Оценка параметров статистических моделей на основе обобщённого обращения матриц

Григорьева Ирина Владимировна, гр. 16.М03-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Статистическое моделирование

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

Санкт-Петербург, 2018г.

Постановка задачи

Цель работы:

Оценка параметров модели дисперсионного анализа на основе барицентрической параметризации обобщенных обратных матриц по А. Г. Барту.

Математические задачи:

- Анализ структуры обобщенного обращения матриц и доказательство ее инвариантности относительно параметров, отвечающих за выбор невырожденного минора.
- Оценивание параметров моделей дисперсионного анализа при отсутствии ограничений на параметры с использованием обобщенных обратных матриц.
- Изучение влияния параметров обращения на свойства оценок моделей и на проверку статистических гипотез.

Классический подход к оценке параметров

Пусть $Y \in \mathbb{R}^n$ — наблюдения, $\mathbf{X} \in \mathbb{R}^{n\times r}$ — матрица плана, $\beta \in \mathbb{R}^r$ — параметры, ε — ошибки н. о. р. $\mathcal{N}(0, \sigma^2)$.

Общая модель имеет вид

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

Используется МНК: $\| Y - \mathbf{X}\hat{\beta} \| \to \min.$

Оценка — решение системы

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\beta} = \mathbf{X}^{\mathrm{T}}Y. \quad (1)$$

- Если $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ не вырождена $\Rightarrow$ (1) имеет единственное решение
$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y.$$
- Если $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ вырождена, т. е. $\mathrm{rank}(\mathbf{X}^{\mathrm{T}}\mathbf{X}) < r \Rightarrow$ (1) имеет бесконечно много решений. Требуется наложение дополнительных линейных ограничений для выбора одного из них.

Примеры линейных ограничений:

- $\sum\limits_{i=0} \beta_i = 0$, т. е. сумма эффектов равна 0.
- $\sum\limits_{i=0} \omega_i\beta_i = 0$, где $\omega_i \geq 0$, т. е. взвешенная сумма эффектов равна 0.

Необходимые определения

Пусть $\mathbf{A} \in \mathbb{R}^{r\times n}$, $\mathbf{A}^*$ — сопряженная матрица,

$$O_1: \ \mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A},$$

$$O_2: \ \mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-,$$

$$O_3: \ (\mathbf{A}\mathbf{A}^-)^* = \mathbf{A}\mathbf{A}^-,$$

$$O_4: \ (\mathbf{A}^-\mathbf{A})^* = \mathbf{A}^-\mathbf{A}.$$

Матрица $\mathbf{A}^- \in \mathbb{R}^{n\times r}$ называется $\{i, j, \ldots, k\}$-**обратной** к A, если она удовлетворяет соотношениям $O_i, O_j, \ldots, O_k$.

Пусть $\nu(l|n) = \{(\nu_1, \ldots, \nu_l) \mid \nu_1 < \nu_2 < \ldots < \nu_l; \ \nu_i \in \mathbb{N}_n\}$.

ν-**частичная матрица** — матрица, составленная из строк $(\mathbb{1}_\nu(l|n))$ или из столбцов $(\mathbb{1}^\nu(l|n))$ матрицы $\mathbb{1}_n$, соответствующих $\nu(l|n)$.

Построение обобщенных обратных матриц

Утверждение (Рао, 1968)

Пусть для матрицы $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathrm{rank}(\mathbf{A}) = r$, *берутся* $\lambda = \lambda(r|n) = \{(\lambda_1, \dots, \lambda_r) \mid 1 \le \lambda_1 < \lambda_2 < \dots < \lambda_r \le n\}$ *и* $\nu = \nu(r|m) = \{(\nu_1, \dots, \nu_r) \mid 1 \le \nu_1 < \nu_2 < \dots < \nu_r \le m\}$, *что* $\mathbf{A}_\lambda^\nu$ *не вырождена. Тогда* $\{1\}$*-обратная имеет вид*

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

где $\mathbf{A}_\lambda$ *и* $\mathbf{A}^\nu$ *полного ранга,* $(\mathbf{A}_\lambda)^- \in (\mathbf{A}_\lambda)\{1,3\}$, $(\mathbf{A}^\nu)^- \in (\mathbf{A}^\nu)\{1,4\}$.

Задача параметризации обобщенных обратных к произвольной матрице сводится к вопросу о параметризации обобщенно обратных к матрице полного ранга.

