НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»
Навчально-науковий фізико-технічний інститут

Кафедра математичного моделювання та аналізу даних

«На правах рукопису»
УДК 519.21

«До захисту допущено»
В.о. завідувача кафедри
__________ Іван ТЕРЕЩЕНКО
«__» __________ 2023 р.

Дипломна робота

на здобуття ступеня бакалавра
за освітньо-професійною програмою
«Математичні методи моделювання, розпізнавання образів та комп'ютерного зору»

зі спеціальності: 113 Прикладна математика
на тему: **«Рівняння авторегресії з марковськими перемиканнями»**

Виконав:
студент 4 курсу, групи ФІ-91
Мочук Олена Володимирівна __________

Керівник:
ст. викл. каф. ММАД
Наказной Павло Олександрович __________

Консультант:
доц. каф. ММЗІ, к. ф.-м. н.
Ніщенко Ірина Іванівна __________

Рецензент:
ст. викл. каф. ММЗІ, к. ф.-м. н.
Рябов Георгій Валентинович __________

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

Студент __________

Київ — 2023

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»
Навчально-науковий фізико-технічний інститут
Кафедра математичного моделювання та аналізу даних

Рівень вищої освіти — перший (бакалаврський)
Спеціальність — 113 Прикладна математика,
ОПП «Математичні методи моделювання, розпізнавання образів та комп'ютерного зору»

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
__________ Іван ТЕРЕЩЕНКО
«__» __________ 2023 р.

ЗАВДАННЯ
на дипломну роботу

Студент: Мочук Олена Володимирівна

1. Тема роботи: «*Рівняння авторегресії з марковськими перемиканнями*», науковий керівник дипломної роботи бакалавра: ст. викл. каф. ММАД Наказной Павло Олександрович,

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

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

3. Вихідні дані до роботи: опубліковані джерела за тематикою дослідження.

4. Об'єкт дослідження: рівняння авторегресії з марковськими перемиканнями.

5. Предмет дослідження: властивості рівняння авторегресії з марковськими перемиканнями та побудова оцінок його параметрів.

6. Перелік завдань:

1) провести пошук та аналіз опублікованих джерел за тематикою дослідження;

2) ознайомитися з прихованими марківськими моделями та з авторегресійними прихованими марківськими моделями;

3) дослідження існування стаціонарного розв'язку рівняння авторегресії з марківськими перемиканнями;
4) побудова оцінок параметрів прихованого марківського ланцюга, що керує перемиканнями у рівнянні;
5) програмна реалізація рівняння авторегресії з марківськими перемиканнями.

7. Орієнтовний перелік графічного (ілюстративного) матеріалу: презентація доповіді.

8. Орієнтовний перелік публікацій: планується доповідь на XXI Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики», публікація статті у збірнику з матеріалами конференції.

9. Дата видачі завдання: 10 вересня 2022 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи бакалавра	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	Дослідження моделі	Листопад-Грудень 2022 р.	Виконано
4	Знаходження оцінок параметрів моделі	Вересень-жовтень 2023 р.	Виконано
5	Програмна реалізація	Січень-березень 2023 р.	Виконано
6	Підготовка апробації	Квітень 2023 р.	Виконано
7	Оформлення результатів	Травень-червень 2023 р.	Виконано

Студент ____________ Олена МОЧУК

Керівник ____________ Павло НАКАЗНОЙ

РЕФЕРАТ

Обсяг роботи: 79 сторінок, 11 рисунків, 1 таблицю, 2 додатки, 8 джерел літератури.

Об'єктом дослідження є рівняння авторегресії першого порядку з марковськими перемиканнями.

Предметом дослідження є властивості рівняння авторегресії з марковськими перемиканнями та побудова оцінок його параметрів.

Вважаючи, що коефіцієнти рівняння, а також дисперсія шуму змінюються разом зі зміною стану деякого неспостережуваного ланцюга Маркова, за скінченним набором спостережень часового ряду побудовано оцінки коефіцієнтів рівняння, дисперсії шуму, а також матриці перехідних ймовірностей прихованого ланцюга Маркова.

