Многомерный анализ повторяющихся неполных наблюдений

Конищева Злата Олеговна, гр. 19.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Научный руководитель: к. ф.-м. н., доцент Алексеева Н. П.
Рецензент: д. ф.-м. н., Ананьевская П. В.

Санкт-Петербург, 2023

Саногенез — комплекс защитно-приспособительных механизмов, направленный на восстановление нарушенной саморегуляции организма.

Модель

Функция, предложенная [Барт, 2003] в качестве аналитической модели кривой саногенеза:

$$S(t) = e^{-\eta t} \cos \tau t$$

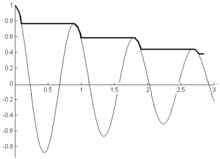


Рис.: Функция $S(t) = e^{-\eta t} \cos \tau t$ и ее двойная правая обратная.

Постановка задачи

Кривая саногенеза $S(t) = e^{-\eta t} \cos \tau t$ может быть рассмотрена в качестве корреляционной функции марковского процесса

Задачи

- Рассмотреть модель корреляционной структуры КМНС процесса и построить оценки параметров кривой саногенеза по методу моментов
- Рассмотреть другую модель КНС процесса так же оценить параметры
- На тех же моделях корреляционной структуры построить оценки параметров по ОМП
- Провести сравнение найденных оценок параметров

Комплексный процесс

- u и v — некоррелированные признаки с повторными наблюдениями по n индивидам и k временным точкам.
- X — комплексный n-мерный вектор наблюдений вида:

$$x_j(t) = u_j(t) + iv_j(t),\ j \in [1, n],\ t \in [1, k].$$

- Вещественная часть ковариационной функции X:

$$\mathcal{B} = \sigma^2 e^{-\eta|t|} cos(\tau t), \eta > 0.$$

Коварицонная матрица

- $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^{\mathrm{T}}$ — вектор средних X.
- Ковариационная матрица X [Бриллинджер, 1980]:

$$\Sigma = \mathbf{E}(X - \mu)\overline{(X - \mu)}^{\mathrm{T}}.$$

- Плотность распределения X:

$$f(X) = \frac{1}{\pi^n |\Sigma|} exp(-\overline{(X - \mu)}^{\mathrm{T}} \Sigma^{-1} (X - \mu)).$$

Рассмотрим корреляционную структуру вида:

$$R = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{k-1} \\ \rho^* & 1 & \rho & \rho^2 & \dots & \rho^{k-2} \\ \rho^{*2} & \rho^* & 1 & \rho & \dots & \rho^{k-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{*k-1} & \rho^{*k-2} & \rho^{*k-3} & \rho^{*k-4} & \dots & 1 \end{pmatrix}.$$

Тогда оценка параметра r будет считаться следующим образом:

$$\widehat{r} = \frac{2}{k(k-1)} \sum_{i=1}^{k} \sum_{j=i+1}^{k-1} \widehat{r}_{ij}^{\frac{1}{j-i}},$$

где $\widehat{r}_{ij}$ — выборочный коэффициент корреляции между i-ой точкой наблюдения и j-ой, k — количество точек наблюдения.

Метод моментов. Вторая модель

Рассмотрим альтернативную корреляционную структуру вида:

$$R = \begin{pmatrix} 1 & \rho & \rho & \rho & \dots & \rho \\ \rho^* & 1 & \rho & \rho & \dots & \rho \\ \rho^* & \rho^* & 1 & \rho & \dots & \rho \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^* & \rho^* & \rho^* & \rho^* & \dots & 1 \end{pmatrix}.$$

Тогда оценка параметра r будет считаться следующим образом:

$$\widehat{r} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \widehat{r}_{ij},$$

где $\widehat{r}_{ij}$ — выборочный коэффициент корреляции между i-ой точкой наблюдения с j-ой, k — количество точек наблюдения.

Оценки $\widehat{\eta}$ и $\widehat{\tau}$

На основании полученной оценки параметра $\widehat{r}$ далее можно посчитать оценки параметров τ и η методом моментов выразив их из обозначений $r = e^{-\eta}\left(\cos\tau - i\ \sin\tau\right)$ [Алексеева, 2012].

Таким образом получим в явном виде оценки $\widehat{\eta}$ и $\widehat{\tau}$:

$$\hat{\eta} = -\ln|\hat{r}|,$$

$$\hat{\tau} = \arctan\frac{\operatorname{Im}\hat{r}}{\operatorname{Re}\hat{r}} + 2\pi n, n \in \mathbb{Z}.$$

Метод максимального правдоподобия. Первая модель

Для оценки по методу максимального правдоподобия найдем определитель корреляционной структуры и ее обратную матрицу. Этот этап представлен в работе [Алексеева, 2012].

Определитель для первой модели

$$\det R = (1 - \rho\rho^*)^{k-1}.$$

Обратная матрица для первой модели

$$R^{-1} = \frac{1}{1 - \rho\rho^*} \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho^* & 1 + \rho\rho^* & -\rho & 0 & \cdots & 0 \\ 0 & -\rho^* & 1 + \rho\rho^* & -\rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Метод максимального правдоподобия. Первая модель

С учетом определителя и обратной матрицы корреляционной структуры функцию правдоподобия можно записать в виде [Алексеева, 2012]:

$$\mathcal{L}(\tau, \eta, \sigma^2) = (\pi\sigma^2)^{-N}(1 - \rho\rho^*)^{-N+n} exp(Q),$$

