

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»
Навчально-науковий фізико-технічний інститут
Кафедра математичних методів захисту інформації

«На правах рукопису»

УДК 519.21

«До захисту допущено»

В.о. завідувача кафедри

_____ Сергій ЯКОВЛЄВ

«__» _____ 2023 р.

Дипломна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою
«Математичні методи криптографічного захисту інформації»

зі спеціальності: 113 Прикладна математика
на тему: «Оцінювання параметрів стохастичного різницевого
рівняння з марковськими перемиканнями та мультиплікативним
шумом»

Виконав:
студент IV курсу, групи ФІ-93
Шифрін Денис Семенович

Керівник:
доцент, к. ф.-м. н.
Ніщенко Ірина Іванівна

Рецензент:
кандидат фіз.-мат. наук,
наук. співр. в. теор. випадк.
проц. Інст. мат. НАН Укр.
Глиняна Катерина Валеріївна

Засвідчую, що у цій дипломній
роботі немає запозичень з праць
інших авторів без відповідних
посилань.

Студент _____

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»

Навчально-науковий фізико-технічний інститут
Кафедра математичних методів захисту інформації

Рівень вищої освіти — перший (бакалаврський)
Спеціальність — 113 Прикладна математика,
ОПП «Математичні методи криптографічного захисту інформації»

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри

_____ Сергій ЯКОВЛЄВ

«__» _____ 2023 р.

ЗАВДАННЯ
на дипломну роботу

Студент: Шифрін Денис Семенович

1. Тема роботи: *«Оцінювання параметрів стохастичного різницевого рівняння з марковськими перемиканнями та мультиплікативним шумом»*, науковий керівник дипломної роботи: доцент, к. ф.-м. н. Ніщенко Ірина Іванівна,

затверджені наказом по університету №__ від «__» _____ 2023 р.

2. Термін подання студентом роботи: «__» _____ 2023 р.

3. Об'єкт дослідження: різницеві рівняння з марковськими перемиканнями та мультиплікативним шумом.

4. Предмет дослідження: пошук ефективних методів оцінювання параметрів рівнянь авторегресії із різноманітними шумами.

5. Перелік завдань:

1) розглянути стохастичні різницеві рівняння із мультиплікативним шумом та марковськими перемиканнями;

2) оцінити параметри рівняння;

3) знайти ефективні методи прогнозу.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу:
Презентація доповіді

7. Орієнтовний перелік публікацій: планується доповідь ³ на всеукраїнській конференції
8. Дата видачі завдання: 10 вересня 2022 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання	Примітка
1	Узгодження теми роботи із науковим керівником	01-15 вересня 2022 р.	Виконано
2	Огляд опублікованих джерел за тематикою дослідження	Вересень- жовтень 2022 р.	Виконано
3	Оцінка параметрів моделі	листопад- грудень 2022 р.	Виконано
4	Імплементация необхідних алгоритмів	Січень-квітень 2023 р.	Виконано
5	Аналіз результатів роботи алгоритмів	Травень 2023 р.	Виконано
6	Оформлення та захист дипломної роботи	Червень 2023 р.	Виконано

Студент _____ Денис ШИФРІН

Керівник _____ Ірина НІЩЕНКО

РЕФЕРАТ

Кваліфікаційна робота містить: 50 стор., 8 рисунків, 3 таблиці, 10 джерел.

В роботі розглянуто рівняння авторегресії з мультиплікативним шумом, коефіцієнти якого змінюються згідно з еволюцією деякого скінченного ланцюга Маркова. В припущенні, що стани цього ланцюга є неспостережуваними, а перехідні ймовірності — невідомими, за скінченим набором спостережень часового ряду оцінено невідомі параметри розподілу, яким моделюється мультиплікативний шум, та знайдено характеристики прихованого ланцюга Маркова.

ПРИХОВАНА МАРКОВСЬКА МОДЕЛЬ, АВТОРЕГРЕСІЙНА
МОДЕЛЬ, МУЛЬТИПЛІКАТИВНИЙ ШУМ, ЧАСОВІ РЯДИ

ABSTRACT

In this paper, we consider an autoregressive equation with multiplicative noise, the coefficients of which change according to the evolution of a finite Markov chain. Under the assumption that the states of this chain are unobservable and the transition probabilities — are unknown, the unknown distribution parameters modeling the multiplicative noise are estimated from a finite set of time series observations and the characteristics of the hidden Markov chain are found.

HIDDEN MARKOV MODEL, AUTOREGRESSIVE MODEL,
MULTIPLICATIVE NOISE, TIME SERIES

ЗМІСТ

Вступ.....	8
1 Опис моделі.....	10
1.1 Опис досліджуваної моделі	10
1.2 Основні відомості про ланцюги Маркова	11
1.3 Означення прихованої марковської моделі.....	13
1.4 Задача оцінювання.....	15
1.5 Задача декодування	16
1.6 Задача навчання	17
1.7 Метод множників Лагранжа	19
1.8 Постановка задачі.....	20
1.9 Висновки до першого розділу	22
2 Побудова оцінок невідомих параметрів.....	23
2.1 Розв'язання задачі оцінювання.....	23
2.2 Розв'язання задачі декодування.....	27
2.3 Розв'язання задачі навчання	27
2.3.1 Знаходження функції правдоподібності.....	29
2.3.2 Знаходження функції квазі-лог правдоподібності	30
2.3.3 Оцінювання параметрів моделі	35
2.4 Висновки до другого розділу	39
3 Аналіз результатів чисельного експерименту	40
3.1 Моделювання поведінки ланцюжка спостережень	40
3.2 Аналіз розв'язку задач оцінювання.....	41
3.3 Аналіз розв'язку задачі декодування	41
3.4 Аналіз розв'язку задачі навчання.....	42
3.5 Побудова прогнозу	45
3.6 Висновки до третього розділу	47
Висновки	48
Перелік посилань	49

ВСТУП

Актуальність дослідження. Зростаюче використання цифрових технологій у всіх сферах життєдіяльності суспільства призводить до експоненційного збільшення обсягу накопичених даних. Велика частина цих даних може бути представлена часовими рядами, тобто впорядкованими за часом спостереженнями. Аналіз таких рядів з метою передбачення майбутніх значень або розуміння процесу їх генерації вимагає належної математичної моделі. Одним з ефективних підходів до аналізу часових рядів є використання авторегресійних моделей. Ці моделі дозволяють визначити, яку частину поточних спостережень можна пояснити попередніми значеннями та застосовні в багатьох сферах. Наприклад, необхідність прогнозування волатильності криптовалют, яка виникає у задачах аналізу фінансових ринків [7], може бути одним зі стимулюючих факторів. Стохастичні апроксимації [5], аналіз потоків даних у вузлах високонавантажених систем та оптимальне стохастичне керування [2] можуть сприяти розвитку робототехніки. Моделі авторегресії можуть також бути корисними при вивченні механізмів поширення захворювання під час епідеміологічних спалахів. Авторегресійні рівняння з мультиплікативним шумом виникають в багатьох застосуваннях теорії сигналів: в каналах із завмиранням, радарах, лідарах, ехолотах [1].

Метою дослідження є пошук ефективних методів прогнозу розвитку авторегресійних моделей із різноманітними шумами. **Задачі дослідження**

- 1) розглянути авторегресійні моделі із мультиплікативним шумом та марковськими перемиканнями;
- 2) побудувати оцінки параметрів авторегресійних моделей із марковськими перемиканнями та мультиплікативним шумом.

Об'єктом дослідження є різницеві рівняння з марковськими

перемиканнями та мультиплікативним шумом.

Предметом дослідження є оцінювання параметрів рівнянь авторегресії із різноманітними шумами.

При розв'язанні поставлених завдань використовувались такі *методи дослідження*: методи теорії імовірностей, математичної статистики, комп'ютерного та статистичного моделювання.

Наукова новизна отриманих результатів полягає в знаходженні ефективних методів аналізу та прогнозу еволюції авторегресійної моделі із мультиплікативним шумом за допомогою теорії прихованих марківських моделей.

Практичне значення результатів полягає в можливості оцінювати невідомі параметри авторегресійних моделей за заданим ланцюжком спостережень та застосовувати такі моделі в прогнозуванні.

Апробація результатів та публікації. Частину результатів даної роботи було представлено на XXI Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених "Теоретичні та прикладні проблеми фізики, математики та інформатики" (11-12 травня 2023 р., м. Київ, Україна).

1 ОПИС МОДЕЛІ

В цьому розділі буде розглянуте стохастичне різницеве рівняння з марковськими перемиканнями та мультиплікативним шумом. Буде дано означення прихованої марковської моделі. Будуть описані задачі теорії прихованих марковських моделей та ефективні методи їх вирішення. Наприкінці розділу буде окреслено питання, які будуть розглядатись в наступних розділах.

1.1 Опис досліджуваної моделі

Розглянемо рівняння авторегресії першого порядку наступного вигляду

$$Y_{n+1} = (C_{X_{n+1}} + \eta_{n+1})Y_n + \epsilon_{n+1}, n \geq 0 \quad (1.1)$$

У цьому рівнянні послідовність $\{X_n\}_{n \geq 1}$ являє собою ланцюг Маркова, стани якого задані на множині $E = \{1, 2\}$, з початковим розподілом $\pi = (P(X_1 = 1), P(X_2 = 2))$ та матрицею перехідних ймовірностей

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Послідовність $\{\epsilon_n\}_{n \geq 1}$ є послідовністю незалежних випадкових величин зі стандартним нормальним розподілом кожна. Мультиплікативний шум моделюємо послідовністю $\{\eta_n\}_{n \geq 1}$ незалежних випадкових величин зі скінченним дискретним розподілом:

$$\forall n \geq 1 : \eta_n = \begin{cases} a_1, & \text{з імовірністю } p_1 \\ a_2, & \text{з імовірністю } p_2, \sum_{i=1}^3 p_i = 1. \\ a_3, & \text{з імовірністю } p_3 \end{cases} \quad (1.2)$$

Вважаємо, що константи C_1, C_2 та a_1, a_2, a_3 є такими, що виконується умова

$$|C_i + a_j| \leq 1, \forall i = \overline{1,2}, j = \overline{1,3}. \quad (1.3)$$

Задача дослідження полягає в тому, щоб за скінченним набором спостережень $y_1, y_2, \dots, y_T$ оцінити параметри π, A прихованого ланцюга $\{X_n\}_{n \geq 1}$ та параметри $p = (p_1, p_2, p_3), a = (a_1, a_2, a_3)$ мультиплікативного шуму $\{\eta_n\}_{n \geq 1}$ рівняння (1.1).

Наведемо деякі відомості про ланцюги Маркова та приховану марковську модель.