Для розв'язування задачі використано математичний апарат прихованих марковських моделей.

ЛАНЦЮГ МАРКОВА, ПРИХОВАНІ МАРКОВСЬКІ МОДЕЛІ, АВТОРЕГРЕСІЙНЕ РІВНЯННЯ, МАРКОВСЬКІ ПЕРЕМИКАННЯ

ABSTRACT

Volume of work: 79 pages, 11 figures, 1 table, 2 appendices, 8 references.

The object of study is a first-order autoregressive equation with Markov switches.

The subject of the study is the properties of the autoregressive equation with Markov switches and the construction of estimates of its parameters.

Assuming that the coefficients of the equation and the noise variance change along with the change in the state of some unobserved Markov chain, estimates of the coefficients of the equation, the noise variance, and the transition probability matrix of the hidden Markov chain are constructed from a finite set of observations of the time series.

The mathematical apparatus of hidden Markov models is used to solve the problem.

MARKOV CHAIN, HIDDEN MARKOV MODELS, AUTOREGRESSIVE EQUATION, MARKOV SWITCHES

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AR — авторегресійна модель (англ. autoregressive model);

MA — модель ковзаючого середнього (англ. moving—average model);

ARMA — модель авторегресії та ковзаючого середнього (англ. autoregressive — moving—average model);

ПММ — прихована марківська модель (англ. hidden Markov model, HMM);

AR ПММ — авторегресійна прихована марківська модель (англ. autoregressive hidden Markov model, AR HMM)

ВСТУП

Для аналізу динаміки багатьох економічних та фінансових змінних як правило застосовують різноманітні моделі часових рядів. Найчастіше використовують такі лінійні моделі як авторегресійну модель (AR), модель ковзаючого середнього (MA) чи змішану ARMA модель [1].

Модель AR(p) записують як:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t,$$

де $\varphi_1, \ldots, \varphi_p$ є параметрами, c є сталою, а випадкова величина ε_t є білим шумом.

Для того щоб ця модель залишалася стаціонарною, для значень цих параметрів необхідні деякі обмеження. Наприклад, послідовність $\{X_t\}_{t\geqslant 0}$ в моделі AR(1) за умови $|\varphi_1| \geqslant 1$ стаціонарною не є.

Позначення MA(q) стосується моделі ковзного середнього порядку q:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i},$$

де $\theta_1, \ldots, \theta_q$ є параметрами моделі, c є математичним сподіванням X_t (що часто вважають рівним 0), а $\varepsilon_t, \varepsilon_{t-1}, \ldots$ є незалежними нормально розподіленими випадковими величинами.

Позначення ARMA(p, q) стосується моделі з p авторегресійними членами та q членами ковзного середнього. Ця модель містить моделі AR(p) та MA(q):

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

Загальну модель ARMA було описано 1951 року в дисертації Пітера Уіттла, який використовував математичний аналіз (ряд Лорана та аналіз

Фур'є) та статистичне висновування. Моделі ARMA було популяризовано книгою 1970 року Джорджа Бокса та Дженкінса, які виклали ітераційний метод (Бокса–Дженкінса).

Попри ефективність у багатьох застосуваннях такі моделі, однак, не спроможні адекватно описати дані у випадку, коли вони демонструють різку зміну у поведінці, зумовлену деякими непередбачуваними подіями [2].

Модель авторегресії з марковськими перемиканнями запропонована Гамільтоном у 1989 році, є однією з найпопулярніших нелінійних моделей, які здатні впоратися з аналізом змінних, динаміка яких суттєво залежить від режиму функціонування деякої системи. Вважається, що зміна режимів керується деяким прихованим ланцюгом Маркова. Відтак тривалість кожного режиму є випадковою величиною і стан прихованого ланцюга визначає параметри авторегресійного рівняння [3].