Для $\mathbf{A}_\lambda$ и $\mathbf{A}^\nu$ используется барицентрическая параметризация.

Барицентрическая параметризация g-обратных полного ранга

Пусть $\mathbf{A} \in \mathbb{R}^{n\times r}$ полного столбцового ранга.

Рассматриваются такие сочетания
$\nu = \nu_t(r|n) = \{(\nu_{t1}, \nu_{t2}, \ldots, \nu_{tr}) \mid 1 \le \nu_{t1} < \ldots < \nu_{tr} \le n\}$, что $\det \mathbf{A}_{\nu_t} \ne 0$, тогда обобщенно обратная к $\mathbf{A}$ имеет вид [Барт, 2003]

$$\mathbf{A}^{-} = \sum_t \alpha_t \mathbf{A}_{\nu_t}^{-1} \mathbb{1}_{\nu_t} + \mathbf{A}_{\triangledown},$$

где полагаем

$\sum\limits_t$ — суммирование по t, для которых $\det \mathbf{A}_{\nu_t} \ne 0$,

$\mathbf{A}_{\triangledown}$ — аннулятор $\mathbf{A}$, α_t — барицентрические параметры, $\sum\limits_t \alpha_t = 1$.

Операцией транспонирования получается параметризация обобщенных обратных к матрицам полного строчного ранга.

Пусть задана матрица $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\operatorname{rank}(\mathbf{A}) = r$,
$\lambda = (\lambda_1, \dots, \lambda_r)$ и $\nu = (\nu_1, \dots, \nu_r)$ — наборы индексов строк и столбцов: $\mathbf{A}_\lambda^\nu \neq 0$. Составим $\mathbf{A}_\lambda \in \mathbb{R}^{r\times m}, \mathbf{A}^\nu \in \mathbb{R}^{n\times r}$.

Построим обобщенную обратную матрицу при барицентрической параметризации

$$\mathbf{A}^-(\mu, \upsilon, \lambda, \nu) = (\mathbf{A}_\lambda)^- \mathbf{A}_\lambda^\nu (\mathbf{A}^\nu)^- = \sum_k \mu_k \mathbb{1}^{\tau_k} (\mathbf{A}_\lambda^{\tau_k})^{-1} \mathbf{A}_\lambda^\nu \sum_d \upsilon_d (\mathbf{A}_{\gamma_d}^\nu)^{-1} \mathbb{1}_{\gamma_d},$$

где $\tau = \tau_k(r|m)$: $\det \mathbf{A}_\lambda^{\tau_k} \neq 0$, $\gamma = \gamma_d(r|n)$: $\det \mathbf{A}_{\gamma_d}^\nu \neq 0$,
υ_d, μ_k — барицентрические параметры левого и правого обращения:

$$\sum_d \upsilon_d = 1, \ \sum_k \mu_k = 1.$$

Утверждение

Для обобщенной обратной матрицы при барицентрической параметризации выполняется равенство $\mathbf{A}^-(\mu, \upsilon, \lambda, \nu) = \mathbf{A}^-(\mu, \upsilon, \psi, \phi)$ *для* $\forall \psi, \phi$, *т. е. она не зависит от выбора* λ *и* ν.

Полученные результаты

Утверждение

Оценки $\hat{\beta} = \mathbf{A}^{-}\mathbf{X}^{\mathrm{T}}Y$ *при* $\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ *инвариантны относительно барицентрических параметров левого обращения* υ *и имеют вид*

$$\hat{\beta} = \sum_{k} \mu_k \mathbb{1}^{\tau_k} (\mathbf{A}_{\lambda}^{\tau_k})^{-1} \mathbf{X}_{\lambda}^{\mathrm{T}} Y.$$

Модель

$$x_{ij} = \beta_0 + \beta_i + \epsilon_{ij},$$

где $i = \overline{1, r}$; $j = \overline{1, n_i}$, $\sum\limits_{i=1}^{r} n_i = n$, ϵ_{ij} — н. о. р. $\mathcal{N}(0, \sigma^2)$.