1.2 Основні відомості про ланцюги Маркова

Означимо скінченну множину $S = \{s_1, s_2, \dots, s_N\}$ та розглянемо послідовність випадкових величин $\{X_n\}_{n \geq 1}$, заданих на одному й тому ж ймовірнісному просторі, які приймають значення із множини S . В такому випадку елементи множини S називатимуться станами, а сама множина — множиною станів послідовності $\{X_n\}_{n \geq 1}$. Тоді введемо поняття ланцюга Маркова.

Означення 1.1. [4] Послідовність випадкових величин $\{X_n\}_{n \geq 1}$ називають *ланцюгом Маркова*, якщо для довільного натурального числа $t \geq 1$ має місце наступна рівність

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_t = x_t \mid X_{t-1} = x_{t-1}), \quad (1.4)$$

для довільних $x_0, x_1, \dots, x_t \in S$.

Властивість 1.4 відображає відсутність пам'яті в марковському процесі, за умови, що відомий поточний стан.

Якщо $X_t = j$, ми кажемо, що ланцюг знаходиться у стані j в момент часу t . Припускаємо, що умовна ймовірність

$$A_{ij} = P(X_t = j \mid X_{t-1} = i), \quad \forall t \geq 1, \quad i, j \in S, \quad (1.5)$$

не залежить від t . Така ймовірність називається перехідною ймовірністю зі стану i в стан j . Сукупність таких перехідних ймовірностей формують квадратну матрицю

$$A = (A_{ij})_{i,j=\overline{1,N}}, \quad (1.6)$$

яка називається матрицею перехідних ймовірностей. Слід зазначити, що i -тий рядок такої матриці задає умовний розподіл ймовірностей X_t при заданому $X_{t-1} = i$. Тому наступна властивість набуває істинності

$$A_{ij} \geq 0, \quad \sum_{j=1}^N A_{ij} = 1. \quad (1.7)$$

Матриця із вказаною властивістю називається стохастичною.

Спираючись на властивість 1.4, можна записати скінченновимірні розподіли ланцюга.

Твердження 1.1. [4] *Якщо послідовність $\{X_n\}_{n \geq 1}$ є ланцюгом Маркова, переходи між станами якого описуються матрицею перехідних ймовірностей A , то*

$$P(X_1 = x_1, \dots, X_T = x_T) = P(X_1 = x_1) \prod_{t=1}^{T-1} A_{x_t x_{t+1}} \quad (1.8)$$

1.3 Означення прихованої марковської моделі

Прихована марковська модель, відома також як ймовірнісна функція від ланцюга Маркова, являє собою згенерований двома взаємопов'язаними механізмами стохастичний процес. В основі лежить марковський ланцюг зі скінченною кількістю станів та набір випадкових функцій, кожна з яких асоціюється з відповідним станом. У дискретні моменти часу процес вважається таким, що перебуває в певному стані, і спостереження генерується випадковою функцією, що відповідає поточному стану. Після цього ланцюг Маркова змінює свій стан відповідно до матриці перехідних ймовірностей.

Спостерігач отримує тільки значення функцій, пов'язаних з кожним станом, і не може безпосередньо спостерігати стани самого ланцюга Маркова. Спостережуваними можуть бути багатовимірні випадкові величини з певною неперервною сумісною функцією щільності розподілу. Оскільки ланцюг Маркова є неспостережуваним, то й назва такої структури — прихований ланцюг Маркова.

Означення 1.2. [4] Послідовність $\{(X_n, Y_n)\}_{n \geq 0}$ називають *прихованою марковською моделлю*, якщо

1. послідовність $\{X_n\}_{n \geq 1}$ — ланцюг Маркова з множиною станів $S = \{s_1, s_2, \dots, s_N\}$, початковим розподілом π та матрицею перехідних ймовірностей A .
2. послідовність $\{(X_n, Y_n)\}_{n \geq 1}$ утворює ланцюг Маркова на $S \times O$, де $O = \{y_0, y_1, \dots, y_M\}$ — множина станів послідовності $\{Y_n\}_{n \geq 1}$.
3. Випадкові величини $Y_1, \dots, Y_T$ є умовно незалежними при заданих $X_1, \dots, X_T$, що виражається наступним чином

$$P(Y_1 = y_1, \dots, Y_T = y_T \mid X_1 = x_1, \dots, X_T = x_T) = \prod_{i=1}^T P(y_i \mid x_1, \dots, x_T),$$

і крім того розподіл Y_i залежить лише від X_i :

$$P(Y_i = y_i \mid X_1 = x_1, \dots, X_T = x_T) = P(Y_i = y_i \mid X_i = x_i).$$

Величини $b_j(i) = P(Y_t = i \mid X_t = j), i = \overline{1, M}, j = \overline{1, M}$ формують матрицю B умовних ймовірностей спостережених станів при заданому прихованому стані.

З означення випливає, що скінченновимірні розподіли послідовності мають $\{(X_n, Y_n)\}_{n \geq 1}$ мають вигляд

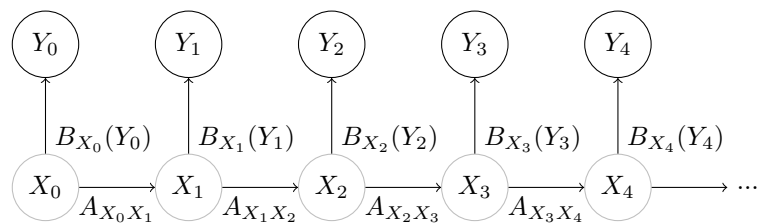
$$P(Y_1 = y_1, \dots, Y_T = y_T, X_1 = x_1, \dots, X_T = x_T) = \pi_{x_1} \prod_{t=1}^{T-1} a_{x_t x_{t+1}} \prod_{t=1}^T b_{x_t}(y_t), \quad (1.9)$$

де π — це початковий розподіл ланцюга Маркова $\{X_n\}_{n \geq 1}$, а матриця A — його матриця перехідних ймовірностей.

Приховану марковську модель, побудовану за π, A, B будемо позначати

$$M = (\pi, A, B).$$

Прихована марковська модель із зазначеними параметрами може бути зображена графом, вершинами якого виступатимуть стани прихованого ланцюга Маркова та спостереження в деякі моменти часу, а ребрами — взаємозв'язки між ними. Граф, який описує модель $M = (\pi, A, B)$ має наступний вигляд



Оскільки всі необхідні поняття вже введені, перейдемо до

розв'язання задач теорії прихованих марковських моделей. Виділяють три задачі: оцінювання, декодування та навчання. Розв'язки цих задач є важливими при дослідженні послідовності змінюваних із плином часу даних, тож далі наведено детальний опис кожної з них.

1.4 Задача оцінювання

Задача оцінювання полягає у визначенні ймовірності

$$\begin{aligned} P_M(Y_1 = y_1, \dots, Y_T = y_T) = \\ = \sum_{(x_1, \dots, x_T)} P_M(Y_1 = y_1, \dots, Y_T = y_T, X_1 = x_1, \dots, X_T = x_T) \end{aligned}$$

отримати послідовність спостережень $y = (y_1, y_2, \dots, y_T)$ за заданою моделлю $M = (\pi, A, B)$. Проблему становить ефективність обчислень за формулою 1.9. Оскільки підсумовування міститиме N^{T+1} можливих послідовностей станів, то повна складність буде $2(T+1)N^{T+1}$ операцій. Ефективним рішенням цієї проблеми є використання алгоритму прямого-зворотнього ходу.

Введемо так звані змінні прямого ходу

$$\alpha_t(j) = P(Y_0 = y_0, \dots, Y_t = y_t, X_t = j), \quad (1.10)$$

які будуть обчислені як ймовірність одночасного отримання ланцюжка спостережень до часу $t \leq T$ та прихованого ланцюга Маркова у стані j у момент часу t , та коефіцієнти алгоритму зворотнього ходу

$$\beta_t(j) = P(Y_{t+1} = y_{t+1}, \dots, Y_T = y_T \mid X_t = j), \quad (1.11)$$

які рівні умовній ймовірності отримати ланцюжок спостережень починаючи з часу $t+1$ і до T при умові, що прихований ланцюг Маркова знаходиться у стані j у момент часу t . В термінах цих змінних функція

правдоподібності обчислюється так

$$P_M(Y = y) = \sum_{j=1}^N \alpha_T(j) = \sum_{j=1}^N \pi_j B_{j1} \beta_1(j). \quad (1.12)$$

Слід зазначити, що як коефіцієнти прямого ходу, так і коефіцієнти зворотнього ходу обчислюються рекурентно. Початкові значення задаються як

$$\begin{aligned} \alpha_1(j) &= \pi_j b_j(1), & j = \overline{1, 2} \\ \beta_T(j) &= 1, & j = \overline{1, 2} \end{aligned}$$

а всі наступні можна отримати використовуючи відповідні рекурентні співвідношення

$$\begin{aligned} \alpha_{t+1}(j) &= \sum_{i=1}^N \alpha_t(i) a_{ij} b_j(t+1), & t = \overline{1, T-1}, j = \overline{1, 2} \\ \beta_{t-1}(j) &= \sum_{i=1}^N a_{ji} b_j(t) \beta_t(i), & t = \overline{T, 2}, j = \overline{1, 2} \end{aligned}$$

1.5 Задача декодування

Задача декодування полягає у пошуку найбільш ймовірної послідовності станів $x = (x_1, x_2, \dots, x_T)$ прихованого ланцюга Маркова $\{X_n\}_{n \geq 1}$, що впливають на спостережувану послідовність $y = (y_1, \dots, y_T)$ для заданої моделі $M = (\pi, A, B)$. Розв'язок цієї задачі дозволяє знайти алгоритм Вітербі, максимізуючи появу наступного ланцюжка станів

$$(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_T) = \arg \max_{x=(x_1, x_2, \dots, x_T)} P_M(X = x \mid Y = y).$$

Нижче описано алгоритм Вітербі [6].

1. Ініціалізація. t — часовий індекс, для кожного стану $j \in S$ означимо

змінні $\psi_t(j)$ та $\delta_t(j)$.

2. Початок. $t = 0$. Для кожного стану $j \in S$ обчислюємо:

$$\delta_0(j) = \pi_j \cdot b_j(y_0),$$

$$\psi_0(j) = \emptyset$$

3. Рекурсія. При $t = 1$ обчислюємо

$$\delta_t(j) = \left[\max_{i=\overline{1,N}} \delta_{t-1}(i) \cdot a_{ij} \right] \cdot b_j(y_t).$$

4. Збереження. Обчислюємо

$$\psi_t = \arg \max_{i=\overline{1,N}} \delta_{t-1}(i) \cdot a_{ij}$$

для всіх $j = \overline{1,N}$. Встановлюємо $t = t + 1$ та повторюємо з кроку 3 доки не виконується рівність $t = T$.