Модель Гамільтона можна записати у вигляді:

$$(Y_t - \mu_{X_t}) = \sum_{i=1}^{s} a_i \cdot (Y_{t-i} - \mu_{X_{t-i}}) + \varepsilon_t, \tag{$*$}$$

де $\{Y_t\}$ — спостережені дані; $a_1, \ldots, a_s$ — деякі константи; X_t — це режим в момент часу t; μ_{X_t} — константа, яка залежить від режиму X_t; $\{\varepsilon_t\}$ — незалежні однаково розподілені випадкові величини з гауссовим розподілом $\varepsilon_t \sim N(o, \sigma^2)$.

У дипломній роботі розглянуто випадок, коли в рівнянні $(*)$ від режиму залежить не лише константа μ_{X_t}, але й коефіцієнти a_i та дисперсія σ^2 шуму. А саме, ми розглядаємо авторегресійне рівняння першого порядку

$$Y_t = \mu_{X_t} + a_{X_t} \cdot Y_{t-1} + \varepsilon_t^{X_t},$$

де $\varepsilon_t^{X_t} \sim N\big(0, \sigma^2_{X_t}\big)$, μ_{X_t}, a_{X_t} — константи, які залежать від режиму X_t у момент часу t [4].

Вважаємо, що X_t — ланцюг Маркова з двома станами $\{1, 2\}$ з матрицею перехідних ймовірностей $(A_{ij})_{i,j=1,2}$ та початковим розподілом

$\pi = (\pi_1, \pi_2)$.

Задача полягає в оцінюванні невідомих констант μ_1, μ_2, a_1, a_2, σ_1^2, σ_2^2 та параметрів $(A_{ij})_{i,j=1,2}$, $\pi = (\pi_1, \pi_2)$ прихованого ланцюга $\{X_t\}$ за скінченним набором спостережень $y_1, \ldots, y_T$.

При розв'язуванні задачі ми застосуємо теорію прихованих марківських моделей (ПММ). Спочатку за ітераційним алгоритмом Баума–Велша знаходимо модель

$$\lambda^* = (\mu_1, \mu_2, a_1, a_2, \sigma_1, \sigma_2, A, \pi),$$

як оцінку максимальної правдоподібності

$$\lambda^* = \arg\max_{\lambda} P_\lambda(y_1, \ldots, y_T).$$

Потім, використовуючи отриману модель λ^*, будуємо прогноз і перевіряємо його якість на даних $(y_{T+1}, \ldots, y_{T+T_1})$, які не використовували для навчання.

Актуальність дослідження. Актуальність даного дослідження зумовлюється широким застосуванням моделей авторегресії при дослідженні часових рядів на фінансових ринках.

Метою дослідження є вивчення властивостей рівняння авторегресії з марковськими перемиканнями. Для досягнення мети необхідно вирішити такі **завдання**:

1. провести пошук та аналіз опублікованих джерел за тематикою дослідження;

2. ознайомитися з прихованими марківськими моделями та з авторегресійними прихованими марківськими моделями;

3. дослідження існування стаціонарного розв'язку рівняння авторегресії з марківськими перемиканнями;

4. побудова оцінок параметрів прихованого марківського ланцюга, що керує перемиканнями у рівнянні;

5. програмна реалізація рівняння авторегресії з марківськими перемиканнями.

Об'єктом дослідження є рівняння авторегресії з марковськими перемиканнями.

Предметом дослідження є властивості рівняння авторегресії з марковськими перемиканнями та побудова оцінок його параметрів.

При розв'язанні поставлених завдань використовувались наступні *методи дослідження*: лінійної алгебри, теорії імовірностей, математичної статистики, теорії випадкових процесів, математичного аналізу.

Наукова новизна отриманих результатів полягає у виведенні оцінок параметрів прихованого ланцюга Маркова та інших параметрів рівняння авторегресії. Також знайдено математичне сподівання, дисперсію та коваріаційну функцію стаціонарного розв'язку рівняння авторегресії з марківськими перемиканнями.

Практичне значення результатів полягає в тому, що вони можуть використовуватися при аналізі різноманітних фінансових показників та для покращення їх прогнозування (наприклад зміни ціни криптовалют).