Оценки параметров имеют вид

$$\hat{\beta}_0 = \sum_{i=1}^{r} \frac{\mu_i}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

$$\hat{\beta}_i = \bar{x}_i - \hat{\beta}_0.$$

Проверка гипотезы $H_0: \ \beta_1 = \ldots = \beta_r = 0.$
Получена статистика

$$F_{-} = \frac{\tilde{Q}_1/2df_1}{Q_2/df_2} \sim \mathcal{F}(df_1, df_2),$$

где $df_1 = r - 1, \ df_2 = n - r,$

$$\tilde{Q}_1 = \sum_{i=1}^{r} n_i(\bar{x}_i - \hat{\beta}_0)^2, \ Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2.$$

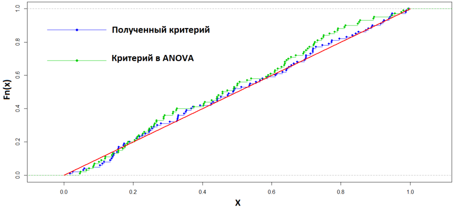

Рис.: Эмпирическая ф. р. $p - value$ полученного критерия и критерия в однофакторном анализе.

Замечание:
Специализированный критерий при $n_1 = \ldots = n_r = m$, $\mu_1 = 1, \mu_k = 0, \ k = \overline{2, r}$ оказывается более мощным и полезным в случае, когда необходимо сравнить одну группу с остальными.

Модель двухфакторного дисперсионного анализа

Рассмотрим модель

$$x_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

где $i = \overline{1,2}$, $j = \overline{1,2}$, $k = \overline{1,K}$, ϵ_{ijk} — н. о. р. $\mathcal{N}(0, \sigma^2)$.

В матричном виде имеет вид

$$Y = \mathbf{X}\beta + \varepsilon,$$

$$Y = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1n_1} \\ x_{21} \\ \vdots \\ x_{2n_2} \\ \vdots \\ x_{41} \\ \vdots \\ x_{4n_4} \end{bmatrix}, \quad \mathbf{X} = \left(\begin{array}{ccccc} 1 & 1 & 0 & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 1 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & 1 & 0 \\ \hline 1 & 0 & 1 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & 0 & 1 \end{array}\right), \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_4 \end{bmatrix}.$$

Оценки параметров модели двухфакторного анализа

Пусть $J = n_{11}^{-1} + n_{12}^{-1} + n_{21}^{-1} + n_{22}^{-1}$.

Параметры модели выражаются в виде линейных комбинаций

$$\begin{aligned}
\hat{\mu} &= \mu_1 M_1 + \ldots + \mu_4 M_4,\\
\hat{\alpha}_1 &= \mu_1 A_1^1 + \mu_2 A_2^1 + \mu_5 A_5^1 + \mu_6 A_6^1 + \mu_7 A_7^1,\\
\hat{\alpha}_2 &= \mu_3 A_3^2 + \mu_4 A_4^2 + \mu_5 A_5^2 + \mu_6 A_6^2 + \mu_8 A_8^2,\\
\hat{\beta}_1 &= \mu_1 B_1^1 + \mu_3 B_3^1 + \mu_5 B_5^1 + \mu_7 B_7^1 + \mu_8 B_8^1,\\
\hat{\beta}_2 &= \mu_2 B_2^2 + \mu_4 B_4^2 + \mu_6 B_6^2 + \mu_7 B_7^2 + \mu_8 B_8^2,
\end{aligned}$$

где

$$\begin{aligned}
M_1 = A_5^2 = B_7^2 &= J^{-1}(-n_{22}^{-1}\bar{x}_{11} + n_{22}^{-1}\bar{x}_{12} + n_{22}^{-1}\bar{x}_{21} + (n_{11}^{-1} + n_{12}^{-1} + n_{21}^{-1})\bar{x}_{22}),\\
M_2 = A_6^2 = B_7^1 &= J^{-1}(n_{21}^{-1}\bar{x}_{11} - n_{21}^{-1}\bar{x}_{12} + (n_{11}^{-1} + n_{12}^{-1} + n_{22}^{-1})\bar{x}_{21} + n_{21}^{-1}\bar{x}_{22}),\\
M_3 = A_5^1 = B_8^2 &= J^{-1}(n_{12}^{-1}\bar{x}_{11} + (n_{11}^{-1} + n_{21}^{-1} + n_{22}^{-1})\bar{x}_{12} - n_{12}^{-1}\bar{x}_{21} + n_{12}^{-1}\bar{x}_{22}),\\
M_4 = A_6^1 = B_8^1 &= J^{-1}((n_{12}^{-1} + n_{21}^{-1} + n_{22}^{-1})\bar{x}_{11} + n_{11}^{-1}\bar{x}_{12} + n_{11}^{-1}\bar{x}_{21} - n_{11}^{-1}\bar{x}_{22}),\\
A_t^1 = -A_d^2 &= J^{-1}((n_{12}^{-1}+n_{22}^{-1})\bar{x}_{11}+(n_{11}^{-1}+n_{21}^{-1})\bar{x}_{12}-(n_{12}^{-1}+n_{22}^{-1})\bar{x}_{21}-(n_{11}^{-1}+n_{21}^{-1})\bar{x}_{22}),\\
B_s^1 = -B_g^2 &= J^{-1}((n_{21}^{-1}+n_{22}^{-1})\bar{x}_{11}-(n_{21}^{-1}+n_{22}^{-1})\bar{x}_{12}+(n_{11}^{-1}+n_{12}^{-1})\bar{x}_{21}-(n_{11}^{-1}+n_{12}^{-1})\bar{x}_{22}),
\end{aligned}$$

