1 \leq \nu_1 < \nu_2 < \cdots < \nu_r \leq n\}$ и $\lambda = \lambda(r|m) = \{(\lambda_1, \lambda_2, \ldots, \lambda_r) \mid 1 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_r \leq m\}$, что $\mathbf{A}_\lambda^\nu$ не вырождена. Тогда $\{1\}$-обратная имеет вид [Барт, 2003]

$$\mathbf{A}^- = (\mathbf{A}_\lambda)^- \mathbf{A}_\lambda^\nu (\mathbf{A}^\nu)^-,$$

где $\mathbf{A}_\lambda$ и $\mathbf{A}^\nu$ полного ранга, для них используются разные виды параметризации (матричная или барицентрическая).

Свойства разложения

Теорема

Свойство $O3$ обобщенно обратной матрицы $\mathbf{A}^-$, полученной в результате разложения

$$\mathbf{A}^- = (\mathbf{A}_\lambda)^- \mathbf{A}_\lambda^\nu (\mathbf{A}^\nu)^-, \tag{1}$$

зависит только от параметров обращения $(\mathbf{A}^\nu)^-$(правых параметров обращения).

Это значит, что матрицы из $\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}$, полученные с использованием разложения (1) параметризуются лишь $(\mathbf{A}_\lambda)^-$.

Однако такое разложение не совсем полно:

Теорема

Матрицы, получающиеся в результате разложения (1) суть $\{1,2\}$-обратные.

Барицентрическая параметризация

Для матрицы $\mathbf{A} \in \mathbb{R}^{n\times r}$ полного столбцового ранга рассматриваются такие $\nu_t = \nu_t(r|n) = \{(\nu_{t1}, \nu_{t2}, \dots, \nu_{tr}) \mid 1 \le \nu_{t1} < \nu_{t2} < \dots < \nu_{tr} \le n\}$, что $\det \mathbf{A}_{\nu_t} \ne 0$.

Тогда обобщенно обратная к $\mathbf{A}$ имеет вид $\mathbf{A}^- = \sum_t \alpha_t \mathbf{A}_{\nu_t}^{-1} \mathbb{1}_{\nu_t} + \mathbf{A}_\triangledown$, где суммирование ведется по всем ν_t, для которых матрица $\mathbf{A}_{\nu_t}$ не вырождена, $\mathbf{A}_\triangledown$ аннулятор $\mathbf{A}$, а α_t — параметры обращения.

Транспонированием получаем параметризацию обобщенных обратных к матрицам полного строчного ранга.

Матричная параметризация

В матричных параметрах, разложение $\mathbf{A}^- = (\mathbf{A}_\lambda)^- \mathbf{A}_\lambda^\nu (\mathbf{A}^\nu)^-$ примет вид

$$(\mathbf{A}_\lambda)^- = \mathbb{1}^\nu (\mathbf{A}_\lambda^\nu)^{-1} + \dot{\mathbf{H}}\mathbf{Q},$$
$$(\mathbf{A}^\nu)^- = (\mathbf{A}_\lambda^\nu)^{-1} \mathbb{1}_\lambda + \mathbf{P}\ddot{\mathbf{H}},$$

где

$\dot{\mathbf{H}}$, $\ddot{\mathbf{H}}$ определяются по $\mathbf{A}$,

$\mathbf{P}$, $\mathbf{Q}$ — параметры обращения.

Модель однофакторного дисперсионного анализа

$$\Omega_1 : \begin{cases} y_{ij} = \beta_0 + \beta_i + \varepsilon_{ij} & i = 1, \dots, r; \; j = 1, \dots, n_i \\ \{\varepsilon_{ij}\} & \text{независимы и распределены } N(0, \sigma^2). \end{cases}$$

В матричной записи $Y = \mathbf{X}\beta + \varepsilon$.

$$\mathbf{X} = \left(\begin{array}{ccccc} 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 \\ \hline 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & \dots & 0 \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline 1 & 0 & 0 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{array} \right)$$

Интерпретация барицентрических параметров

Барицентрическая параметризация левых параметров имеет вид

$$(\mathbf{X}_\lambda)^- = \sum_t \alpha_t \mathbf{X}_{\nu_t}^{-1} \mathbb{1}_{\nu_t} + \mathbf{X}_\triangledown.$$

