БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра математического моделирования и анализа данных

Отчет

о прохождении производственной (преддипломной) практики

Толочко Александра Викторовича

студента 4 курса,
специальность Прикладная математика»

Руководитель практики:

кандидат физ.-мат. наук,
доцент И.А. Бодягин

Минск, 2025

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра ________________________________
(наименование кафедры)

Утверждаю

Заведующий кафедрой ________________________
(подпись)(фамилия, инициалы)

«__»____________20__г.

Задание на практику

по специальности «Прикладная математика»

Студенту ________________________
(фамилия, инициалы)

1. Тема практики: ____
(наименование темы дипломной работы)

2. Список рекомендуемой литературы:

2.1. Андерсон, Т. Статистический анализ временных рядов

2.2. Дженкинс, Г., Ваттс, Д. Спектральный анализ и его приложения: в 2 т.

2.3. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление

2.4. Харин, Ю.С. Оптимальность и робастность в статистическом прогнозировании

3. Перечень подлежащих разработке вопросов или краткое содержание расчетно-пояснительной записки:

3.1. Исследовать применимость метода игнорирования пропусков в случае пропусков, зависящих от данных

3.2. Показать, как модель неслучайных пропусков влияет на работу стандартных методов оценивания параметров

3.3. Получить верные оценки параметров распределения в случае пропусков, зависящих от данных

4. Примерный календарный график:

– **февраль (1-ая неделя)** – ознакомление с условиями работы, изучение основных теоретических вопросов, получение задания.

– **февраль (2-3-я неделя)** – построение модели для описания различных видов пропусков.
– **март (4-5-ая неделя)** – исследование метода максимального правдоподобия для получения оценки в случае пропусков, зависящих от данных.
– **март (6-7-ая неделя)** – исследование метода моментов для получения оценки в случае пропусков, зависящих от данных.
– **апрель (8-9-ая неделя)** – проведение компьютерных экспериментов для проверки полученных результатов.
– **апрель (10-ая неделя)** – обобщение результатов и подготовка отчета

5. Руководители практики:

от предприятия ________________________________
(ФИО)

от кафедры ________________________________
(ФИО)

6. Дата выдачи задания ________________

7. Срок сдачи отчета ________________

Руководитель_____________ ________________________
(от кафедры) (подпись) (инициалы, фамилия)

Подпись студента ________________________
(подпись, дата)

Содержание

Введение

Временной ряд представляет собой набор наблюдений, полученных путем регулярного измерения одной переменной в течение некоторого периода времени. Наблюдения представляют собой набор из одного или нескольких значений, зафиксированных в определенный момент времени. Элементами наблюдения являются действительные числа — значения непрерывных или дискретных переменных.

Временные ряды — один из наиболее важных инструментов в аналитике данных. Они используются во многих областях, включая финансы, производство, социальные и экономические исследования, климатологию и другие. Примеры временных рядов могут включать данные о продажах продукции в определенный день или данные о температуре на определенной территории в различные временные промежутки.

На практике зачастую часть значений переменных во временном ряду по какой-то причине отсутствует. Например, часть респондентов, участвующих в обследовании семей, может отказаться сообщить размер дохода. Пропуски также могут быть вызваны факторами, не связанными с самим экспериментальным процессом, например, ограничения или неисправности оборудования, собирающего данные.

Существуют множество видов пропусков, но в литературе наиболее часто встречаются следующие:

- Случайные пропуски — пропуски, не зависящие от данных и от самого эксперимента (второй пример из приведенных выше)
- Цензурирование — пропуски, зависящие от данных, при которых пропускаются лишь значения из определенной области. Цензурирование называется полным, если пропускаются все значения из области (в качестве примера можно взять исследование сроков наступления эффекта лекарства, в котором все значения, большие срока проведения эксперимента, будут пропущены). Цензурирование называется частичным если лишь часть значений из области пропускается (пример с размером дохода в опросе семей).

