Метод существенной выборки для доверительного оценивания некоторых параметров пуассоновского процесса

Веселова Влада Валерьевна, гр.18.Б04-мм

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

Научный руководитель: д.ф.-м.н., профессор М.С.Ермаков
Рецензент: Старший научный сотрудник, ИПМаш РАН
к. ф.-м. н. В. А. Проурзин

Санкт-Петербург, 2022

Введение

В задаче оценивания параметров модели часто возникает вопрос нахождения вероятности редкого события.

При вычислении малых вероятностях возникают сложности:

- Большая трудоемкость;
- Независимость псевдослучайных чисел на множествах, имеющих малую вероятность, сложно проверить.

Метод существенной выборки позволяет обойти эти сложности.

Введение: Пуассоновский процесс

Функция интенсивности — $\lambda : \mathbb{R}^d \to [0, \infty)$ такая, что для любого ограниченного множества B: $\Lambda(B) = \int\limits_B \lambda(x)dx < \infty$.

Пуассоновский процесс — функция $X(\cdot)$ такая, что для любых непересекающихся ограниченных борелевских множеств $B_1, \ldots, B_m$ случайные величины $X(B_1), \ldots, X(B_m)$ независимы и распределены по закону Пуассона:
$P(X(B) = k) = \dfrac{\Lambda(B)^k}{k!} \exp(-\Lambda(B))$.

В работе рассматривается следующая функция интенсивности:

$$\lambda(a, x) = \frac{a}{2}[1 + \cos(\omega x + \varphi)] + \lambda_0.$$

Введение: оценка минимального расстояния

Оценка минимального расстояния параметра θ процесса Пуассона $X(\cdot)$ с интенсивностью Λ_θ на семействе множеств $\{A_t,\ t \in [0,1]\}$ — это $\arg\inf_\theta \int\limits_0^1 (X(A_t) - \Lambda_\theta(A_t))^2 \mu(dt)$, где μ — некоторая конечная мера на $[0,1]$.

Пусть X_i, $i = 1, \ldots n$ — наблюдаемые события пуассоновского процесса с функцией интенсивности $\lambda(a, x)$.

Оценка минимального расстояния параметра амплитуды:

$\hat{a}_n = 6 \int\limits_0^1 \left[\frac{1}{N} N([0, Nt]) - \lambda_0 t\right] t\ dt\ (1 + O(N^{-1}))$ при $N \to \infty$.

1. Состоятельная;
2. Асимптотически нормальная:
   $\mathcal{L}_a\{\sqrt{N}(\hat{a}_n - a)\} \Rightarrow N(0, \sigma^2)$, $\sigma^2 = 2.4a\left(1 + 2\frac{\lambda_0}{a}\right)$.

Пусть $\mathbb{X} = X_1, \ldots, X_n$ имеет плотность распределения f_0 — наблюдаемый на интервале $[0, N]$ пуассоновский процесс интенсивности Λ_0.
Интересна ситуация: $m = \Lambda_0([0, N]) \to \infty, \ N \to \infty$.

Задача: оценить вероятность $\omega_{n,m} = P((\hat{a}_{n,m} - a) > b_m)$

1. $\hat{a}_{n,m}$ — оценка минимального расстояния параметра a;
2. $\{b_m\}$ — некоторая последовательность;
3. Возьмем $b_m = \dfrac{\sqrt{D\hat{a}_{n,m}}}{\sqrt{n}} x$,
   $x = x_\gamma$ — квантиль стандартного нормального распределения.

Будем моделировать k независимых выборок $\mathbb{Y}_i = y_{i1}, \ldots, y_{in}$, $i = 1, \ldots, k$ с плотностью распределения f_n.

В качестве оценки вероятности ω_m возьмем

$$\hat{\omega}_{n,m} = \frac{1}{k} \sum_{i=1}^{k} \mathbb{I}_{\{(\hat{a}_{n,m} - a) > b_m\}} \prod_{j=1}^{n} \frac{f_0(y_{ij})}{f_n(y_{ij})} \cdot \frac{\Lambda_0([0,N])^n e^{-\Lambda_0([0,N])}}{\Lambda_n([0,N])^n e^{-\Lambda_n([0,N])}},$$

Λ_n — интенсивность пуассоновского процесса с плотностью f_n.

В работе показана несмещенность данной оценки.

Асимтотическая эффективность метода существенной выборки

Процедура существенной выборки называется асимптотичеки эффективной, если

$$\overline{\lim_{m\to\infty}} \frac{\ln S_{n,m}}{\ln \omega_{n,m}} = 1,$$

где

$$S_{n,m}^2 = E \sum_{n=1}^{\infty} \mathbb{I}_{\{(\hat{a}_{n,m}-a)>b_m\}} \prod_{j=1}^{n} \frac{f_0^2(y_{ij})}{f_n^2(y_{ij})} \cdot \left(\frac{\Lambda_0([0,N])^n e^{-\Lambda_0([0,N])}}{\Lambda_n([0,N])^n e^{-\Lambda_n([0,N])}} \right)^2 .$$

В работе будет построена асимптотически эффективная процедура существенной выборки.

Постановка задачи

Цели работы:

1. Написать алгоритм построения доверительного интервала для оценки параметра a — амплитуды неоднородного пуассоновского процесса, используя асимптотически эффективный метод существенной выборки;
2. Проверить качество работы алгоритма при разных мешающих факторах модели.