Апробація результатів та публікації. Апробація результатів даної роботи була здійснена на XXI Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики» та була опублікована як стаття у збірнику з матеріалами конференції.

1 ОСНОВНІ ВІДОМОСТІ ПРО ПММ

У цьому розділі ми наведемо основні поняття прихованої марковської моделі та сформулюємо основні задачі, які розв'язуються в рамках цієї моделі.

1.1 Означення ПММ

Прихована марковська модель або скрочено ПММ (англ. hidden Markov model, HMM) — це статистична марковська модель, у якій система, що моделюється, розглядається як марковський процес із прихованими (неспостережуваними) станами. Основним представником простіших марковських моделей є ланцюги Маркова. У них стан є безпосередньо видимим спостерігачеві, і тому ймовірності переходу станів є єдиними параметрами. У прихованій марковській моделі стан не є видимим безпосередньо, але вихідні дані, як залежать від стану, видимими є. Кожен стан має ймовірнісний розподіл усіх можливих вихідних значень. Отже, послідовність символів, згенерована ПММ, дає якусь інформацію про послідовність станів. Прикметник «прихований» стосується послідовності станів, якою проходить модель, а не параметрів моделі; модель все одно називають «прихованою» марковською моделлю, навіть якщо ці параметри відомі точно [5].

Нехай $\{X_n\}_{n\geqslant 0}$ — послідовність випадкових величин, заданих на одному й тому ж ймовірнісному просторі і які приймають не більш ніж зліченну кількість значень.

Означення 1.1. Послідовність випадкових величин $\{X_n\}_{n\geqslant 0}$ називається ланцюгом Маркова, якщо $\forall n \geqslant 1$ і $\forall i_0, i_1, \ldots, i_n \in E$ має місце наступна рівність:

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_n = i_n \mid X_{n-1} = i_{n-1}).$$

Випадкову величину X_n інтерпретують як стан системи в момент часу n.

Означення 1.2. Ймовірність $p_{ij} = P(X_{n+1} = j \mid X_n = i)$ переходу зі стану i в стан j називають перехідною ймовірністю зі стану i в стан j. Матрицю $A = (p_{ij})_{i \in E, j \in E}$ називають матрицею перехідних ймовірностей ланцюга Маркова.

Матриця A є стохастичною, тобто

$$p_{ij} \geqslant 0 \quad \forall i, j \in E$$

та

$$\sum_{j \in E} p_{ij} = 1 \quad \forall i \in E$$

Означення 1.3. Нехай $\{X_n\}_{n \geqslant 0}$ — задана послідовність випадкових величин на скінченній множині станів $E = \{x_1, \ldots, x_N\}$. Нехай $\{Y_n\}_{n \geqslant 0}$ — задана послідовність випадкових величин на скінченній множині станів $F = \{y_1, \ldots, y_M\}$. Послідовність $\{(X_n, Y_n)\}_{n \geqslant 0}$ називається прихованою марковською моделлю, якщо виконуються наступні умови:

1) послідовність $\{(X_n, Y_n)\}_{n \geqslant 0}$ є ланцюгом Маркова на $E \times F$;

2) $\{X_n\}_{n \geqslant 0}$ є ланцюгом Маркова з початковим розподілом π і матрицею перехідних ймовірностей A;

3) $\forall n \geqslant 0$ випадкові величини $Y_0, Y_1, \ldots, Y_n$ є умовно незалежними при заданих $X_0, X_1, \ldots, X_n$, тобто

$$\forall n \geqslant 1 \ \forall y_0, \ldots, y_n \in F \text{ та } \forall x_0, \ldots, x_n \in E :$$
$$P(Y_0 = y_0, \ldots, Y_n = y_n \mid X_0 = x_0, \ldots, X_n = x_n) =$$
$$= \prod_{i=0}^{n} P(Y_i = y_i \mid X_0 = x_0, \ldots, X_n = x_n)$$