При работе с пропусками, зависящими от данных, могут возникнуть серьезные проблемы, так как в этом случае большинство рядовых методов, используемых для статистического анализа временных рядов, как игнорирование пропусков, оказываются неэффективными, что будет подтверждено в ходе курсового проекта. Поэтому необходимо использование иных методов, основывающихся на некоторой определенной модели пропусков или на более общей модели.

Математическая модель и постановка задачи

Пусть наблюдается выборка

$$X = (x_1, x_2, \dots, x_T), \quad x_i \in \mathbb{R},$$

из некоторого абсолютно непрерывного закона распределения вероятности с плотностью p_ξ, заданной с точностью до параметров. И пусть некоторые значения выборки пропущены, то есть известно, что на данном месте должно наблюдаться некоторое значение, но само оно неизвестно. В литературе чаще всего встречаются два основных вида пропусков во временных рядах по зависимости от данных:

1. Случайные пропуски (не зависят от данных)
2. Цензурирование (зависят от данных)

Обобщим следующим образом. Пусть нам дана выборка вида

$$X = (x_1, x_2, \dots, x_T), \quad x_i \in \mathbb{R} \cup \{NA\},$$

где NA — пропущенное значение. Также введем шаблон пропусков

$$Obs = (obs_1, obs_2, \dots, obs_T), obs_i \in \{0, 1\},$$

где $obs_i = 0$ значит, что $x_i = NA$, а $obs_i = 1$ значит, что $x_i \neq NA$, и заранее известную функцию вероятности пропуска элемента со значением x

$$m(x) = P\{obs = 0 \mid \xi = x\}, \quad x \in \mathbb{R}.$$

Таким образом, получим $P\{obs_i = 1\} = \int\limits_{-\infty}^{+\infty} p(x)(1 - m(x))\,\mathrm{d}x.$

В рамках вышеописанной модели *случайными пропусками* будем называть случай $m(x) = \text{const}$. *Полным цензурированием справа* будем называть случай

$$m(x) = \begin{cases} 0, & x \leq c, \\ 1, & x < c \end{cases}, \quad c \in \mathbb{R}.$$

Аналогично определим *полное цензурирование слева.* Кроме данных выше моделей будет рассматриваться модель с кусочно-постоянной функцией $m(x)$.

В рамках курсовой работы была поставлена задача: по имеющимся наблюдениям оценить параметры распределения выборки и показать, что стандартные методы, использующиеся для оценки параметров распределения по выборке без пропусков, не дают результатов в случае пропусков, зависящих от данных. Для проверки работы различных методов проводились компьютерные эксперименты.

Компьютерные эксперименты

Все компьютерные эксперименты проводятся по единому шаблону. Рассмотрим ход эксперимента.

Создается случайная выборка размера T с некоторым распределением p_ξ, зависящим от параметров. Затем моделируется некоторая ситуация пропусков данных, описанная функцией $m(x)$.

После происходит оценка параметров. В этом этапе содержатся основные отличия методов, испытываемых в курсовой работе, поэтому их будем рассматривать индивидуально.

Вышеописанная процедура проводится $n = 500$ раз для размеров

$$T = \overline{50, 100, \ldots, 450, 500}.$$

При этом для каждого T считается выборочное среднее оценки параметра

$$E\{\theta\} = \frac{1}{500}\sum_{k=1}^{500}\hat{\theta}_k$$

и выборочное среднее вариации оценки

$$E\{\mathrm{var}(\theta)\} = \frac{1}{500}\sum_{k=1}^{500}(\hat{\theta}_k - \theta)^2.$$

Метод игнорирования пропусков

Посмотрим, как себя покажет метод «игнорированя пропусков» в случае пропусков, зависящих от данных. Для этого сначала теоретически исследуем свойство несмещенности и асимптотической несмещенности оценки, полученной по этому методу, в случае цензурирования. Также исследуем зависимость смещения оценки от порога цензурирования c.