Результаты: моделирование пуассоновского процесса

В работе используется метод моделирования, основанный на утверждении:
$\lambda(\cdot)$ — неотрицательная функция на $\mathbb{R}^d$. Неоднородный пуассоновский процесс с функцией интенсивности λ в области $A \subset \mathbb{R}^d$ может быть получен в два этапа:

1. Выбирается случайное число N из распределения Пуассона с параметром $\Lambda(A) = \int_A \lambda(s)ds$;
2. Выбирается N точек на A, плотность вероятности которых пропорциональна $\lambda(s)$.

Первый этап выполняется с помощью метода обратных функций.
Второй этап выполняется с помощью метода отбора с мажорирующей плотностью равномерного распределения.

Результаты: Схема численного моделирования

1. Фиксируем интервал $[0, N]$, на котором будет производиться моделирование;
2. Вычисляем $m = \Lambda_0([0, N])$;
3. Моделируем выборку $\mathbb{Y}_i$ с плотностью

$$f_n(x) = f_0(x) + \frac{b_m}{\sqrt{m}\sigma^2} g(x),$$

где

$$f_0(x) = \frac{\lambda(x)}{\int\limits_0^N \lambda(s) ds}; \quad g(x) = 3(1 - x^2);$$

4. Вычисляем оценку $\hat{a}_{n,m} = 6\int\limits_0^1 (\frac{n}{m} - \lambda_0 t)tdt$;

5. Вычисляем

$$\prod_{j=1}^{n} \frac{f_0(y_{ij})}{f_n(y_{ij})} \cdot \frac{\Lambda_0([0,N])^n e^{-\Lambda_0([0,N])}}{\Lambda_n([0,N])^n e^{-\Lambda_n([0,N])}};$$

6. Повторяем пункты 3–5 k раз;

7. Вычисляем $\hat{\omega}_{n,m}$.

Оценка параметра a и доверительные интервалы для нее в зависимости от размера выборки k для разных наборов параметров и длиной промежутка моделирования $N = 2$.

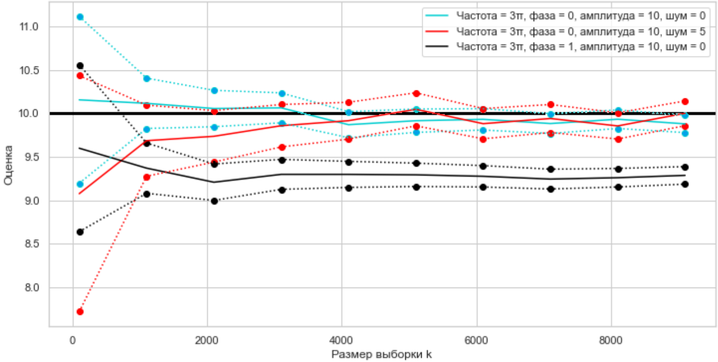


Рис.: Доверительные интервалы для оценки a уровня 0.05

График зависимости оценки вероятности $\hat{\omega}_{n,m}$ от выбранного значения квантиля x. Размер выборки $k = 1000$.

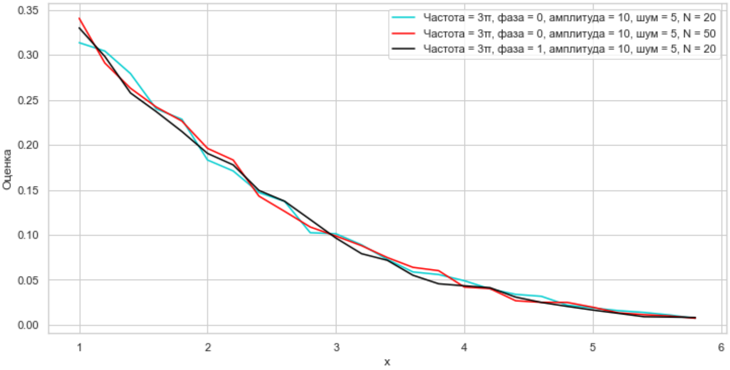
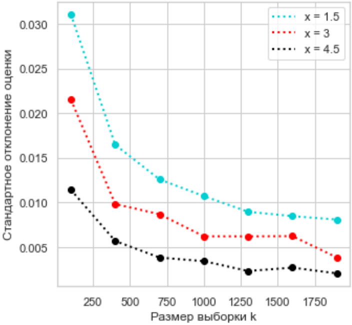


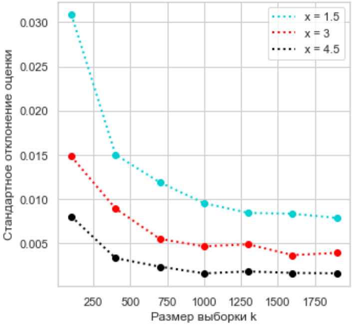
Рис.: Оценка вероятности в зависимости от x, $k = 1000$

Результаты

График стандартного отклонения оценки $\hat{\omega}_{n,m}$ в зависимости от размера выборки k для разных значений x.
$\varphi = 1$, $\lambda_0 = 5$, $\omega = 3\pi$, $a = 10$.



$N = 20$



$N = 2$

Рис.: Графики стандартного отклонения, размер выборки оценок вероятностей $= 50$

Заключение

- Получены оценки истиных вероятностей границ доверительных интервалов для параметра амплитуды;
- Показано, что при различных параметрах модели можно оценивать границы доверительных интервалов параметра при достаточно больших x и N с приемлемой точностью;
- Других исследований, связанных с вычислениями малых вероятностей для пуассоновских процессов, насколько нам известно, не проводилось.