і до того ж

$$P(Y_i = y_i \mid X_0 = x_0, \dots, X_n = x_n) = P(Y_i = y_i \mid X_i = x_i)$$

З означення 1.3 випливає, що скінченновимірні розподіли $\{(X_n, Y_n)\}_{n \geqslant 0}$ мають такий вигляд:

$$\begin{gathered} P(Y_0 = y_0, \dots, Y_n = y_n, X_0 = x_0, \dots, X_n = x_n) = \\ = P(X_0 = x_0, \dots, X_n = x_n) \times \\ \times P(Y_0 = y_0, \dots, Y_n = y_n \mid X_0 = x_0, \dots, X_n = x_n) = P(X_0 = x_0) \times \\ \times P(X_1 = x_1 \mid X_0 = x_0) \cdot \ldots \cdot P(X_n = x_n \mid X_{n-1} = x_{n-1}) \times \\ \times \prod_{i=0}^{n} P(Y_i = y_i \mid X_i = x_i). \end{gathered}$$

Якщо ввести наступні позначення:

$$\begin{gathered} B_{xy} = P(Y_k = y \mid X_k = x), \\ A_{x_i x_j} = P(X_k = x_j \mid X_{k-1} = x_i), \\ \pi = (P(X_0 = x_1), \dots, P(X_0 = x_N)). \end{gathered}$$

то скінченновимірні розподіли матимуть вигляд:

$$\begin{gathered} P(Y_0 = y_0, \dots, Y_n = y_n, X_0 = x_0, \dots, X_n = x_n) = \\ = \pi_{x_0} \cdot \prod_{i=0}^{n-1} A_{x_i x_{i+1}} \cdot \prod_{j=0}^{n} B_{x_j y_j}. \end{gathered}$$

Трійку $\lambda = (\pi, A, B)$ використовують для позначення ПММ, тут π — вектор початкового розподілу ПММ, A — матриця перехідних ймовірностей ПММ, B — матриця, що пов'язує приховані стани спостережень із самими спостереженнями [6].

1.2 Три основні задачі ПММ

Для прихованих марковських моделей існує 3 основних задачі [7, 8]:

1) *задача оцінювання*: за заданою моделлю $\lambda = (\pi, A, B)$ і послідовністю спостережень $y = (y_1, \dots, y_T)$ визначити ймовірність $P_\lambda(Y = y)$;

2) *задача декодування*: за заданою моделлю $\lambda = (\pi, A, B)$ і послідовністю спостережень $y = (y_1, \dots, y_T)$ визначити оптимальну послідовність станів $x = (x_1, \dots, x_T)$ для прихованого ланцюга Маркова;

3) *задача навчання*: за $y = (y_1, \dots, y_T)$ — заданою послідовністю спостережень визначити модель $\lambda = (\pi, A, B)$, яка максимізує ймовірність $P_\lambda(Y = y)$ спостереження цієї послідовності.

Розв'язок задачі 1:

Нехай маємо модель $\lambda = (\pi, A, B)$ і послідовність спостережень $y = (y_0, \dots, y_T)$, $y_t \in 1, 2, \dots, M$, за якими хочемо визначити ймовірність $P_\lambda(Y = y)$, де $Y = (Y_0, \dots, Y_T)$. Позначимо $X = (X_0, \dots, X_T)$, $x = (x_0, \dots, x_T)$, $x_t \in 1, 2, ..., N$.

Оскільки

$$P_\lambda(Y = y) = \sum_x P_\lambda(Y = y, X = x),$$

то

$$P_\lambda(Y = y) = \sum_{x_0,\dots,x_T} \pi_{x_0} \prod_{i=0}^{T-1} A_{x_i x_{i+1}} \prod_{j=0}^{T} B_{x_j y_j}.$$

Пряме обчислення за цією формулою,взагалі кажучи, є неможливим, оскільки кількість добутків є приблизно $(T+1)N^{T+1}$, де T є як правило великим і $N \geqslant 2$.

Але однією з суттєвих переваг ПММ є те, що існує ефективний алгоритм - алгоритм прямого і зворотного ходу - для обчислення цієї ймовірності. Цей алгоритм вимагає лише $N^2(T+1)$ добутків.