5. Припинення.

$$P^* = \max_{i=\overline{1,N}} \delta_T(i)$$

$$j_*(T) = \arg \max_{i=\overline{1,N}} \delta_T(i)$$

6. Відстежування. Найкраще рішення буде знайдене як

$$j_*(t) = \psi_{t+1}(j_*(t+1)), \quad t = \overline{T-1, 0}.$$

1.6 Задача навчання

Задачею навчання виступає оцінка параметрів моделі M за заданим ланцюгом спостережених даних $y = (y_1, y_2, \dots, y_T)$, якою він міг бути згенерований. Одним із розв'язків даної задачі може виступати максимізація функції правдоподібності за допомогою алгоритму Баума-Велша, який є аналогом ЕМ алгоритму для скінченних сумішей.

Шукана оцінка при такій максимізації буде

$$\hat{M}_{ML} = \arg \max_M P_M(Y = y).$$

Позначимо індексом s вектор значень $x = (x_1, \dots, x_T)$, що визначається як послідовність станів прихованого ланцюга Маркова. Тоді введемо такі змінні при всіх можливих наборах значень s :

$$u_s = P_M(Y = y, X = x), \quad v_s = P_{M^*}(Y = y, X = x),$$

де u_s буде рівна функції правдоподібності для моделі M , а v_s — для моделі M^* при фіксованих x .

Замість максимізації функції правдоподібності будемо максимізувати допоміжну функцію — функцію квазі-логарифмічної правдоподібності, яка визначається як

$$Q(M, M^*) = \sum_s u_s \ln v_s. \quad (1.13)$$

Алгоритм подальших дій буде наступним:

- 1) продиференціювати функцію $Q(M, M^*)$ по кожному із параметрів моделі M^* та, прирівнявши похідні до нуля, знайти оцінки параметрів.
- 2) виразити всі невідомі величини через параметри поточної моделі та коефіцієнти алгоритмів прямого та оберненого ходу.
- 3) ітеративно переоцінювати параметри моделі за принципом ЕМ алгоритму доти, доки не буде виконано критерій зупинки.

Оскільки елементи матриці A та початкового вектора π пов'язані лінійною залежністю, то для знаходження оцінок буде необхідно застосувати метод множників Лагранжа.

1.7 Метод множників Лагранжа

Перехід до методу множників Лагранжа виконується задля того, щоб перейти до функції з більшим числом змінних, але разом з тим без додаткових умов, що значно полегшує оцінювання невідомих параметрів.

Означення 1.3. [4] *Функція Лагранжа* для оцінки параметрів $l_1^*, l_2^*, \dots, l_n^*$ виглядає як

$$\mathcal{L}(l_1^*, l_2^*, \dots, l_n^*, \theta) = Q(M, M^*) - \theta \Phi(l_1^*, \dots, l_n^*),$$

де функція $\Phi(l_1^*, \dots, l_n^*)$ відображає як залежать параметри $l_1^*, \dots, l_n^*$ один від одного.

Для оцінки необхідного параметра знаходиться $\nabla \mathcal{L}(l_1^*, l_2^*, \dots, l_n^*, \theta)$ та прирівнюється до нуля, звідки отримується система

$$\left\{ \begin{array}{l} \frac{\partial \mathcal{L}}{\partial l_1} = 0 \\ \frac{\partial \mathcal{L}}{\partial l_1} = 0 \\ \dots \\ \frac{\partial \mathcal{L}}{\partial l_n^*} = 0 \\ \frac{\partial \mathcal{L}}{\partial \theta} = 0 \end{array} \right.$$

Зауваження. При переоцінці параметрів слід пам'ятати, що матриця перехідних ймовірностей A та вектор початкового розподілу стохастичні, що і буде додатковим обмеженням при максимізації функції $Q(M, M^*)$.

1.8 Постановка задачі

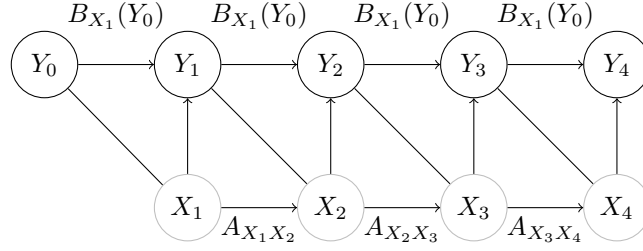
Дослідження авторегресійних моделей є вкрай важливим для багатьох галузей науки. Так, наприклад, одна з таких моделей була застосована для глибокого навчання роботів для подальшого розпізнавання ними об'єктів реального світу [10]. Схоже застосування різницевих рівнянь можна знайти в дослідженні [3], в якому розглядалось двостороннє керування системою, яка складалася із група роботів. Керуючий робот обробляв надані людиною команди та надавав чіткі інструкції підпорядкованим йому машинам. Аналіз ринку криптовалют вимагає побудови прогнозу часових рядів [9]. Таким чином, авторегресійні моделі застосовні у аналізі та передбаченні динаміки цієї валюти із плином часу.

Розглянуто авторегресійну модель з марковськими перемиканнями

$$Y_{n+1} = (C_{X_{n+1}} + \eta_{n+1})Y_n + \epsilon_{n+1}, n \geq 0$$

відрізняють від звичайної прихованої марківської моделі умови, які пов'язують спостереження і приховані стани. У звичайній прихованій марковській моделі припускається, що спостереження є умовно незалежними при заданих станах ланцюга Маркова, натомість в авторегресійній моделі з марковськими перемиканнями спостереження в деякий момент часу залежать як від поточного стану прихованого ланцюга, так і від спостережень в попередні моменти часу. Граф, який описує взаємозв'язки між станами прихованого ланцюга Маркова та спостереженнями для досліджуваної авторегресійної моделі (1.1), зображається наступним чином

Зауважимо, що умовний розподіл випадкової величини Y_n при заданих значеннях $X_n = i, \eta_n = a_k$ та відомому значенні $Y_{n-1} = y_{n-1}$ має нормальний розподіл $\mathcal{N}((C_i + a_k)y_{n-1}, 1)$ з математичним сподіванням $(C_i + a_k)y_{n-1}$ та одиничною дисперсією. Позначимо через $b_{ik}(y \mid y_{n-1})$



щільність цього розподілу

$$\begin{aligned}
 b_{ik}(y \mid y_{n-1}) &= P_{Y_n | X_n, \eta_n, Y_{n-1}}(y \mid i, a_k, y_{n-1}) = \\
 &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y - (C_i + a_k)y_{n-1})^2}
 \end{aligned} \tag{1.14}$$

Тоді умовний розподіл Y_n при заданому значенні $X_n = i$ та відомому значенні $Y_{n-1} = y_{n-1}$ є сумішшю нормальних розподілів із щільностями $b_{i1}(y \mid y_{n-1})$, $b_{i2}(y \mid y_{n-1})$, $b_{i3}(y \mid y_{n-1})$ з коефіцієнтами змішування p_1, p_2, p_3 :

$$\begin{aligned}
 b_i(y \mid y_{n-1}) &\equiv P_{Y_n | X_n, Y_{n-1}}(y \mid i, y_{n-1}) = \\
 &= p_1 b_{i1}(y \mid y_{n-1}) + p_2 b_{i2}(y \mid y_{n-1}) + p_3 b_{i3}(y \mid y_{n-1})
 \end{aligned} \tag{1.15}$$

Набір $(\pi, A, C_1, C_2, p = (p_1, p_2, p_3), a = (a_1, a_2, a_3))$ визначає авторегресійну модель з марковськими перемиканнями, в якій π та A є початковим розподілом та матрицею перехідних ймовірностей прихованого ланцюга Маркова $\{X_n\}_{n \geq 1}$; C_1, C_2 є деякими відомими константами. Вектори $p = (p_1, p_2, p_3)$ та $a = (a_1, a_2, a_3)$ є параметрами мультиплікативного шуму $\{\eta_n\}_{n \geq 1}$. Цей набір визначає умовні розподіли спостережень $\{Y_n\}_{n \geq 1}$ за формулами (1.14), (1.15).

Метою роботи виступає оцінка параметрів π, A прихованого ланцюга $\{X_n\}_{n \geq 1}$ та параметрів $p = (p_1, p_2, p_3), a = (a_1, a_2, a_3)$ мультиплікативного шуму $\{\eta_n\}_{n \geq 1}$ рівняння (1.1) за скінченним набором спостережень $y_1, y_2, \dots, y_T$.

1.9 Висновки до першого розділу

В розділі було наведено означення ланцюга Маркова та деякі ключові поняття теорії прихованих марковських моделей. Було означено три задачі прихованих марковських моделей та алгоритми для їх розв'язання, які будуть детальніше розглянуті в наступному розділі. Було означено досліджувану авторегресійну модель з описом всіх її параметрів, та параметрів мультиплікативного шуму.

2 ПОБУДОВА ОЦІНОК НЕВІДОМИХ ПАРАМЕТРІВ

Цей розділ міститиме основні розрахунки для розв'язання трьох задач теорії прихованих марковських моделей для заданої авторегресійної моделі.

2.1 Розв'язання задачі оцінювання

Нехай маємо скінчений набір спостережень $y = (y_0, y_1, \dots, y_T)$, за яким прагнемо оцінити початковий розподіл, перехідні ймовірності прихованого ланцюга Маркова $\{X_n\}_{n \geq 1}$ та параметри $p = (p_1, p_2, \dots, p_K)$, $a = (a_1, a_2, \dots, a_K)$ при $K = 3$ мультиплікативного шуму. Надалі використовуватимемо позначення $x = (x_1, x_2, \dots, x_T)$ для послідовності станів прихованого ланцюга Маркова $\{X_n\}_{n \geq 1}$.

Метою задачі оцінювання виступає пошук ймовірності отримати заданий ланцюжок спостережень при відомій моделі $M = (\pi, A, p, a)$. Для ефективного обчислення, як вже було зазначено раніше, можна використовувати алгоритм прямого та зворотнього ходу. Для застосування цього алгоритму використаємо змінні прямого ходу. Почнемо виведення рекурентних співвідношень для коефіцієнтів прямого ходу. Означимо

$$\alpha_t(i) = p(y_0, y_1, \dots, y_t, X_t = i).$$

Використовуючи формулу повної ймовірності продовжимо наступним чином

$$\alpha_{t+1}(j) = \sum_{i=1}^2 p(y_0, y_1, \dots, y_{t+1}, X_t = i, X_{t+1} = j) =$$

$$\begin{aligned}
&= \sum_{i=1}^2 p(y_0, y_1, \dots, y_t, X_t = i) P(X_{t+1} = j \mid X_t = i) \times \\
&\quad \times P(y_{t+1} \mid x_{t+1} = j, y_0, y_1, \dots, y_t) = \\
&= \sum_{i=1}^2 \alpha_t(i) a_{ij} b_j(y_{t+1} \mid y_t), \tag{2.1}
\end{aligned}$$

де $b_j(y \mid z)$ задано в (1.15).