Смещение будем искать по определению:

$$b(n, \theta_0, c) = E_{\theta_0}\{\hat{a}\} - \mu,$$

где θ_0 — истинное значение параметра распределения, μ — матожидание x_i. Также обозначим за $p_\xi(x, \theta)$ плотность распределения элементов выборки. получим

$$b(n, \theta_0, c) = \int\limits_{\mathbb{R}^n} T(X) p_{\boldsymbol{\xi}}(X, \theta_0)\, \mathrm{d}X - \mu,$$

где

$$p_{\boldsymbol{\xi}}(X, \theta) = \prod_{i=1}^{n} p_\xi(x_i, \theta).$$

В случае цензурирования вид статистики $T(X)$ будет зависеть от значений выборки, поэтому представим пространство $\mathbb{R}^n$ как сумму областей

$$A_\alpha = \{X \in \mathbb{R}^n : x_i > c \Leftrightarrow i \in \alpha\}, \quad \alpha \subseteq \{1, \ldots, n\}.$$

Тогда по свойствам интеграла

$$E_{\theta_0}\{\hat{a}\} = \sum_{\alpha} \int\limits_{A_\alpha} T(X) p_{\boldsymbol{\xi}}(X, \theta_0)\, \mathrm{d}X.$$

Рассмотрим $A_\varnothing = \{X \in \mathbb{R}^n : x_1 \le c, \ldots, x_n \le c\}$. В ней статистика имеет вид

$$T(X) = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

$$\int\limits_{A_\varnothing} T(X) p_{\boldsymbol{\xi}}(X, \theta_0)\, \mathrm{d}X = \frac{1}{n} \sum_{i=1}^{n} \int\limits_{A_\varnothing} x_i p_{\boldsymbol{\xi}}(X, \theta_0)\, \mathrm{d}X =$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int\limits_{-\infty}^{c} p_\xi(x_1, \theta_0)\, \mathrm{d}x_1 \int\limits_{-\infty}^{c} \cdots \int\limits_{-\infty}^{c} x_i p_\xi(x_i, \theta_0)\, \mathrm{d}x_i \int\limits_{-\infty}^{c} \cdots \int\limits_{-\infty}^{c} p(x_n, \theta_0)\, \mathrm{d}x_n =$$

$$= \int\limits_{-\infty}^{c} x p(x, \theta_0)\, \mathrm{d}x \cdot \left(\int\limits_{-\infty}^{c} p(x, \theta_0)\, \mathrm{d}x \right)^{n-1}.$$

Теппрь рассмотрим $A_i = \{X \in \mathbb{R}^n : x_1 \leq c, \ldots, x_{i-1}\}$, где статистика принимает вид

$$T(X) = \frac{x_1 + \ldots + x_{i-1} + x_{i+1} + \ldots + x_n}{n-1},$$

и рассуждая аналогично, получаем

$$\int\limits_{A_i} T(X) p_{\boldsymbol{\xi}}(X, \theta_0)\,\mathrm{d}X = \int\limits_{-\infty}^{c} x p(x,\theta_0)\,\mathrm{d}x \cdot \left(\int\limits_{c}^{+\infty} p(x,\theta_0)\,\mathrm{d}x\right)\left(\int\limits_{-\infty}^{c} p(x,\theta_0)\,\mathrm{d}x\right)^{n-2}$$

Проведя такие же преобразования, получим общую формулу для A_α:

$$\int\limits_{A_\alpha} T(X) p_{\boldsymbol{\xi}}(X, \theta_0)\,\mathrm{d}X = \int\limits_{-\infty}^{c} x p(x,\theta_0)\,\mathrm{d}x \left(\int\limits_{c}^{+\infty} p(x,\theta_0)\,\mathrm{d}x\right)^{|\alpha|}\left(\int\limits_{-\infty}^{c} p(x,\theta_0)\,\mathrm{d}x\right)^{n-1-|\alpha|}$$

Примем $T(X) = 0$ при $X \in A_{\{1,\ldots,n\}}$, обозначим интегралы для краткости

$$a = \int\limits_{-\infty}^{c} x p_\xi(x,\theta_0)\,\mathrm{d}x, \quad b = \int\limits_{-\infty}^{c} p_\xi(x,\theta_0)\,\mathrm{d}x$$