Параметры модели β в этом случае выражаются явно:

$$\begin{aligned}
\hat{\beta}_0 &= \sum_{i=1}^{r} \frac{\alpha_i}{n_i} \sum_{j=1}^{n_i} x_{ij},\\
\hat{\beta}_1 &= \bar{x}_i - \hat{\beta}_0,\\
&\vdots\\
\hat{\beta}_r &= \bar{x}_i - \hat{\beta}_0.
\end{aligned}$$

Некоррелированность оценок

Ковариационная матрица оценок $\mathrm{cov}(\hat{\beta}) = \sigma^2 \mathbf{X}^-(\mathbf{X}^-)^{\mathrm{T}}$.

$$\begin{cases} \hat{\beta} = \mathbf{X}^{(1,3)} Y \\ \sum\limits_{0 \le i,j \le r; i \ne j} |\sigma_{ij}| \to \min\limits_{\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}}, \end{cases}$$

Барицентрическими вектором $\alpha = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$, приводится к диагональному виду

$$\mathrm{cov}(\hat{\beta}) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \frac{1}{n_r} \end{pmatrix}.$$

Минимальная норма решения

Теорема (Penrose, 1956)

Пусть $\mathbf{A} \in \mathbb{R}^{k \times n}$ *и* $b \in \mathbb{R}^k$, *тогда МНК решение уравнения* $\mathbf{A}x = b$ $x = \mathbf{A}^+ b$ *— одно из решений с минимальной нормой. Обратно, если для любого* b $x = \mathbf{X}b$ *— решение* $\mathbf{A}x = b$ *с минимальной нормой, то* $\mathbf{X} = \mathbf{A}^+$.

Этот же результат можно получить, минимизируя след ковариационной матрицы:

$$\begin{cases} \hat{\beta} = \mathbf{X}^{(1,3)} Y \\ \operatorname{Tr}(\mathbf{X}^{(1,3)} Y) \to \min\limits_{\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}} \end{cases}.$$

В барицентрических параметрах этому соответствует $\alpha_i = \frac{1}{r+1}, i = 0, 1, \ldots, r$.

Расширенная двухфакторная модель

Вместо обычной модели двухфакторного дисперсионного анализа с категориальными признаками,

$$\Omega_2 : \begin{cases} y_{ij} = \beta_0 + \beta_1 \alpha_{ij}^{(1)} + \beta_2 \alpha_{ij}^{(2)} + \beta_3 (\alpha^{(1)} \alpha^{(2)})_{ij} + \varepsilon_{ij} \\ \{\varepsilon_{ij}\} \text{ независимы и распределены } N(0, \sigma^2). \end{cases}$$

рассматривалась расширенная модель, где в качестве признаков — не нулевые линейные комбинации над $\mathbb{F}_2$.
В расширенной модели 7 признаков.
Матричная запись такая же $Y = \mathbf{X}\beta + \varepsilon$.
Для оценки параметров применялся тот же подход с минимизационной функцией.

Некоррелированность в двухфакторной модели

Ставится задача

$$\begin{cases} \hat{\beta} = \mathbf{X}^{(1,3)} Y \\ \sum\limits_{0 \le i,j \le 7; i \ne j} |\sigma_{ij}| \to \min\limits_{\mathbf{X}^{(1,3)} \in \mathbf{X}\{1,3\}} \end{cases},$$

где $\operatorname{cov}(\hat{\beta}) = (\sigma_{ij})_{0 \le i,j \le 7}$.

Полной некоррелированности достичь нельзя.

Был написан алгоритм для численного поиска минимума.

Итоги

- Найдено выражение для МНК оценок параметров в общей линейной модели.
- Предложен подход для оценки параметров через обобщенно обратные матрицы и наложение нелинейных ограничений.
- Найдено явное выражение параметров при барицентрической параметризации.
- В рамках однофакторной модели были получены некоторые конкретные оценки.
- Для расширенной двухфакторной модели был написан алгоритм для минимизации ковариаций оценок.