Оскільки ми вважаємо, що $Y_0 \sim \mathcal{N}(0, 1)$ і не залежить від $\{X_n\}_{n \geq 1}$, то із рівняння (2.1) остаточно отримуємо рекурентні співвідношення для коефіцієнтів алгоритму прямого ходу

$$\begin{aligned}
\alpha_1(j) &= P(Y_0 = y_0) P(X_1 = j) P(Y_1 = y_1 \mid Y_0 = y_0, X_1 = j) = \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{y_0^2}{2}} \pi_j b_j(y_1 \mid y_0), \quad j = \overline{1, 2}
\end{aligned}$$

$$\alpha_{t+1}(j) = \sum_{i=1}^2 \alpha_t(i) \cdot a_{ij} b_j(y_{t+1} \mid y_t), \quad t = \overline{1, T-1}, j = \overline{1, 2} \tag{2.2}$$

Обчисливши всі коефіцієнти алгоритму прямого ходу, ми можемо отримати розв'язок задачі оцінювання. Тобто ми можемо записати, що ймовірність отримати ланцюжок спостережень $y = (y_0, y_1, \dots, y_T)$ буде рівна

$$P(Y_0 = y_0, Y_1 = y_1, \dots, Y_T = y_T \mid M) = \sum_{j=1}^2 \alpha_T(j). \tag{2.3}$$

Оскільки задача має бути розв'язана за допомогою чисельних методів [6], то коефіцієнти прямого ходу під час рекурсивного обчислення

слід масштабувати в кожний момент часу. Це унеможливить виникнення проблеми роботи із надмалими величинами, які сучасні обчислювальні машини не мають змоги обробити.

Відшкальовані змінні прямого ходу будемо позначати як $\hat{\alpha}_t(j), t = \overline{1, T}, j = \overline{1, 2}$. Тоді маємо, що

$$\hat{\alpha}_1(j) = c_1 \cdot \alpha_1(j), \quad j = \overline{1, 2},$$

де c_1 коефіцієнт шкалювання в момент часу $t = 1$, визначений наступним чином

$$c_1 = \frac{1}{\sum_{j=1}^2 \alpha_1(j)}.$$

Позначимо через c_τ коефіцієнт шкалювання у моменти часу $\tau = \overline{2, T}$, які визначаються як

$$c_\tau = \frac{1}{\sum_{j=1}^2 \hat{\alpha}_\tau(j)}, \quad \tau = \overline{1, T}. \quad (2.4)$$

Рекурентні співвідношення для коефіцієнтів алгоритму прямого ходу після шкалювання мають наступний вигляд

$$\hat{\alpha}_{t+1}(j) = c_t \sum_{i=1}^2 \hat{\alpha}_t(i) \cdot a_{ij} b_j(y_{t+1} | y_t), \quad t = \overline{1, T-1}, j = \overline{1, 2} \quad (2.5)$$

Тоді розв'язок задачі оцінювання, тобто обчислення ймовірності отримати заданий ланцюжок спостережень $y = (y_0, y_1, \dots, y_T)$, буде обчислюватись наступним чином

$$P(Y_0 = y_0, Y_1 = y_1, \dots, Y_T = y_T | M) = \frac{1}{\prod_{\tau=1}^T c_\tau}. \quad (2.6)$$

Хоч задачу оцінювання було розв'язано, проте треба вивести рекурентні співвідношення для коефіцієнтів зворотнього ходу, які буде використано при розв'язанні задачі навчання. Означимо

$$\beta_t(j, y_t) = p_{Y_{t+1}, \dots, Y_T | X_t, Y_t}(y_{t+1}, \dots, y_T \mid X_t = j, Y_t = y_t)$$

— як умовну щільність набору $(y_{t+1}, y_{t+2}, \dots, y_T)$ при заданих значеннях $X_t = j$ та $Y_t = y_t$. Зробивши всеможливі припущення щодо значення X_t , отримаємо

$$\begin{aligned} \beta_{t-1}(j, y_{t-1}) &= \sum_{i=1}^2 P(y_t, \dots, y_T, X_t = i \mid X_{t-1} = j, Y_{t-1} = y_{t-1}) = \\ &= \sum_{i=1}^2 a_{ji} b_j(y_t \mid y_{t-1}) \beta_t(j, y_t). \end{aligned} \quad (2.7)$$

Тут ми врахували те, що припущення про умовну незалежність спостережень при заданих станах ланцюга для авторегресійного рівняння не виконується. Використовуючи рівняння (2.7) остаточно матимемо

$$\beta_T(j, y_T) = 1, \quad j = \overline{1, 2}$$

$$\beta_{t-1}(j, y_{t-1}) = \sum_{i=1}^2 a_{ji} b_i(y_t \mid y_{t-1}) \beta_t(i, y_t), \quad t = \overline{2, T}, j = \overline{1, 2} \quad (2.8)$$

Не зайвим буде вказати, що ймовірність отримати ланцюжок спостережень $y = (y_0, y_1, \dots, y_T)$ можна обчислити за допомогою коефіцієнтів алгоритму зворотнього ходу також. Наступне співвідношення відображає цю можливість

$$P(Y_0 = y_0, Y_1 = y_1, \dots, Y_T = y_T \mid M) = p(y_0) \sum_{j=1}^2 \pi_j b_j(y_1 \mid y_0) \beta_1(j, y_1). \quad (2.9)$$

Як і при обчисленні коефіцієнтів прямого ходу, при обчисленні ми маємо використовувати нормуючі множники. У випадку із коефіцієнтами зворотнього ходу ми маємо використовувати саме ті множники, що були для коефіцієнтів алгоритму прямого ходу. Отже, остаточний вигляд рекурентних співвідношень буде таким:

$$\hat{\beta}_{t-1}(j, y_{t-1}) = c_{t-1} \cdot \sum_{i=1}^2 a_{ji} b_i(y_t | y_{t-1}) \hat{\beta}_t(i, y_t), \quad t = \overline{2, T}, j = \overline{1, 2}. \quad (2.10)$$

2.2 Розв'язання задачі декодування

Розв'яжемо задачу декодування за допомогою алгоритму Вітербі (1.5). Знайдена послідовність $\hat{X}$ станів прихованого ланцюга Маркова $\{X_n\}_{n \geq 1}$ матиме вигляд

$$\hat{X} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_T) = \arg \max_{x=(x_1, x_2, \dots, x_T)} P_M(X = x | Y = y), \quad (2.11)$$

та буде використана для прогнозування поведінки заданої моделі $M = (\pi, A, B)$.

2.3 Розв'язання задачі навчання

Розв'язком задачі навчання є оцінена за заданим ланцюжком спостережень $y = (y_0, y_1, \dots, y_T)$ модель $M^* = (\pi^*, A^*, p^*, a^*)$. Задля отримання такої моделі треба побудувати оцінки кожного з її параметрів. Для початку означимо вектори значень випадкових величин мультиплікативного шуму як $z = (z_1, z_2, \dots, z_T)$.

Оптимальну модель $M^* = (\pi^*, A^*, p^*, a^*)$ шукаємо як оцінку максимальної правдоподібності

$$M^* = \arg \max_M P_M(Y = y), \quad (2.12)$$

де $Y = (Y_1, \dots, Y_T)$, $y = (y_1, \dots, y_T)$ – спостережені значення.

Зауважимо, що функцію правдоподібності $P_M(Y = y)$ можна подати у вигляді

$$P_M(Y = y) = \sum_{\substack{x=(x_1, \dots, x_T) \\ e=(e_1, \dots, e_T)}} P_M(Y = y, X = x, \eta = e),$$

де $X = (X_1, \dots, X_T)$, $\eta = (\eta_1, \dots, \eta_T)$. Зважаючи на незалежність ланцюга Маркова $\{X_n\}_{n \geq 1}$ та послідовності $\{\eta_n\}_{n \geq 1}$, можемо записати:

$$\begin{aligned} P_M(Y = y) &= \sum_{\substack{x=(x_1, \dots, x_T) \\ e=(e_1, \dots, e_T)}} p_M(X = x) P_M(\eta = e) \times \\ &\quad \times p_M(Y = y \mid X = x, \eta = e). \end{aligned}$$

Оскільки $P_M(X = x) = \pi_{x_1} \prod_{i=1}^{T-1} a_{x_i x_{i+1}}$, $P_M(\eta = e) = \prod_{i=1}^T \prod_{k=1}^3 p_k^{\mathbb{I}(e_i = a_k)}$ і $P_M(Y = y \mid X, \eta) = \prod_{i=1}^T b_{x_i e_i}(y_i \mid y_{i-1})$, то

$$\begin{aligned} &= \sum_{\substack{x=(x_1, \dots, x_T) \\ e=(e_1, \dots, e_T)}} \pi_{x_1} \prod_{i=1}^{T-1} a_{x_i x_{i+1}} \times \prod_{i=1}^T \prod_{k=1}^3 p_k^{\mathbb{I}(e_i = a_k)} \prod_{i=1}^T b_{x_i e_i}(y_i \mid y_{i-1}) \end{aligned}$$

Максимізацію функції правдоподібності $P_M(Y = y)$ такого складного вигляду можна здійснити за допомогою ітераційного алгоритму навчання Баума-Велша [6]. Задавши деяке початкове наближення моделі $M^{(0)}$, кожен наступну модель $M^{(n+1)}$, $n \geq 0$ шукаємо з умови

$$M^{(n+1)} = \arg \max_M Q(M^{(n)}, M), \quad (2.13)$$

де

$$Q(M^{(n)}, M) = \sum_{\substack{x=(x_1, \dots, x_T) \\ e=(e_1, \dots, e_T)}} P_{M^{(n)}}(Y = y, X = x, \eta = e) \times \\ \times \ln(P_M(Y = y, X = x, \eta = e)). \quad (2.14)$$

Відомо [4], що ітераційний алгоритм (2.13), (2.14) є збіжним і дозволяє знайти аргумент, при якому досягається локальний максимум функції правдоподібності $P_M(Y = y)$.

