Метод существенной выборки для оценивания границ доверительных интервалов в задачах параметрической нелинейной регрессии с неоднородным шумом

Смирнов Иван Александрович, 22.М03-мм

Санкт-Петербургский Государстенный Университет
Математика-механический факультет
Кафедра статистического моделирования

Научный руководитель: д.ф.-м.н., п. Ермаков М.С.
Рецензент: к.ф.-м.н., с.н.с. Проурзин В.А.

Санкт-Петербург 2024г.

Введение

Суть работы:

- Исследование метода существенной выборки для построения границ доверительного интервала в задаче оценивания параметра нелинейной регрсессии с неоднородным гаусовским шумом

Что было сделано?

- Построена асимптотически эффективная процедура метода существенной выборки;
- Доказана ее эффективность;
- Метод существенной выборки исследован на примере оценки параметра в формуле суммарной намагниченности ферромангетика.

Постановка задачи

Пусть наблюдаются величины $x_i \in R^1$ следующего вида:

$$x_i = S(t_i, \theta) + \sigma_i \xi_i, \text{ при } 1 \le i \le n \tag{1}$$

где $S(t, \theta)$ — нелинейная функция,
$t_1, ..., t_n$ — n точек равномерно взятых на отрезке $[0, 1]$,
$\theta = (\theta^1, ..., \theta^l)$ — вектор неизвестных параметров,
$\sigma_i \xi_i$ — неоднородный гауссовский шум.
Истинное значение вектора неизвестных параметров будем обозначать θ_0.
Цель работы – исследовать метод существенной выборки для оценивания вероятностей умеренных уклонений:

$$V_n = P((\hat{\theta}_n - \theta_0) > b_n). \tag{2}$$

Метод существенной выборки

Произведем замену меры для метода существенной выборки. Обозначим выбранную плотность $g_i(x)$. Моделируем k независимых выборок, где $Y_j^{(i)}$ имеет плотность g_j.

$$Y_1^{(i)}, \cdots, Y_n^i, 1 \leq i \leq k.$$

Оценка $V_n = P((\hat{\theta}_n - \theta_0) > b_n)$ имеет вид:

$$\hat{V}_n = \frac{1}{k} \sum_{i=1}^{k} \chi(\hat{\theta}_n > \theta_0 + b_n) \prod_{j=1}^{n} q_n^{-1}(y_i^{(j)}), \tag{3}$$

где $q_n(y_i^{(j)}) = \frac{g_i(y_i^{(j)})}{p_{i,\theta_0}(y_i^{(j)})}$ – нормирующий множитель.

В примере численного моделирования будет показано, что если выбрать замену $g_i(x) = p_{i,\theta_n}(x) = f(x - S(t_i, \theta_0 + b_n))$, то в большем числе случаев индикатор будет ненулевым

Асимтотическая эффективность процедуры существенной выборки

Математическое ожидание оценки (3) (оценки вероятности уклонения):

$$\omega_n = E\hat{V}_n = V_n.$$

Дисперсия оценки:

$$\mathrm{Var}[\hat{V}_n] = U_n - \omega_n^2, \tag{4}$$

где

$$U_n = E\left(\frac{1}{k}\sum_{i=1}^{k}\chi(\hat{\theta}_n^{(i)} - \theta_0 > b_n)\prod_{j=1}^{n} q_n(y_i^{(j)})\right)^2.$$

Определение

Процедура называется асимтотически эффективной (в смысле логарифмической асимптотики), если

$$\lim_{n\to\infty}\frac{\log U_n}{2\log\omega_n} = 1.$$

Основной результат в виде теормы

Теорема

Пусть выполняются следующие условия:

1. $nb_n^2 \to \infty, nb_n^{2+\alpha} \to 0$ при $n \to \infty$.
2. $\sup_{t,\theta}|S(t,\theta)| < C$.
3. Для $\forall t$ и $\forall \theta$ функция $S(t,\theta)$ дифференцируема по θ и верно соотношение $|S(t,\theta+b_n) - S(t,\theta) - b_n S'_\theta(t,\theta)| < cb_n^{1+\frac{\alpha}{2}}$.
4. Функция $\sigma(t_i)$ ограничена и $\sigma(t_i) > c > 0$ для всех $t_1, ..., t_n$, где c – const.

тогда процедура существенной выборки с заменой меры $g_i(x) = f(x - S(t_i, \theta_0 + b_n)))$ является асимтотически эффективной.