и получим

$$E_{\theta_0}\{\hat{a}\} = \sum_{i=0}^{n-1} a C_n^i b^{n-1-i}(1-b)^i = ab^{n-1}\sum_{i=0}^{n-1} C_n^i \left(\frac{1-b}{b}\right)^i =$$
$$= ab^{n-1}\left(\left(\frac{1-b}{b}+1\right)^n - \left(\frac{1-b}{b}\right)^n\right) = \frac{a}{b}(1-(1-b)^n).$$

Учитывая, что $0 < b < 1$,

$$\lim_{n\to+\infty} \frac{a}{b}(1-(1-b)^n) = \frac{a}{b} = \frac{\int_{-\infty}^{c} x p_\xi(x,\theta_0)\,\mathrm{d}x}{\int_{-\infty}^{c} p_\xi(x,\theta_0)\,\mathrm{d}x}$$

Для примера возьмем экспоненциальное распределение с параметром λ_0. Получим

$$\lim_{n\to+\infty} b(n,\lambda_0,c) = \lim_{n\to+\infty} \frac{\int_0^c x\lambda_0 e^{-\lambda_0 x}\,\mathrm{d}x}{\int_0^c \lambda_0 e^{-\lambda_0 x}\,\mathrm{d}x} - \frac{1}{\lambda_0} = -\frac{c}{1-e^{c\lambda_0}}$$

Эта функция стремится к 0 по модулю при $c \to +\infty$, но никогда не равна нулю, следовательно наша оценка будет смещенной для всех c и λ.

Теперь получим оценки по методу «игнорирования пропусков» в компьютерном эксперименте. Создадим выборку с гауссовским распределением с параметрами $a = 1$ и $\sigma^2 = 2$. Затем смоделируем 4 ситуации пропусков данных:

1. Без пропусков
2. Случайные пропуски, $m(x) = 0.5$
3. Полное цензурирование, $c = 1.5$
4. Частичное цензурирование, случай $m(x) = \begin{cases} 0.5, & x \geq 1.5 \\ 0, & x < 1.5 \end{cases}$.

Теперь удалим пропуски из выборки и будем вести дальнейшую работу с полученной выборкой размером $T' \leq T$. Будем рассчитывать оценку матожидания вида

$$\hat{a} = \frac{1}{T'} \sum_{i=1}^{T'} x_i.$$

На рисунках 1 и 2 показаны сами оценки и их вариация соответственно для различных размеров выборки T и для четырех ситуаций с пропусками данных.