2.3.1 Знаходження функції правдоподібності

Лема 2.1. *Функція правдоподібності $L_M(y, x, z)$ для набору даних $y_0, y_1, \dots, y_T$ має вигляд*

$$L_M(y, x, z) = \pi_{x_1} \prod_{i=1}^{T-1} a_{x_i x_{i+1}} \prod_{t=1}^T \prod_{k=1}^3 p_k^{\mathbb{I}(z_t = a_k)} \frac{1}{\sqrt{2\pi}} e^{-\frac{y_0^2}{2}} \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_t - (C_{x_t} + z_t)y_{t-1})^2}$$

Доведення. Запишемо функцію правдоподібності для повних даних як ймовірність спостерігати їх при невідомій послідовності станів прихованого ланцюга Маркова x та значеннях шуму z . Ця функція матиме такий вигляд

$$L_M(y, x, z) = P(y, x, z \mid M). \quad (2.15)$$

Зі зроблених вище припущень про модель можемо записати таку формулу для функції правдоподібності

$$L_M(y, x, z) = P(x)P(z)P(y_0) \prod_{t=1}^T P(y_t \mid x_t, z_t, y_{t-1}) \quad (2.16)$$

Скінченновимірний розподіл прихованого ланцюга Маркова описується вектором початкового розподілу та матрицею перехідних

ймовірностей наступним чином

$$P(x) = \pi_{x_1} \prod_{t=1}^{T-1} a_{x_t x_{t+1}}. \quad (2.17)$$

Ймовірність $P(z_1, z_2, \dots, z_T)$ спостерігати послідовність z знаходиться за формулою

$$P(z) = \prod_{t=1}^T p_1^{\mathbb{I}(z_t=a_1)} p_2^{\mathbb{I}(z_t=a_2)} p_3^{\mathbb{I}(z_t=a_3)}. \quad (2.18)$$

Оскільки $Y_0 \sim \mathcal{N}(0, 1)$, то ймовірність $P(y_0)$ можна записати як

$$P(y_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_0^2}{2}}. \quad (2.19)$$

Для того, щоб виразити ймовірність спостереження ланцюжка $y = (y_0, y_1, \dots, y_T)$ при заданому наборі станів прихованого ланцюга Маркова, скористаймось означеними змінними (1.15) переходу від спостереження до спостереження. Тож матимемо наступні формули

$$\prod_{t=1}^T p(y_t \mid x_t, z_t, y_{t-1}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_t - (C_{x_t} + z_t)y_{t-1})^2} \quad (2.20)$$

Отже, функція правдоподібності матиме вигляд

$$L_M(y, x, z) = \pi_{x_1} \prod_{i=1}^{T-1} a_{x_i x_{i+1}} \prod_{t=1}^T \prod_{k=1}^3 p_k^{\mathbb{I}(z_t=a_k)} \frac{1}{\sqrt{2\pi}} e^{-\frac{y_0^2}{2}} \prod_{t=1}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_t - (C_{x_t} + z_t)y_{t-1})^2} \quad (2.21)$$

□

2.3.2 Знаходження функції квазі-лог правдоподібності

Лема 2.2. *Функції квазі-логарифмічної правдоподібності для набору даних $y_0, y_1, \dots, y_T$ має вигляд*

$$\begin{aligned}
Q(M, M^*) &= \sum_{x,z} L_M(y, x, z) \ln L_{M^*}(y, x, z) = \sum_{i=1}^2 \ln \pi_i^* P_M(y, x_1 = i) + \\
&+ \sum_{i=1}^2 \sum_{t=1}^{T-1} \ln a_{ij}^* P_M(y, x_t = i, x_{t+1} = j) + \sum_{k=1}^3 \sum_{t=1}^T \ln p_k^* P_M(y, z_t = a_k) - \\
&- ((T+1) \ln \sqrt{2\pi} + \frac{y_0^2}{2}) \cdot P(y) - \\
&- \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^2 \sum_{k=1}^3 (y_t - (C_i + a_k^*) y_{t-1})^2 P_M(y, x_t = i, z_t = a_k).
\end{aligned}$$

Доведення. Запишемо логарифм функції правдоподібності. Задля спрощення підрахунку зробимо це частинами. Логарифмуючи формулу (2.17) отримаємо

$$\ln P(x) = \sum_{i=1}^2 \ln \pi_i \mathbb{I}(x_1 = i) + \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^{T-1} \ln a_{ij} \mathbb{I}(x_t = i, x_{t+1} = j) \quad (2.22)$$

Наступною прологарифмуємо формулу (2.18) та отримаємо

$$\ln P(z) = \sum_{k=1}^3 \sum_{t=1}^T \ln p_k \mathbb{I}(z_t = a_k) \quad (2.23)$$

Завершимо логарифмування формулами (2.19) та (2.20)

$$\begin{aligned}
\ln P(y_0) P(y_1, \dots, y_T \mid x, y_0) &= -((T+1) \ln \sqrt{2\pi} + \frac{y_0^2}{2}) - \\
&- \frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^3 \sum_{t=1}^T (y_t - (c_i a_k) y_{t-1})^2 \mathbb{I}(x_t = i, z_t = a_k)
\end{aligned} \quad (2.24)$$

Отже, об'єднавши результати обчислень формул (2.22)-(2.24), отримаємо, що логарифм функції правдоподібності матиме вигляд

$$\begin{aligned} \ln L_M(y, x, z) = & \sum_{i=1}^2 \ln \pi_i \mathbb{I}(x_1 = i) + \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^{T-1} \ln a_{ij} \mathbb{I}(x_t = i, x_{t+1} = j) + \\ & + \sum_{k=1}^3 \sum_{t=1}^T \ln p_k \mathbb{I}(z_t = a_k) - ((T+1) \ln \sqrt{2\pi} + \frac{y_0^2}{2}) - \\ & - \frac{1}{2} \sum_{i=1}^2 \sum_{k=1}^3 \sum_{t=1}^T (y_t - (c_i + a_k) y_{t-1})^2 \mathbb{I}(x_t = i, z_t = a_k) \quad (2.25) \end{aligned}$$

Тепер ми можемо записати явний вигляд функції квазі-логарифмічної правдоподібності як

$$\begin{aligned} Q(M, M^*) = & \sum_{x,z} L_M(y, x, z) \ln L_{M^*}(y, x, z) = \sum_{i=1}^2 \ln \pi_i^* P_M(y, x_1 = i) + \\ & + \sum_{i=1}^2 \sum_{t=1}^{T-1} \ln a_{ij}^* P_M(y, x_t = i, x_{t+1} = j) + \sum_{k=1}^3 \sum_{t=1}^T \ln p_k^* P_M(y, z_t = a_k) - \\ & - ((T+1) \ln \sqrt{2\pi} + \frac{y_0^2}{2}) \cdot P(y) - \\ & - \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^2 \sum_{k=1}^3 (y_t - (C_i + a_k^*) y_{t-1})^2 P_M(y, x_t = i, z_t = a_k). \end{aligned}$$

□

Для того, щоб отримати оцінку параметрів π_i^* при $i = \overline{1,2}$ скористаймось методом множників Лагранжа. Функція Лагранжа в такому випадку матиме вигляд

$$\mathcal{L}(\pi_1^*, \pi_2^*, \theta) = Q(M, M^*) - \theta \cdot (\pi_1^* + \pi_2^* - 1).$$

Звідси маємо систему рівнянь

$$\begin{cases} \frac{\partial \mathcal{L}(\pi_1^*, \pi_2^*, \theta)}{\partial \pi_1^*} = 0 \\ \frac{\partial \mathcal{L}(\pi_1^*, \pi_2^*, \theta)}{\partial \pi_2^*} = 0 \\ \frac{\partial \mathcal{L}(\pi_1^*, \pi_2^*, \theta)}{\partial \theta} = 0 \end{cases}$$

Продиференціювавши функцію по відповідним параметрам, отримаємо наступну систему

$$\begin{cases} \frac{P_M(y, x_1=1)}{\pi_1^*} = \theta \\ \frac{P_M(y, x_1=1)}{\pi_2^*} = \theta \\ \pi_1^* + \pi_2^* = 1 \end{cases}$$

З цієї системи нескладними перетвореннями отримаємо формулу для переоцінки параметрів вектора початкового розподілу для $i = \overline{1, 2}$

$$\pi_i^* = \frac{P_M(y, X_1 = i)}{P_M(y)}. \quad (2.26)$$

Перейдемо до оцінки параметрів матриці перехідних ймовірностей. Для цього задамо функцію Лагранжа наступним чином

$$\mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2) = Q(M, M^*) - \theta_1 \cdot (a_{11}^* + a_{12}^* - 1) - \theta_2 \cdot (a_{21}^* + a_{22}^* - 1).$$

Маємо систему рівнянь

$$\begin{cases} \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial a_{11}^*} = 0 \\ \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial a_{12}^*} = 0 \\ \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial a_{21}^*} = 0 \\ \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial a_{22}^*} = 0 \\ \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial \theta_1} = 0 \\ \frac{\partial \mathcal{L}(a_{11}^*, a_{12}^*, a_{21}^*, a_{22}^*, \theta_1, \theta_2)}{\partial \theta_2} = 0 \end{cases}$$

Продиференціювавши, отримаємо наступну систему

$$\begin{cases} \frac{\sum_{t=1}^{T-1} P_M(y, X_t=1, X_{t+1}=1)}{\sum_{t=1}^{T-1} P_M(y, X_t=1)} = \theta_1 \\ \frac{\sum_{t=1}^{T-1} P_M(y, X_t=1, X_{t+1}=2)}{\sum_{t=1}^{T-1} P_M(y, X_t=1)} = \theta_1 \\ \frac{\sum_{t=1}^{T-1} P_M(y, X_t=2, X_{t+1}=1)}{\sum_{t=1}^{T-1} P_M(y, X_t=2)} = \theta_2 \\ \frac{\sum_{t=1}^{T-1} P_M(y, X_t=2, X_{t+1}=2)}{\sum_{t=1}^{T-1} P_M(y, X_t=2)} = \theta_2 \\ a_{11}^* + a_{12}^* = 1 \\ a_{21}^* + a_{22}^* = 1 \end{cases}$$

З останньої системи отримаємо наступні співвідношення для переоцінки елементів матриці перехідних ймовірностей для будь-яких $i, j = \overline{1, 2}$

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} P_M(y, X_t = i, X_{t+1} = j)}{\sum_{t=1}^{T-1} P_M(y, X_t = i)} \quad (2.27)$$

Параметри мультиплікативного шуму p_k^* отримаємо диференціюванням функції квазі-логарифмічної правдоподібності

$$p_k^* = \frac{\sum_{t=1}^T P_M(y, z_t = a_k)}{\sum_{k=1}^3 \sum_{t=1}^T P_M(y, z_t = a_k)} \quad (2.28)$$

Продиференціювавши функцію квазі-лог правдоподібності за параметрами a_k^* та прирівнявши до нуля, отримаємо наступне співвідношення

$$\frac{\partial Q(M, M^*)}{\partial a_k^*} = \sum_{t=1}^T \sum_{i=1}^2 (y_t - (C_i + a_k^*) y_{t-1}) y_{t-1} \cdot P_M(y, X_t = i, z_t = a_k) = 0.$$

З цього співвідношення отримаємо наступні оцінки при $k = \overline{1,3}$

$$a_k^* = \frac{\sum_{t=1}^T \sum_{i=1}^2 (y_t - (C_i + a_k) y_{t-1}) y_{t-1} \cdot P_M(y, X_t = i, z_t = a_k)}{\sum_{t=1}^T y_{t-1}^2 \cdot P_M(y, z_t = a_k)} \quad (2.29)$$

2.3.3 Оцінювання параметрів моделі

Ми отримали співвідношення для оцінки параметрів заданої моделі $M = (\pi, A, p, a)$, проте в такій формі їх буде все ще складно обчислити за допомогою чисельних методів. Для підвищення ефективності обчислень параметрів пропонується виразити оцінки через відомі параметри моделі та коефіцієнти алгоритму прямого та зворотнього ходу.