Численное моделирование

Проведем численное моделирование на примере модели

$$x_i = c \cdot \mathrm{th}(\theta t_i) + \sigma_i \xi_i \tag{5}$$

где c — некоторая константа,
$c \cdot \mathrm{th}(\theta t_i)$ — криволинейная функция,
$\sigma_i \xi_i$ — неоднородный гаусовский шум.
Модель описывает суммарную намагниченность парамагнетика.
Оценку θ будем вычислять методом наименьших квадратов:

$$Q(\theta) = \sum_{i=1}^{n} \omega_i (S(t_i, \theta) - y_i)^2 \to \min_{\theta}.$$

где ω_i — вес i-ого параметра.

Численное моделирование. Зависимость от k

Соответсвие x и b_n:

x	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5
b_n	0.23	0.29	0.35	0.42	0.48	0.50	0.53	0.56	0.61	0.67

Изобразим полученные оценки при разных k на рисунке 1.

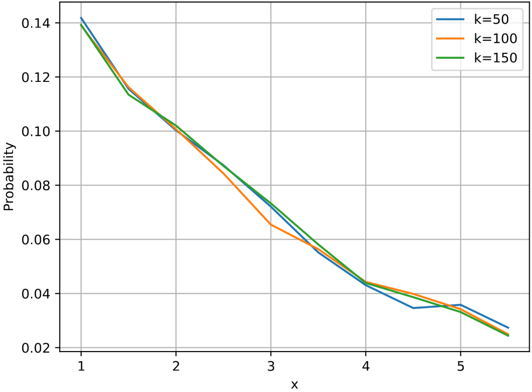


Рис. 1: Оценки вероятности при $k = 50, 100, 150$

Численное моделирование. Зависимость от n

Изобразим полученные оценки при разных n на рисунке 2.

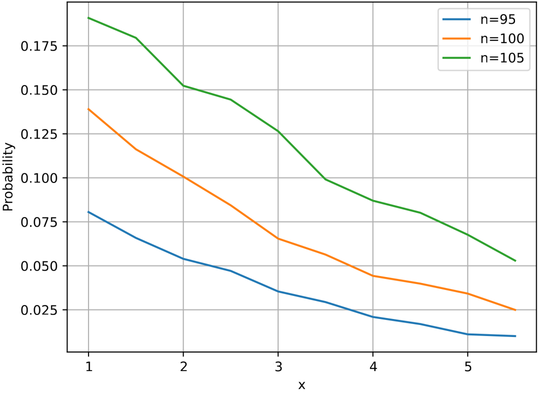


Рис. 2: Оценки вероятности при $n = 95, 100, 105$

Численное моделирование. Доверительный интервал

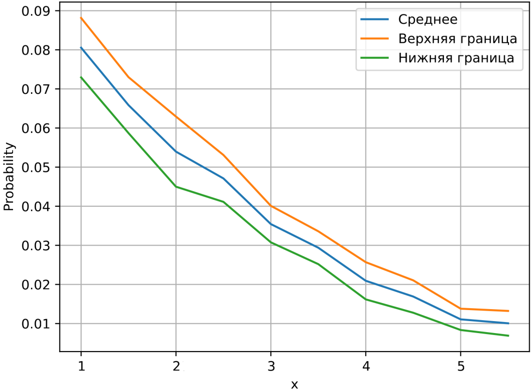
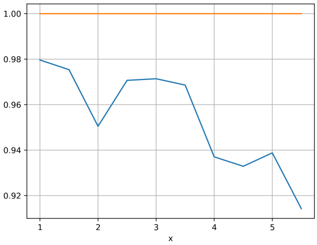
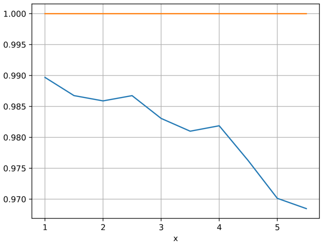


Рис. 3: Доверительные интервалы оценки при $k = 100, n = 95$

Численное моделирование. Асимптотическая эффективность



(a) k = 100, n = 95



(b) k = 100, n = 105

Рис. 4: Асимптотическая эффективность

Заключение:

Полученные результаты:

1. Был применен метод вычисления вероятности редких событий (метод существенной выборки) для оценивания доверительных интервалов в задачах параметрической нелинейной регрессии с неоднородным шумом;
2. Доказана асимптотическая эффективность предложенного варианта процедуры существенной выборки в зоне вероятностей умеренных уклонений;
3. Осуществлена численная реализация процедуры существенной выборки, в результате которой построены доверительные интервалы для оценки вероятности и сосчитана её эффективность. Численное моделирование показало, что метод хорошо применим в зоне малых вероятностей (вероятностей умеренных уклонений).