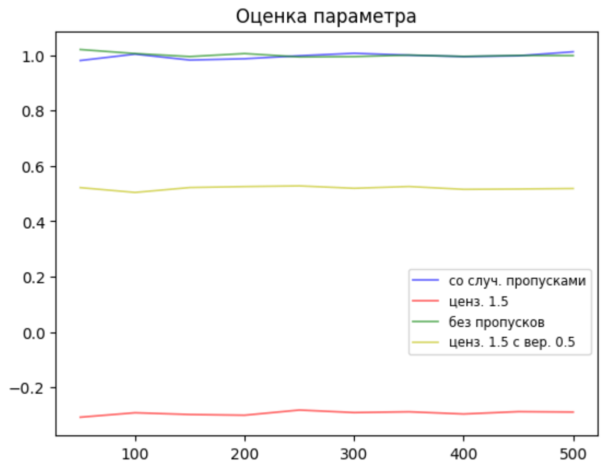

Рисунок 1

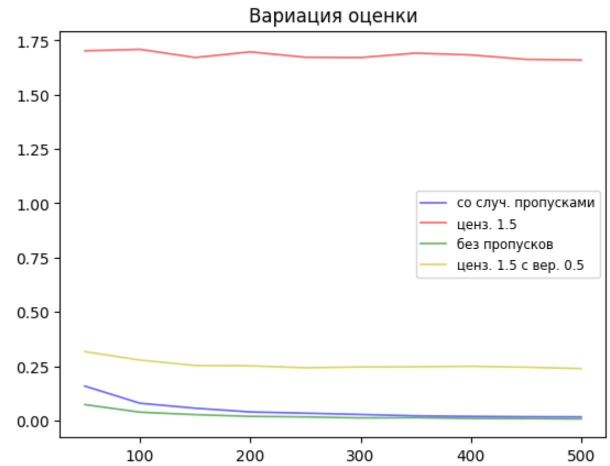

Рисунок 2

Метод максимального правдоподобия

Теперь попробуем применить метод максимального правдоподобия и посмотрим, насколько он усложняется в случае пропусков, зависящих от данных.

Пусть дана выборка размера n значений с пропусками случайной величины с плотностью вероятности $p_\xi(x,\theta)$, где θ — неизвестный параметр. И пусть модель пропусков описывается функцией $m(x)$, Obs — шаблон пропусков данной выборки, k — число наблюдаемых значений выборки. Тогда функция правдоподобия (ФП) будет выглядеть следующим образом:

$$L(\theta) = \prod_{i=1}^{n} (p_\xi(x,\theta))^{obs_i} G(p_\xi, m)^{1-obs_i},$$

где $G(p_\xi, m) = \int_{-\infty}^{+\infty} p_\xi(x,\theta) m(x)\,\mathrm{d}x$. Логарифмическая ФП (ЛФП) будет иметь вид

$$\begin{aligned} l(\theta) = ln(L(\theta)) &= \sum_{i=1}^{n} obs_i \ln(p_\xi(x_i,\theta)) + (1 - obs_i)\ln(G(p_\xi, m)) = \\ &= \sum_{i=1}^{n} obs_i \ln(p_\xi(x_i,\theta)) + (n-k)\ln(G(p_\xi, m)), \end{aligned}$$

где n — величина выборки, k — количество наблюдаемых значений.

Рассмотрим три случая пропусков

1. Случайные пропуски. Рассматриваем $m(x) \equiv c,\ 0 < c < 1$ и получаем функцию G и ЛФП:

$$G(p_\xi, m) = \int\limits_{-\infty}^{+\infty} c p_\xi(x,\theta)\,\mathrm{d}x = c,$$

$$l(\theta) = \sum_{i=1}^{n} obs_i \ln\big(p_\xi(x_i,\theta)\big) + (n-k)\ln c.$$

При случайных пропусках $G(\cdot)$ не зависит от θ и второе слагаемое никак не влияет на поиск максимума функции, поэтому в этом случае получение ММП-оценки никак не отличается от случая без пропусков.

2. Цензурирование. Рассмотрим функцию $m(x)$ вида

$$m(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}.$$

Тогда

$$G(p_\xi, m) = \int\limits_c^{+\infty} p_\xi(x, \theta)\,\mathrm{d}x,$$

$$l(\theta) = \sum_{i=1}^{n} obs_i \ln\big(p_\xi(x_i, \theta)\big) + (n-k)\ln\big(G(p_\xi, m)\big).$$

Теперь функция $G(\cdot)$ зависит от θ, что усложняет нахождение максимума функции.

3. Кусочно-постоянная функция пропусков. Для примера рассмотрим

$$m(x) = \begin{cases} 0, & x < a \\ \frac{1}{2}, & a \le x < b \\ 1, & x \ge b \end{cases}.$$

Тогда получаем ситуацию, аналогичную случаю 2:

$$G(p_\xi, m) = \frac{1}{2}\int\limits_a^{b} p_\xi(x, \theta)\,\mathrm{d}x + \int\limits_b^{+\infty} p_\xi(x, \theta)\,\mathrm{d}x,$$

$$l(\theta) = \sum_{i=1}^{n} obs_i \ln\big(p_\xi(x_i, \theta)\big) + (n-k)\ln\big(G(p_\xi, m)\big).$$

Теперь рассмотрим 2 конкретных примера:

1. Возьмем экспоненциальное распределение

$$p(x, \theta) = \left\{ \begin{array}{ll} \theta e^{-\theta x}, & x \ge 0 \\ 0, & x < 0 \end{array} \right., \quad \theta > 0$$