Ймовірність того, що ланцюжок прихованих станів розпочнеться зі стану i може бути переозначена наступним чином

$$P_M(y, X_1 = i) = \alpha_1(i) \beta_1(i, y_1).$$

Ймовірність отримання ланцюжка спостережень y , починаючи із будь-якого стану прихованого ланцюга Маркова визначатиметься наступним співвідношенням

$$P_M(y) = \sum_{i=1}^2 \alpha_T(i).$$

Виражаючи ймовірність переходу від стану i в момент часу t до стану j в наступний момент часу $t + 1$, можна отримати наступну рівність

$$P_M(y, X_t = i, X_{t+1} = j) = \alpha_t(i) a_{ij} b_j(y_{t+1} | y_t) \beta_{t+1}(j, y_{t+1}).$$

Ймовірність отримати ланцюжок спостережень y при тому, що у момент часу t прихований ланцюг Маркова знаходитиметься у стані i виражається наступним чином

$$P_M(y, X_t = i) = \alpha_t(i)\beta_t(i, y_t).$$

Розглядаючи ймовірність отримати ланцюжок спостережень y зі значення мультиплікативного шуму в момент часу t рівним a_k більш детально, можемо отримати таку рівність

$$\begin{aligned} P_M(y, z_t = a_k) &= \\ &= \sum_{i=1}^2 \sum_{j=1}^2 P_M(y_0, \dots, y_{t-1}, x_{t-1} = i, x_t = j, y_t, z_t = a_k, y_{t+1}, \dots, y_T) = \\ &= \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) \cdot a_{ij} \cdot p_k \cdot b_{jk}(y_t \mid y_{t-1}) \beta_t(j, y_t), \end{aligned}$$

де, нагадаємо, $b_{jk}(y \mid z)$ задаються в (1.15).

В свою чергу, ймовірність отримати ланцюжок спостережень y зі значенням мультиплікативного шуму рівним a_k та станом j прихованого ланцюга Маркова в момент часу t обчислюється наступним чином

$$\begin{aligned} P(y, x_t = i, z_t = a_k) &= \sum_{j=1}^2 P_M(y, x_{t-1} = j, x_t = i, z_t = a_k) = \\ &= \sum_{j=1}^2 \alpha_{t-1}(j) \cdot a_{ji} \cdot p_k \cdot b_{jk}(y_t \mid y_{t-1}) \beta_t(i, y_t). \end{aligned}$$

Враховуючи всі знайдені ймовірності, можна тепер записати формули для переоцінки параметрів у більш зручному для використання вигляді. Формула для переоцінки елементів вектора початкового розподілу при $i = \overline{1,2}$ буде такою

$$\pi_i^* = \frac{\alpha_1(i)\beta_1(i, y_1)}{\sum_{j=1}^2 \alpha_T(j)}.$$

Переоцінені елементи матриці перехідних ймовірностей при $i, j = \overline{1,2}$ виглядатимуть наступним чином

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(y_{t+1} | y_t) \beta_{t+1}(j, y_{t+1})}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i, y_t)}.$$

Переписуючи формули переоцінки параметрів мультиплікативного шуму, отримаємо наступні вирази для $k = \overline{1,3}$

$$p_k^* = \frac{\sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} \cdot p_k \cdot b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}{\sum_{k=1}^3 \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} \cdot p_k \cdot b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}$$

$$a_k^* = \frac{\sum_{t=1}^T \sum_{i=1}^2 (y_t - C_i y_{t-1}) y_{t-1} \cdot \sum_{n=1}^2 \alpha_{t-1}(n) a_{ni} p_k b_{i,k}(y_t | y_{t-1}) \beta_t(i, y_t)}{\sum_{t=1}^T y_{t-1}^2 \cdot \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} \cdot p_k \cdot b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}$$

Отримані формули описують, як будуть переоцінюватись параметри моделі M за допомогою немасштабованих коефіцієнтів прямого та зворотнього ходу. Саме тому треба знайти їх вигляд в термінах відшкальованих змінних.

Спочатку знайдемо елементи вектора початкового розподілу π при $i = \overline{1,2}$ та отримаємо наступні результати

$$\pi_i^* = \frac{\hat{\alpha}_1(i) \hat{\beta}_1(i)}{\sum_{j=1}^2 \hat{\alpha}_T(j)} = \frac{\left(\prod_{\tau=1}^1 c_\tau\right) \alpha_1(i) \left(\prod_{\tau=2}^T c_\tau\right) \beta_1(i, y_1)}{\left(\prod_{\tau=1}^T c_\tau\right) \sum_{j=1}^2 \alpha_T(j)} = \frac{\alpha_1(i) \beta_1(i, y_1)}{\sum_{j=1}^2 \alpha_T(j)}.$$

Отже, перейдемо до оцінювання елементів матриці перехідних ймовірностей для $i, j = \overline{1,2}$

$$a_{ij}^* = \frac{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) a_{ij} b_j(y_{t+1} | y_t) \hat{\beta}_{t+1}(j, y_{t+1})}{\sum_{t=1}^{T-1} \hat{\alpha}_t(i) \hat{\beta}_t(i, y_t)} =$$

$$\begin{aligned}
&= \frac{\sum_{t=1}^{T-1} \left(\prod_{\tau=1}^t c_{\tau} \right) \alpha_t(i) a_{ij} b_j(y_{t+1} | y_t) \left(\prod_{\tau=t+1}^T c_{\tau} \right) \beta_{t+1}(j, y_{t+1})}{\sum_{t=1}^{T-1} \left(\prod_{\tau=1}^t c_{\tau} \right) \alpha_t(i) \left(\prod_{\tau=t+1}^{T-1} c_{\tau} \right) \beta_t(i, y_t)} = \\
&= \frac{\left(\prod_{\tau=1}^T c_{\tau} \right) \sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(y_{t+1} | y_t) \beta_{t+1}(j, y_{t+1})}{\left(\prod_{\tau=1}^T c_{\tau} \right) \sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i, y_t)} = \\
&= \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(y_{t+1} | y_t) \beta_{t+1}(j, y_{t+1})}{\sum_{t=1}^{T-1} \alpha_t(i) \beta_t(i, y_t)}.
\end{aligned}$$

Тепер оцінемо параметри мультиплікативного шуму. Формули для параметрів p_k^* при $k = \overline{1,3}$ отримаємо таким чином

$$\begin{aligned}
p_k^* &= \frac{\sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \hat{\alpha}_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \hat{\beta}_t(j, y_t)}{\sum_{k=1}^3 \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \hat{\alpha}_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \hat{\beta}_t(j, y_t)} = \\
&= \frac{\sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \left(\prod_{\tau=1}^T c_{\tau} \right) \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}{\sum_{k=1}^3 \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \left(\prod_{\tau=1}^T c_{\tau} \right) \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)} = \\
&= \frac{\left(\prod_{\tau=1}^T c_{\tau} \right) \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}{\left(\prod_{\tau=1}^T c_{\tau} \right) \sum_{k=1}^3 \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)} = \\
&= \frac{\sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}{\sum_{k=1}^3 \sum_{t=1}^T \sum_{i=1}^2 \sum_{j=1}^2 \alpha_{t-1}(i) a_{ij} p_k b_{jk}(y_t | y_{t-1}) \beta_t(j, y_t)}.
\end{aligned}$$

Оцінка параметрів a_k^* відбувається за схожим принципом. Отже, оцінки застосовні для алгоритму Баума-Велша знайдено.

2.4 Висновки до другого розділу

В цьому розділі було знайдено вигляд функції правдоподібності та квазілогарифміної функції правдоподібності для отриманих даних та отримано явний вигляд формул для оцінювання параметрів авторегресійної моделі, які будуть застосовані в алгоритмі навчання Баума-Велша.

3 АНАЛІЗ РЕЗУЛЬТАТІВ ЧИСЕЛЬНОГО ЕКСПЕРИМЕНТУ

В цьому розділі наведено результати проведення чисельних обрахунків розв'язків задач теорії прихованих марковських моделей. Також досліджено вплив довжини ланцюга спостережень на якість розв'язання поставлених задач та проведено прогноз його еволюції.

3.1 Моделювання поведінки ланцюжка спостережень

В залежності від параметрів авторегресійної моделі, характер ланцюжка спостережень буде змінюватись. Так на графіку 3.1 наведено фрагмент спостережень двох різних часових рядів, поведінку яких було згенеровано наступними моделями:

$$M = (\pi = (0.14, 0.86), A = \begin{pmatrix} 0.78 & 0.22 \\ 0.372 & 0.628 \end{pmatrix}, \\ C = (0.23, 0.76), a = (0.05, 0.14, 0.22), \\ p = (0.66, 0.14, 0.20))$$

$$M_1 = (\pi = (0.63, 0.37), A = \begin{pmatrix} 0.31 & 0.69 \\ 0.22 & 0.78 \end{pmatrix}, \\ C = (0.68, 0.52), a = (0.16, -0.14, 0.26), \\ p = (0.22, 0.46, 0.32))$$

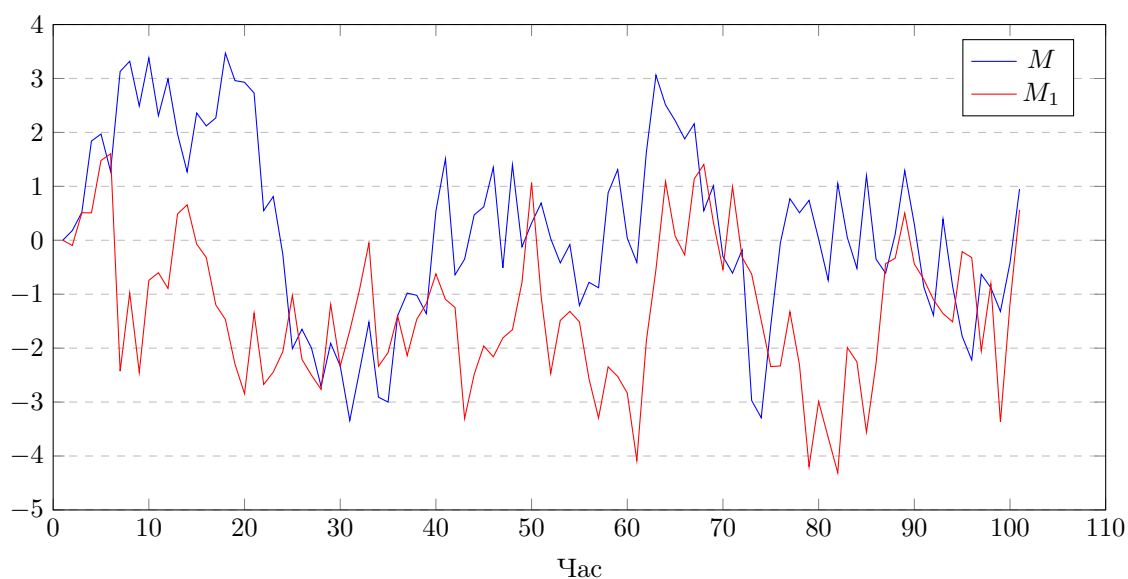


Рисунок 3.1 – Приклад спостережень при різних параметрах моделей.