Щоб ефективно обчислити $P_\lambda(Y=y)$, покладемо для $t=0,1,\ldots,T$ та $x_t=1,\ldots,N$

$$\alpha_t(x_t) = P_\lambda(Y_0=y_0,\ldots,Y_t=y_t,X_t=x_t),$$

$\alpha_t(x_t)$ — це сумісна ймовірність часткової послідовності спостережень до моменту часу t і стану $x_t \in 1,2,\ldots,N$ ланцюга Маркова в момент часу t.

Виявляється, що ці коефіцтєнти можуть бути обчислені рекуретно за допомогою так званого алгоритму прямого ходу, який вимагає лише $N^2(T+1)$ добутків.

В основі цього алгоритму лежить наступний ланцюжок співвідношень для рекурентного обчислення коефіцієнтів:

$$\forall x_0 \quad 1 \leqslant x_0 \leqslant N:$$
$$\alpha_0(x_0) = P_\lambda(Y_0=y_0,X_0=x_0) = P_\lambda(X_0=x_0)\cdot P_\lambda(Y_0=y_0 \mid X_0=x_0) =$$
$$= \mu_{x_0}\cdot B_{x_0}(y_0),$$
$$t+1=1,\ldots,T \quad x_{t+1}=1,\ldots,N:$$
$$\alpha_{t+1}(x_{t+1}) = P_\lambda(Y_0=y_0,\ldots,Y_{t+1}=y_{t+1},X_{t+1}=x_{t+1}) =$$
$$= \sum_{x_t=1}^{N}[P_\lambda(Y_0=y_0,\ldots,Y_t=y_t,X_t=x_t)\cdot P_\lambda(X_{t+1}=x_{t+1} \mid X_t=x_t)\times$$
$$\times P_\lambda(Y_{t+1}=y_{t+1} \mid X_{t+1}=x_{t+1})] = \sum_{x_t=1}^{N}\alpha_t(x_t)\cdot A_{x_t x_{t+1}}\cdot B_{x_{t+1}y_{t+1}}.$$

Якщо всі коефіцієнти прямого ходу обчислено, то:

$$P_\lambda(Y_0=y_0,\ldots,Y_T=y_T) = \sum_{x_T=1}^{N}\alpha_T(x_T),$$
$$\alpha_T(x_T) = P_\lambda(Y_0=y_0,\ldots,Y_T=y_T,X_T=x_T).$$

Цю ж ймовірність можна обчислити за допомогою так званого алгоритму зворотного ходу. Означимо умовні ймовірності послідовності спостережень від моменту часу $t+1$ до T при умові, що прихованим є стан x_t в момент часу t:

$$\beta_t(x_t) = P_\lambda(Y_{t+1} = y_{t+1}, \ldots, Y_T = y_T \mid X_t = x_t),$$
$$0 \leqslant t \leqslant T-1, \quad 1 \leqslant x_t \leqslant N$$

і покладемо

$$\beta_T(x_T) = 1,$$
$$1 \leqslant x_T \leqslant N.$$

Тоді для $t = 0, ..., T-1$ і $1 \leqslant x_t \leqslant N$ отримуємо

$$\beta_t(x_t) = P_\lambda(Y_{t+1} = y_{t+1}, \ldots, Y_T = y_T \mid X_t = x_t) =$$
$$= \sum_{x_{t+1}=1}^{N} [P_\lambda(X_{t+1} = x_{t+1} \mid X_t = x_t) \cdot P_\lambda(Y_{t+1} = y_{t+1} \mid X_{t+1} = x_{t+1}) \times$$
$$\times P_\lambda(Y_{t+2} = y_{t+2}, \ldots, Y_T = y_T \mid X_{t+1} = x_{t+1})] =$$
$$= \sum_{x_{t+1}=1}^{N} A_{x_t x_{t+1}} \cdot B_{x_{t+1} y_{t+1}} \cdot \beta_{t+1}(x_{t+1}).$$