$$t = 1, 2, 7; d = 3, 4, 8; s = 1, 3, 5; g = 2, 4, 6.$$

Частный случай

Пусть $\mu_5 = \ldots = \mu_8 = 0$, $\mu_1 + \mu_2 + \mu_3 + \mu_4 = 1$,

$$J_{.k} = \frac{n_{1k}^{-1} + n_{2k}^{-1}}{n_{11}^{-1} + n_{12}^{-1} + n_{21}^{-1} + n_{22}^{-1}}, \quad J_{k.} = \frac{n_{k1}^{-1} + n_{k2}^{-1}}{n_{11}^{-1} + n_{12}^{-1} + n_{21}^{-1} + n_{22}^{-1}}, \quad k = 1, 2.$$

Получаем следующие оценки

$$\hat{\alpha}_1 = (\mu_1 + \mu_2) J_{.1} J_{.2} \left(\frac{\bar{x}_{11} - \bar{x}_{21}}{J_{.1}} + \frac{\bar{x}_{12} - \bar{x}_{22}}{J_{.2}} \right),$$

$$\hat{\alpha}_2 = (\mu_3 + \mu_4) J_{.1} J_{.2} \left(\frac{\bar{x}_{21} - \bar{x}_{11}}{J_{.1}} + \frac{\bar{x}_{22} - \bar{x}_{12}}{J_{.2}} \right),$$

$$\hat{\beta}_1 = (\mu_1 + \mu_3) J_{1.} J_{2.} \left(\frac{\bar{x}_{11} - \bar{x}_{12}}{J_{1.}} + \frac{\bar{x}_{21} - \bar{x}_{22}}{J_{2.}} \right),$$

$$\hat{\beta}_2 = (\mu_2 + \mu_4) J_{1.} J_{2.} \left(\frac{\bar{x}_{12} - \bar{x}_{11}}{J_{1.}} + \frac{\bar{x}_{22} - \bar{x}_{21}}{J_{2.}} \right).$$

При $n_{ij} = m$ и $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 1/4$ получаем известные оценки

$$\hat{\mu} = \bar{x} + \frac{(\mu_1 - \mu_4)(\bar{x}_{22} - \bar{x}_{11})}{2} + \frac{(\mu_2 - \mu_3)(\bar{x}_{21} - \bar{x}_{12})}{2} \Rightarrow \hat{\mu} = \bar{x},$$

$$\hat{\alpha}_1 = \bar{x}_{1.} - \bar{x}, \quad \hat{\beta}_1 = \bar{x}_{.1} - \bar{x},$$

$$\hat{\alpha}_2 = \bar{x}_{2.} - \bar{x}, \quad \hat{\beta}_2 = \bar{x}_{.2} - \bar{x}.$$

Итоги

1. На языке «R» реализован алгоритм обобщенного обращения матриц с использованием барицентрической параметризации.
2. Сформулированы и доказаны утверждения:
   - об инвариантности обобщенной обратной матрицы при барицентрической параметризации относительно выбора невырожденного минора максимального ранга.
   - об инвариантности оценок параметров относительно барицентрических параметров левого обращения.
3. Изучено влияние параметров обращения на проверку гипотез в случае однофакторного дисперсионного анализа и получен обобщенный критерий для проверки гипотезы равенства средних.
4. Получены оценки параметров модели двухфакторного дисперсионного анализа через параметры обобщенной обратной к информационной матрице.