3.2 Аналіз розв’язку задач оцінювання

Отримати значення функції правдоподібності із ростом кількості спостережень стає дедалі складніше, через значне зменшення досліджуваної величини. Натомість пропонується розглянути логарифм цієї функції. Шукана ймовірність обчислюється згідно формул 2.3 та 2.9 за допомогою коефіцієнтів алгоритмів прямого та зворотнього ходу.

3.3 Аналіз розв’язку задачі декодування

Найімовірніша послідовність станів прихованого ланцюга Маркова для заданої моделі при певному наборі спостережень оцінюється за допомогою алгоритму Вітербі. Точність оцінки такого алгоритму залежить від довжини ланцюжка спостережень. Для подальшого використання моделей позначимо їх індексом, що відповідає степеню десяти в довжині ланцюга. Так, наприклад, оцінена на 10^4 ітераціях модель позначатиметься як M_4 . Таблиця 3.1 відображає залежність точності оцінки послідовності $\hat{X}$ від довжини ланцюжка спостережень.

Таблиця 3.1 – Порівняльна таблиця точності оцінок алгоритму Вітербі

Кількість спостережень	10^2	10^3	10^4	10^5
Точність оцінки	57%	72%	77%	84%

Точність оцінки визначаємо як величину

$$\delta = \frac{1}{T} \sum_{t=1}^T \mathbb{I}(X_t = \hat{X}_t) \cdot 100\%.$$

3.4 Аналіз розв’язку задачі навчання

Вважаючи параметри моделі M шуканими та спираючись на ефективні методи обчислень коефіцієнтів прямого та зворотнього ходу, засобами мови програмування Python було реалізовано алгоритм навчання Баума-Велша [8].

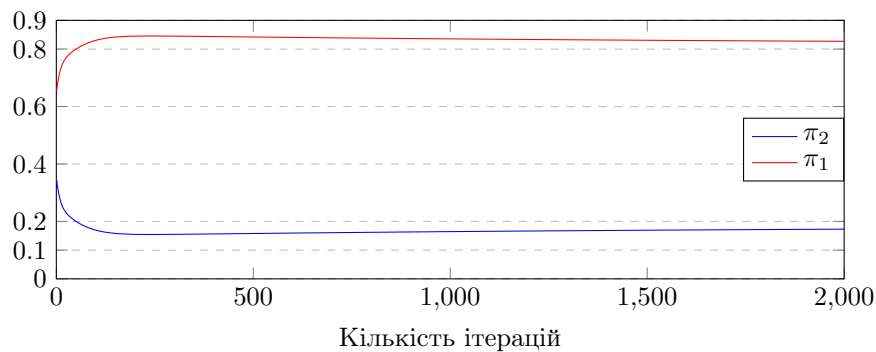


Рисунок 3.2 – Графік збіжності початкового розподілу.

Розглянемо збіжність оцінених параметрів моделі M до істинних на прикладі ланцюжка спостережень довжиною 10^5 на графіках, представлених на рисунках (3.2)-(3.5).

Щоб продемонструвати точність отриманих оцінок позначимо через $\delta_\pi, \delta_A, \delta_p$ та δ_a похибки вимірювання величин $\pi_i, A_{ij}, p_k, a_k, i = 1, 2, j = 1, 2, k = \overline{1, 3}$, що обчислюються за формулами

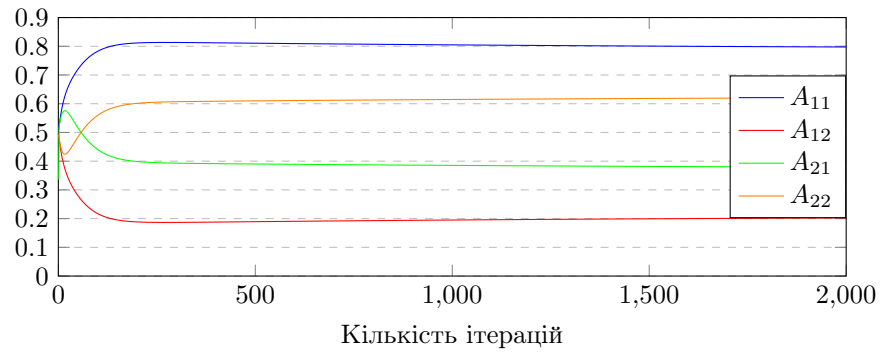


Рисунок 3.3 – Графік збіжності матриці перехідних ймовірностей.

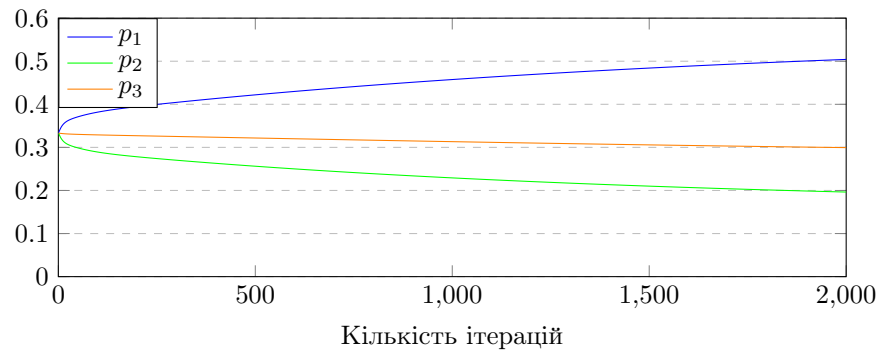


Рисунок 3.4 – Графік збіжності параметра p .

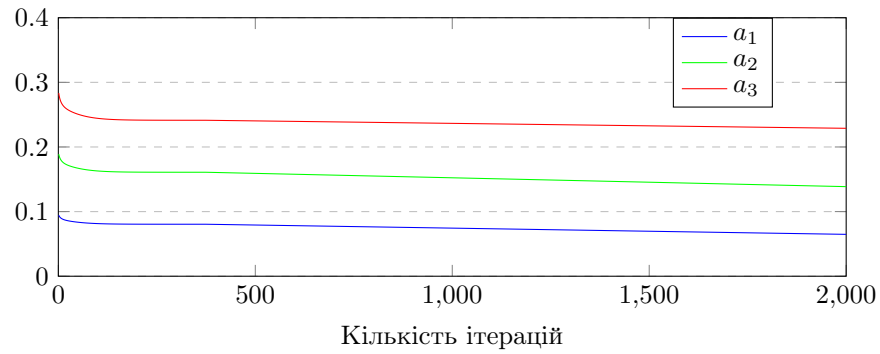


Рисунок 3.5 – Графік збіжності параметра a .

$$\delta_{\pi} = (| \pi_i - \hat{\pi}_i |), \delta_A = (| A_{ij} - \hat{A}_{ij} |)$$

$$\delta_p = (| p_k - \hat{p}_k |), \delta_a = (| a_i - \hat{a}_i |)$$

В таблиці 3.2 наведено значення похибок в залежності від довжини ланцюга спостережень.

Таблиця 3.2 – Порівняльна таблиця похибок оцінювання параметрів суміші

Модель	T	δ_π	δ_A	δ_p	δ_a
M_2	10^2	(0.29035, 0.29036)	$\begin{pmatrix} 0.00173 & 0.00172 \\ 0.09718 & 0.09717 \end{pmatrix}$	$\begin{pmatrix} 0.09706 \\ 0.0039 \\ 0.10094 \end{pmatrix}$	$\begin{pmatrix} 0.08871 \\ 0.0226 \\ 0.02161 \end{pmatrix}$
M_3	10^3	(0.17785, 0.17786)	$\begin{pmatrix} 0.02523 & 0.02524 \\ 0.06092 & 0.06093 \end{pmatrix}$	$\begin{pmatrix} 0.0139 \\ 0.00803 \\ 0.02191 \end{pmatrix}$	$\begin{pmatrix} 0.02907 \\ 0.01815 \\ 0.01722 \end{pmatrix}$
M_4	10^4	(0.08484, 0.08485)	$\begin{pmatrix} 0.00029 & 0.0003 \\ 0.10519 & 0.1052 \end{pmatrix}$	$\begin{pmatrix} 0.05091 \\ 0.00182 \\ 0.05272 \end{pmatrix}$	$\begin{pmatrix} 0.03312 \\ 0.02624 \\ 0.02936 \end{pmatrix}$
M_5	10^5	(0.00444, 0.00445)	$\begin{pmatrix} 0.02491 & 0.02492 \\ 0.01328 & 0.01329 \end{pmatrix}$	$\begin{pmatrix} 0.20252 \\ 0.0892 \\ 0.1133 \end{pmatrix}$	$\begin{pmatrix} 0.02441 \\ 0.01231 \\ 0.01653 \end{pmatrix}$

Також було проведено ряд випробувань з метою дослідження впливу довжини ланцюжка спостережень (T) та кількості ітерацій алгоритму навчання (I) на якість оцінок. Було взято моделі із різними початковими значеннями для отримання еквівалентних до M після оцінки. Введемо для кожної такої моделі позначення для однозначної ідентифікації, які разом з оцінками наведено у таблицях 3.3 та 3.4.