и модель пропусков

$$m(x) = \left\{ \begin{array}{l} 0, \ \ x < 0 \\ \frac{1}{2}, \ \ 0 \le x \le 3 \\ 0, \ \ x > 3 \end{array} \right..$$

Тогда

$$G(p, m) = \int\limits_0^{3} \frac{\theta}{2} e^{-\theta x}\,\mathrm{d}x = \frac{1}{2}(1 - e^{-3\theta})$$

и

$$L(\theta) = \theta^k e^{-\theta \sum_{obs_i = 1}} \frac{1}{2^{n-k}} (1 - e^{-3\theta})^{n-k}.$$

Прологарифмируем функцию правдоподобия и вычислим производную для поиска точки максимума

$$l(\theta) = \ln L(\theta) = k \ln \theta - \theta \sum_{obs_i=1} x_i + (n-k) \ln \left(\frac{1}{2}(1 - e^{-3\theta}) \right),$$

$$l'(\theta) = \frac{k}{\theta} - \sum_{obs_i=1} x_i + 3(n-k) \left(\frac{1}{1 - e^{-3\theta}} - 1 \right).$$

Перенесем все слагаемые с θ влево и получим уравнение для нахождения МП-оценки параметра

$$\frac{k}{\theta} + \frac{1}{1 - e^{-3\theta}} = \sum_{obs_i=1} x_i - 3(n-k).$$

Уравнение уже слишком сложное для аналитического решения, параметр можно оценить только численными методами, приближенно.

2. Возьмем нормальное распределение

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и ту же модель пропусков, что и в прошлом примере. Тогда

$$G(p, m) = \frac{1}{2\sqrt{2\pi}\sigma} \int\limits_0^3 e^{-\frac{(x-a)^2}{2\sigma^2}} \, \mathrm{d}x,$$

$$L(a, \sigma) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \right)^{obs_i} \left(\frac{1}{2\sqrt{2\pi}\sigma} \int\limits_0^3 e^{-\frac{(x-a)^2}{2\sigma^2}} \, \mathrm{d}x \right)^{1-obs_i} =$$
$$= \left(\frac{1}{3} \right)^{n-k} \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{n} e^{-\sum_{obs_i} \frac{(x_i-a)^2}{2\sigma^2}}$$

Функция правдоподобия уже стала настолько сложной, что даже взятие производной затруднительно. Поэтому здесь оценка параметров является препятствием даже для численных методов.

Посмотрим, какими методами можно найти значения параметров проще.

Метод моментов

Первым из методов, которые мы рассмотрим, будет метод моментов. Он основан на приравнивании теоретических характеристик распределения к выборочным. В случае пропусков, зависящих от данных, удобнее всего использовать следующие два равенства

$$P\{obs_i = 1\} = \int\limits_{-\infty}^{+\infty} p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x = \frac{k}{n},$$

$$E\{x_i\} = \frac{\int_{-\infty}^{+\infty} x p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x}{\int_{-\infty}^{+\infty} p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x} = \frac{\sum_{obs_i=1} x_i}{k},$$

где k — количество наблюдаемых значений в выборке.

Сначала посмотрим, как изменится вид уравнений в различных случаях с пропусками.

1. Случайные пропуски. В этом случае $m(x) \equiv c,\ 0 < c < 1$ и уравнения примут следующий вид

$$\frac{k}{n} = \int\limits_{-\infty}^{+\infty} p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x = (1-c)\int\limits_{-\infty}^{+\infty} p_\xi(x,\theta)\,\mathrm{d}x = 1-c$$

$$\frac{\sum_{obs_i=1} x_i}{n} = \frac{\int_{-\infty}^{+\infty} x p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x}{\int_{-\infty}^{+\infty} p_\xi(x,\theta)(1-m(x))\,\mathrm{d}x} = E\{\xi\}$$

Из первого уравнения пропала θ, поэтому его нельзя использовать. Вместо него можно составить уравнение для момента порядка 2 или выше. Второе же уравнение в итоге ничем не отличается от соответствующего уравнения в случае без пропусков.