где

$$N = nk,$$

$$Q = \frac{-A_1 - rr * A_2 + rA_3 + r * A_3^*}{\sigma^2(1 - \rho\rho^*)},$$

$$A_1 = \sum_{j=1}^{n} \sum_{i=1}^{k} x_j(i)^* x_j(i), A_2 = \sum_{j=1}^{n} \sum_{i=2}^{k-1} x_j(i)^* x_j(i),$$

$$A_3 = \sum_{j=1}^{n} \sum_{i=1}^{k-1} x_j(i)^* x_j(i+1).$$

Метод максимального правдоподобия. Вторая модель

Утверждение. Определитель второй модели

$$\det R = 1 + \sum_{i=0}^{k-2} (-1)^{k+i+1} C_k^i \sum_{l=1}^{k-i-1} \rho^l \rho^{*k-i-1}.$$

Введем замену:

$$a = \rho - 1,\ b = \rho^* - 1.$$

Утверждение. Обратная матрица второй модели

$$R^{-1} = \frac{1}{\det R_k} \begin{pmatrix} \det R_{k-1} & r b^{k-2} a^0 & \cdots & r b^0 a^{k-2} \\ r^* a^{k-2} b^0 & \det R_{k-1} & \cdots & r b^1 a^{k-3} \\ \vdots & \vdots & \ddots & \vdots \\ r^* a^0 b^{k-2} & \cdots & r^* a^{k-2} b^0 & \det R_{k-1} \end{pmatrix}.$$

Метод максимального правдоподобия. Вторая модель

С учетом определителя и обратной матрицы корреляционной структуры функцию правдоподобия можно записать в виде:

$$\mathcal{L} = \frac{1}{\pi^N (\det R_k)^n} \exp(Q),$$

где

$$Q = \frac{-\sum_{j=0}^{k-2} \rho b^{k-2-j} a^j \ A_1 - \sum_{j=0}^{k-2} \rho^* a^{k-2-j} b^j \ A_2 - \det R_{k-1} \ A_3}{\det R_k}.$$

Статистики имеют вид:

$$A_1 = \sum_{m=0}^{k-2} \sum_{i=1}^{n} \sum_{j=1}^{k-1-m} x_{ij}^* x_{i\ j+m+1}, \quad A_2 = \sum_{m=0}^{k-2} \sum_{i=1}^{n} \sum_{j=1}^{k-1-m} x_{ij} x_{i\ j+m+1}^*,$$

$$A_3 = \sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}^* x_{ij}.$$

Применение метода на реальных данных

Пример решения задачи оценки параметров кривой саногенеза производится на многолетние кардиологические данные из ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России.

Используемые данные

В течение определенного периода были повторно собраны различных метрики о людях с сердечными заболеваниями.

- 23 индивида.
- 6 временных точек.

Выбор u и v

- RA — размер правого предсердия в мм.
- RV — размер правого желудочка в мм.

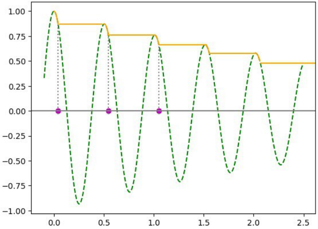
Оценки параметров $\widehat{\tau}$ и $\widehat{\eta}$ методом моментов

Таблица: Доверительные интервалы для вещественной и мнимой части оценки $\hat{r}$, найденной методом моментов

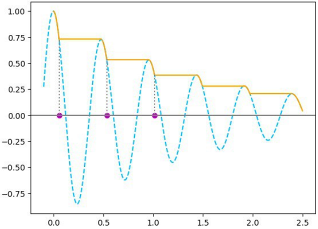
Модель	$\hat{r}$	$\mathrm{Re}\,\hat{r}$	$\mathrm{Im}\,\hat{r}$
Первая модель	$0.44 + \mathbf{i}0.26$	$(-0.58, 0.92)$	$(-0.69, 0.89)$
Вторая модель	$0.35 + \mathbf{i}0.06$	$(-0.45, 0.9)$	$(-0.59, 0.83)$

Таблица: Значение ошибки MSE полученных оценок методом моментов $\hat{r}$ и реальных данных

Модель	$\hat{r}$	MSE
Первая модель	$0.44 + \mathbf{i}0.26$	0.061
Вторая модель	$0.35 + \mathbf{i}0.06$	0.053



(a) Марковская



(b) Альтернативная

Рис.: Графики кривых саногенеза построенных по первой и второй корреляционным структурам. Желтый цвет - кривая дожития, зеленый/синий - кривые саногенеза, красный - критические точки.

Таблица: Абсолютная разница значений критических точек, полученных из анализа кривых дожития

Первая точка	Вторая точка	Третья точка
0.016	0.011	0.037

Заключение

- Были рассмотрены две корреляционные структуры: марковская и альтернативная.
- Проведена оценка параметров модели с использованием методов моментов и оценки максимального правдоподобия.
- Проведен сравнительный анализ результатов для двух корреляционных структур.
- Результаты эксперимента показали, что разница между марковской и альтернативной корреляционными структурами незначительна, что указывает на качество альтернативной структуры.

Список литературы

- Алексеева Н. П. Анализ медико-биологических систем. — Издательство Санкт-Петербургского государственного университета, 2012.
- Барт А. Г. Анализ медико-биологических систем. — Издательство Санкт-Петербургского государственного университета, 2002.
- Бриллинджер Д. Временные ряды. Обработка данных и теория. — Издательство «МИР», 1980.
- Тихов М.С. Современные методы оценивания статистических параметров. — Нижегородский государственный университет им. Н.И. Лобачевского, 2016.
- Ф.З. Меерсон. Адаптация, стресс, профилактика. — Издательство «Наука», 1981. — 279 с.