Таблиця 3.3 – Порівняльна таблиця істинних та оцінених параметрів моделі

Модель	T	I	π	A	p	a
M_2	10^2	10^2	(0.49533, 0.50466)	$\begin{pmatrix} 0.75010 & 0.24989 \\ 0.26305 & 0.73694 \end{pmatrix}$	$\begin{pmatrix} 0.38537 \\ 0.30719 \\ 0.30742 \end{pmatrix}$	$\begin{pmatrix} 0.03712 \\ 0.07424 \\ 0.11137 \end{pmatrix}$
		10^3	(0.45035, 0.54964)	$\begin{pmatrix} 0.77827 & 0.22172 \\ 0.27482 & 0.72517 \end{pmatrix}$	$\begin{pmatrix} 0.56294 \\ 0.13610 \\ 0.30094 \end{pmatrix}$	$\begin{pmatrix} 0.13871 \\ 0.16260 \\ 0.19839 \end{pmatrix}$

Таблиця 3.4 – Порівняльна таблиця істинних та оцінених параметрів моделі

Модель	T	I	π	A	p	a
M_3	10^3	10^2	$(0.33607, 0.66392)$	$\begin{pmatrix} 0.77641 & 0.22358 \\ 0.49256 & 0.50743 \end{pmatrix}$	$\begin{pmatrix} 0.30730 \\ 0.30388 \\ 0.38881 \end{pmatrix}$	$\begin{pmatrix} 0.08458 \\ 0.16917 \\ 0.25375 \end{pmatrix}$
		10^3	$(0.33785, 0.66214)$	$\begin{pmatrix} 0.80523 & 0.19476 \\ 0.43292 & 0.56707 \end{pmatrix}$	$\begin{pmatrix} 0.64610 \\ 0.13197 \\ 0.22191 \end{pmatrix}$	$\begin{pmatrix} 0.07907 \\ 0.15815 \\ 0.23722 \end{pmatrix}$
M_4	10^4	10^2	$(0.32216, 0.67783)$	$\begin{pmatrix} 0.78252 & 0.21747 \\ 0.54223 & 0.45776 \end{pmatrix}$	$\begin{pmatrix} 0.42140 \\ 0.27548 \\ 0.30310 \end{pmatrix}$	$\begin{pmatrix} 0.08356 \\ 0.16712 \\ 0.25068 \end{pmatrix}$
		10^3	$(0.24484, 0.75515)$	$\begin{pmatrix} 0.78029 & 0.21970 \\ 0.47719 & 0.52280 \end{pmatrix}$	$\begin{pmatrix} 0.60909 \\ 0.13818 \\ 0.25272 \end{pmatrix}$	$\begin{pmatrix} 0.08312 \\ 0.16624 \\ 0.24936 \end{pmatrix}$
M_5	10^5	10^2	$(0.16965, 0.83034)$	$\begin{pmatrix} 0.78260 & 0.21739 \\ 0.44048 & 0.55951 \end{pmatrix}$	$\begin{pmatrix} 0.38183 \\ 0.28916 \\ 0.32900 \end{pmatrix}$	$\begin{pmatrix} 0.08143 \\ 0.16287 \\ 0.24430 \end{pmatrix}$
		10^3	$(0.16444, 0.83555)$	$\begin{pmatrix} 0.80491 & 0.19508 \\ 0.38528 & 0.61471 \end{pmatrix}$	$\begin{pmatrix} 0.45748 \\ 0.22920 \\ 0.31330 \end{pmatrix}$	$\begin{pmatrix} 0.07441 \\ 0.15231 \\ 0.23653 \end{pmatrix}$
Істинні:			$(0.16, 0.84)$	$\begin{pmatrix} 0.78 & 0.22 \\ 0.372 & 0.628 \end{pmatrix}$	$\begin{pmatrix} 0.66 \\ 0.14 \\ 0.20 \end{pmatrix}$	$\begin{pmatrix} 0.05 \\ 0.14 \\ 0.22 \end{pmatrix}$

3.5 Побудова прогнозу

За оціненими параметрами моделі M побудовано прогноз із використанням моделі M_5 . Для навчання моделі було використано 70% від спостережених даних, а іншу частину — для перевірки прогнозу. Прогноз побудовано із використанням послідовності $\hat{X}$ — найбільш

ймовірної послідовності станів прихованого ланцюга Маркова, обрахованої за допомогою алгоритму Вітербі. Наявність такої послідовності дає змогу розпочати прогноз не у випадковому стані, а у найбільш ймовірному, що може покращити точність. Спрогнозовані значення обчислюються як математичне сподівання наступного вигляду

$$\hat{y}_t = \sum_{k=1}^3 (C_{x_t} + a_k) p_k \hat{y}_{t-1}.$$

Відповідність прогнозу такою моделлю подальшої поведінки дійсної моделі пропонуємо відстежити на наступному графіку 3.6. Значення середньоквадратичного відхилення $MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ рівне 1.09.

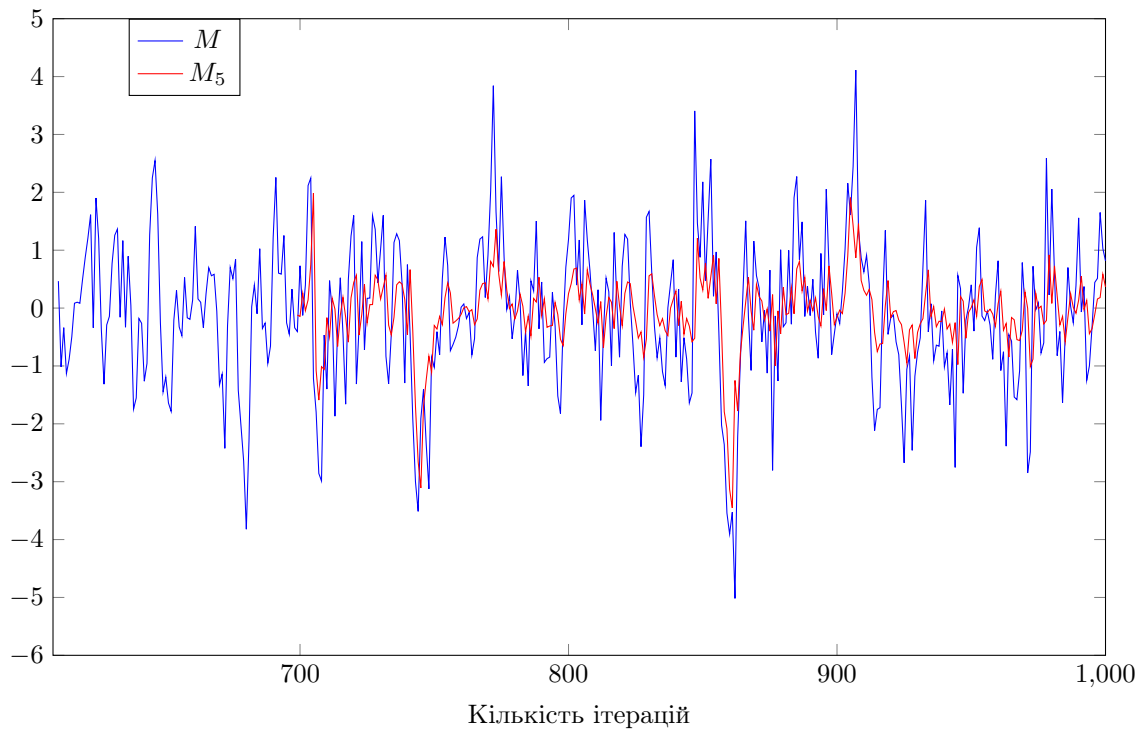


Рисунок 3.6 – Порівняння спостереженої та прогнозованої поведінки часового ряду.

3.6 Висновки до третього розділу

В цьому розділі були наведені результати роботи алгоритмів прямого та зворотнього ходу в поєднанні з алгоритмом Баума-Велша для розв'язання задач оцінювання та навчання. Також були представлені дані щодо застосування алгоритму Вітербі в розв'язанні задачі декодування. Показано збіжність досліджуваних параметрів до істинних, наведено таблиці, які демонструють якість оцінки в залежності від довжини ланцюга спостережень та кількості ітерацій алгоритму навчання. З'ясовано величину похибок оцінювання моделей в залежності від вищезгаданих змінних. Наприкінці розділу наведено графік із прогнозованою та дійсною послідовністю значень ланцюга спостережень.

ВИСНОВКИ

У ході даної роботи був проведений аналіз наявних джерел за даною тематикою, який показав те, що авторегресійні моделі застосовуються у багатьох напрямках науки. Особлива увага спрямована на дослідження впливу мультиплікативного шуму на них.

В роботі було знайдено оцінки максимальної правдоподібності параметрів рівняння та характеристик прихованого ланцюга Маркова, який керує перемиканням режимів авторегресійного рівняння. Для побудови оцінок було використано ітераційний алгоритм навчання Баума-Велша, алгоритм прямого-зворотнього ходу та алгоритм Вітербі.

Наведені результати чисельного експерименту демонструють збіжність отриманих оцінок до істинних значень параметрів та допомагають побудувати прогноз еволюції ланцюга спостережень із плином часу.

Надалі розподіл мультиплікативного шуму пропонується змінити із дискретного на неперервний, що дасть можливість більш загально описувати залежності між спостереженнями.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] Martial Coulon, Jean-Yves Tourneret та Ananthram Swami. «Detection of Multiplicative Noise in Stationary Random Processes Using Second- and Higher Order Statistics». В: (вер. 2000), с. 10. ISSN: 1053-587X. DOI: 10.1109/78.852007.
- [2] Zhiqiang Gao та Ronald J Patton. «Model-based adaptive fault detection for a class of uncertain nonlinear systems». В: *Sadhana* 24.4-5 (1999), с. 425—452.
- [3] Kazuki Hayashi та ін. «A New Autoregressive Neural Network Model with Command Compensation for Imitation Learning Based on Bilateral Control». В: бер. 2021, с. 1—7. DOI: 10.1109/ICM46511.2021.9385691.
- [4] D. Timo Koski. *Hidden Markov Models for Bioinformatics*. Boca Raton, FL: CRC Press, 2001, с. 83—365.
- [5] Harold J. Kushner та G. George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. 2-е вид. Springer, 2003.
- [6] Mikael Nilsson. *First Order Hidden Markov Model: Theory and Implementation Issues*. 2005, с. 24—33.
- [7] Federal Reserve Bank of San Francisco. «Autoregressive Models for Time-Series Analysis». В: *Economic Letter* 2014.14 (2014), с. 1—6.
- [8] Shyfrin. *Implementing the Baum-Welch Algorithm for Hidden Markov Models*. GitHub repository. Code available at <https://github.com/Schidev/supplement>. 2023.
- [9] Sang-Ha Sung та ін. «A Study on Cryptocurrency Log-Return Price Prediction Using Multivariate Time-Series Model». В: *Axioms* 11.9 (2022), с. 448. DOI: 10.3390/axioms11090448.

- [10] Joshua Peter Tobin. «Real-World Robotic Perception and Control Using Synthetic Data». ProQuest ID: Tobin_berkeley_0028E_18893. Merritt ID: ark:/13030/m5sv2sr8. Дис. . . . док. UC Berkeley, 2019.