2. Цензурирование. Рассматриваем функцию вероятности пропуска

$$m(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}.$$

В таком случае уравнения примут вид

$$\int\limits_{-\infty}^{c} p_\xi(x,\theta)\,\mathrm{d}x = \frac{k}{n},$$

$$\frac{\int_{-\infty}^{c} x p_\xi(x,\theta)\,\mathrm{d}x}{\int_{-\infty}^{c} p_\xi(x,\theta)\,\mathrm{d}x} = \frac{\sum_{obs_i=1} x_i}{n}$$

3. Кусочно-постоянная функция пропусков. К примеру, возьмем функцию

$$m(x) = \begin{cases} 0, & x < a \\ \frac{1}{2}, & a \le x < b \\ 1, & x \ge b \end{cases}.$$

Подставив в уравнения баланса, получим

$$\int\limits_{-\infty}^{a} p_\xi(x,\theta)\,\mathrm{d}x + \frac{1}{2}\int\limits_{a}^{b} p_\xi(x,\theta)\,\mathrm{d}x = \frac{k}{n}$$

$$\frac{\int_{-\infty}^{a} x p_\xi(x,\theta)\,\mathrm{d}x + \frac{1}{2}\int_a^b x p_\xi(x,\theta)\,\mathrm{d}x}{\int_{-\infty}^{c} p_\xi(x,\theta)\,\mathrm{d}x + \frac{1}{2}\int_a^b p_\xi(x,\theta)\,\mathrm{d}x} = \frac{\sum_{obs_i=1} x_i}{n}$$

Для решения сначала применим метод к распределению с одним параметром, например к экспоненциальному

$$p(x,\theta) = \begin{cases} \theta e^{-\theta x}, & x \ge 0 \\ 0, & x < 0 \end{cases}, \quad \theta > 0.$$

В случаях 2 и 3 уравнения усложняются, в особенности тем, что во втором уравнении слева имеем частное двух интегралов, зависящих от θ.

Теперь проверим метод моментов на практике. Возьмем параметр $\theta = 1$ и функцию модели пропусков

$$m(x) = \begin{cases} 1, & x < 1 \\ \frac{1}{2}, & 1 \le x \le 2 \\ 0, & x > 2 \end{cases}$$

и проведем компьютерный эксперимент, аналогичный тому, который проводился для метода «игнорирования пропусков». Замечу, что здесь достаточно только одного из двух уравнений метода моментов. В таком случае можно решать его методом дихотомии, предполагая, что функция зависимости интеграла $\int_{-\infty}^{+\infty} p_\xi(x,\theta) m(x)\,\mathrm{d}x$ от параметра θ монотонна.

На рисунках 3 и 4 показаны оценки и их вариация для разных объемов выборки.

