Оценка метода максимального правдоподобия. Решение задачи 24

Яковлев Д.М.
ST095998@STUDENT.SPBU.RU

29 марта 2025 г.

Постановка задачи об интервальном цензурировании смешанного типа [2].
(Интервальное цензурирование смешанного типа) Пусть

- $K : (\Omega, \mathcal{F}, P) \to (\mathbb{Z}, \mathcal{B}_{\mathbb{R}})$ — положительная целочисленная случайная величина;
- T — набор случайных величин $\{T_{k,j},\ j = 1, \dots, k;\ k = 1, \dots, +\infty\}$ таких, что $0 = T_{k,0} < T_{k,1} < \dots < T_{k,k} < T_{k,k+1} = +\infty$;
- $X,\ (K, T)$ — независимы;
- $Y = (\Delta_K, T_K, K)$, где T_k — k-я строка треугольного массива T, $\Delta_k = (\Delta_{k,1}, \dots, \Delta_{k,k+1})$ и $\Delta_{k,j}$ — индикатор события $X \in (T_{k,j-1}, T_{k,j}]$;
- Случайная величина X имеет распределение Вейбулла:

$$X \sim W(k, \lambda) \leftrightarrow p_X(x) = \begin{cases} \dfrac{k}{\lambda} \left(\dfrac{x}{\lambda}\right)^{k-1} e^{-\left(\dfrac{x}{\lambda}\right)^k}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Исследуйте зависимость ширины доверительного интервала для оценок параметров от объёма выборок.

Задание можно разделить на несколько частей:

1. Моделирование выборки из n независимых реализаций Y;
2. Написание функции правдоподобия;
3. Написание процедуры оценивания параметров (с учётом естественных ограничений на области допустимых значений параметров);
4. Исследование указанных свойств оценок параметров.

1 Моделирование Y

Итак, мы знаем вид случайной величины Y. В условиях задачи было сказано, что можно положить

$$T_{k,j} = \sum_{i=1}^{j} Z_i, \quad K = \sup_{j \geqslant 1} \left\{ \sum_{i=1}^{j} Z_i < T \right\},$$

где T — константа, а $\{Z_i\}$ — независимые одинаково распределённые случайные величины. Поскольку в Y также имеются Δ_k, то $\{Z_i\}$ должны быть неотрицательными.

Естественный пример, который приходит в голову — моделлирование пуассоновского процесса. Известно, что если

$$Z_i \sim U(0, T) \leftrightarrow Z_{(1)}, \dots, Z_{(k)},$$

где Z_i упорядочены в порядке возрастания равносильны пуассоновскому процессу с k скачками в интервале $(0, T)$. Таким образом, случайная величина Y — распределение пуассоновского процесса при фиксированном времени $t = T$:

$$K \sim \Pi(\lambda T) \leftrightarrow P(K = k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}.$$

Поскольку K принимает значения от 0, имеет смысл рассматривать случайную величину $K + 1$. Пришли к следующему моделированию выборки:

for $i = 1 \to n$ **do**
    $k \leftarrow K(\cdot)$
    $Z \leftarrow U(0, T; k)$
    Сортировка по возрастанию Z
    Сохранение выборки с учётом k и Z
end for

2 Функция правдоподобия L

Для построения будем пользоваться материалами [1]. Функция правдоподобия будет выглядеть следующим образом

$$L(k, \lambda | Y_1, ..., Y_n) = \prod_{i=1}^{n} P(X_i \in \Delta_{K_i, j_i}) = \prod_{i=1}^{n} \left(\exp\left[-\left(\frac{T_{K_i, j_i - 1}}{\lambda} \right)^k \right] - \exp\left[-\left(\frac{T_{K_i, j_i}}{\lambda} \right)^k \right] \right),$$

То есть, у нас есть информация о том, какие интервалы цензурирования есть у Y_i и то, в какой из интервалов попадает X_i. Поэтому можно генерировать двойки (Y_i, X_i), а сохранять то, в какой интервал попадает i-ое значение выборки.

3 Процедура оценивания параметров

Использование встроенной функции optim с методом Nelder-Mead (так как я его реализовывал в прошлом семестре).

4 Исследование свойств оценки параметров

Исследуйте зависимость ширины доверительного интервала для оценок параметров от объёма выборок.

Список литературы

[1] Звонарёв Н. М-оценки (M-estimators). — 2024. — [Online: accessed 16-March-2025]. Режим доступа: `https://statmod.ru/wiki/_media/study:spring2024:slides.pdf`.

[2] Коробейников А. Оценки метода максимального правдоподобия. — 2023-03-07. — [Online: accessed 16-March-2025]. Режим доступа: `https://statmod.ru/wiki/_media/study:spring2023:compstat:mle.pdf`.