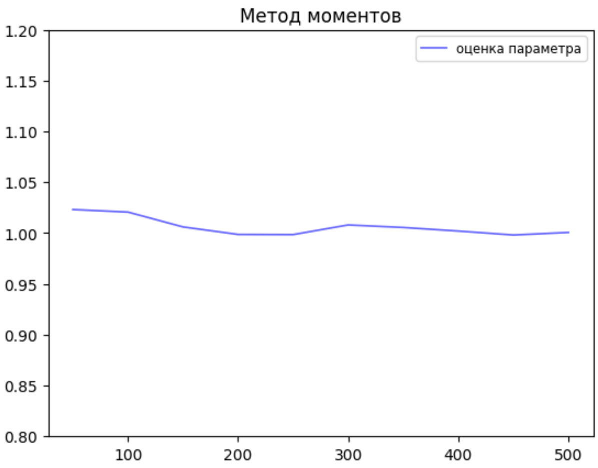

Рисунок 3

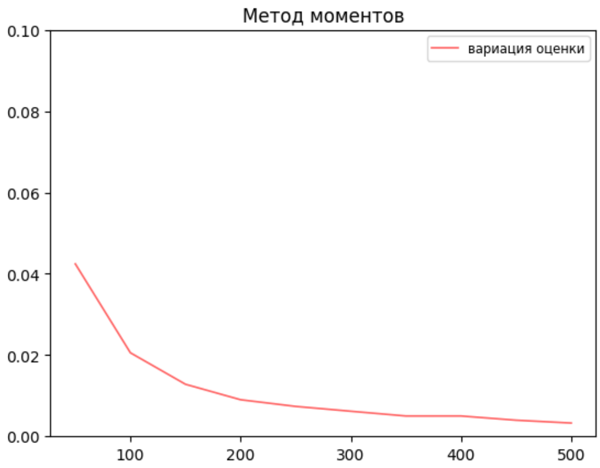

Рисунок 4

Как видим, метод достаточно точен и даже при небольшом объеме выборки средняя вариация не превысила значения 0.05.

Теперь попробуем таким же образом применить этот метод к нормальному распределению с параметрами $a = 0.7$ и $\sigma = 0.8$. Так как параметра два, придется решать систему из двух уравнений, приведенных в начале главы. Это гораздо сложнее, чем в случае с одним параметром, поэтому для нахождения решения использовались инструменты языка Python (в котором реализован метод сопряженных направлений).

На рисунках 5 и 6 показаны оценка и вариация соответственно для параметров распределения. Видим, что для нормального распределения метод тоже работает.

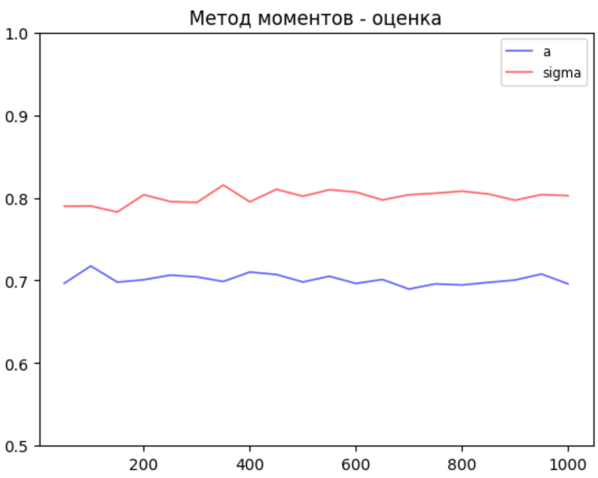

Рисунок 5

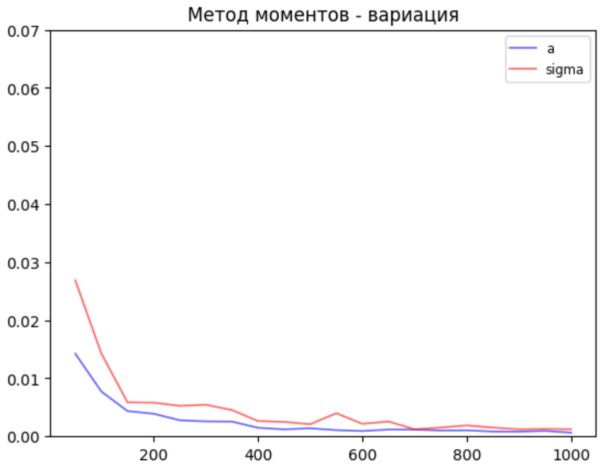

Рисунок 6

Заключение

Таким образом, в курсовой работе получены следующие основные результаты:

1. Произведено сравнение поведения оценок параметров распределения вероятности, полученных методом игнорирования пропусков, для различных видов пропусков во временном ряду. Показана несостоятельность метода в случае неслучайных пропусков.
2. На примере метода максимального правдоподобия исследовано влияние, оказываемое на стандартные методы получения оценок параметров смешением плотности распределения вероятности с функцией модели пропусков, зависящих от данных.
3. Исследованы методы, позволяющие получить верные оценки параметров распределения для случая неслучайных пропусков.