Якщо всі коефіцієнти зворотного ходу обчислено, то:

$$P_\lambda(Y_0 = y_0, \ldots, Y_T = y_T) = \sum_{x_0=1}^{N} [P_\lambda(X_0 = x_0) \cdot P_\lambda(Y_0 = y_0, \mid X_0 = x_0) \times$$
$$\times P_\lambda(Y_1 = y_1, \ldots, Y_T = y_T \mid X_0 = x_0) = \sum_{x_0=1}^{N} \mu_{x_0} \cdot B_{x_0 y_0} \cdot \beta_0(x_0).$$

Алгоритм зворотного ходу, як і алгоритм прямого ходу, вимагає обчисленя $N^2(T+1)$ добутків.

Розв’язок задачі 2:

За заданою моделлю $\lambda = (\pi, A, B)$ і послідовністю спостережень $y = (y_0, \ldots, y_T)$ ми хочемо знайти послідовність станів $x = (x_0, \ldots, x_T)$ прихованого ланцюга, яка максимізує умовну ймовірність

$P_\lambda(X = x \mid Y = y)$. Це еквівалентно пошуку послідовності, яка максимізує ймовірність $P_\lambda(X = x, Y = y)$ сумісної появи ланцюжків. Алгоритм, який дозволяє ефективно розв'язати задачу декодування, називається алгоритмом Вітербі.

Алгоритм Вітербі знаходження найбільш ймовірного ланцюжка прихованих станів є наступним:

1. Покладаємо
$$\delta_0(x_0) = \mu_{x_0} \cdot B_{x_0 y_0}, \quad 1 \leqslant x_0 \leqslant N$$

2. Рекурентно обчислюємо для $t = 1, 2, \ldots, T$ і $1 \leqslant x_t \leqslant N$
$$\delta_t(x_t) = \max_{1 \leqslant x_{t-1} \leqslant N} [\delta_{t-1}(x_{t-1}) \cdot A_{x_{t-1} x_t}] \cdot B_{x_t y_t}, \quad 1 \leqslant x_0 \leqslant N,$$
$$\psi_t(x_t) = \arg\max_{1 \leqslant x_{t-1} \leqslant N} [\delta_{t-1}(x_{t-1}) \cdot A_{x_{t-1} x_t}].$$

3. Після знаходження масивів $\delta_t(x_t)$, $\psi_t(x_t)$ покладаємо
$$\delta^* = \max_{1 \leqslant x_T \leqslant N} [\delta_T(x_T)],$$
$$\psi^* = \arg\max_{1 \leqslant x_{t-1} \leqslant N} [\delta_T(x_T)].$$

4. Врешті, ми отримуємо оптимальну послідовність станів (справа наліво):
$$x_T^* = \psi^*,$$
$$x_t^* = \psi_{t+1}(x_{t+1}^*),$$
$$t = T-1, T-2, \ldots, 0$$

Зауваження. Оптимальний ланцюжок прихованих станів не обов'язково є єдиним.

Розв'язок задачі 3:

Ми хочемо знайти параметри моделі, які найкраще пояснюють отримані спостереження. Ми вважаємо, що кількість N станів прихованого ланцюга і кількість M можливих станів спостережень є відомими, а елементи вектора початкового розподілу π, матриць A, B треба визначити. Задачу можна сформулювати так:

$$\lambda^* = \arg\max_{\lambda} P_\lambda(Y = y)$$

Введемо наступні позначення:

$$\gamma_t(i,j) = P_\lambda(X_t = i, X_{t+1} = j | Y_0^T = y_0^T),$$
$$\gamma_t(i) = P_\lambda(X_t = i | Y_0^T = y_0^T) = \sum_{j=1}^{N} \gamma_t(i,j),$$

$\gamma_t(i{,}j)$ — це умовна ймовірність того, що в момент часу t станом був i та з нього відбувся перехід в стан j в наступний момент часу при заданих $Y_0, \ldots, Y_T$. Зауважимо, що цю ймовірність можна виразити через коефіцієнти прямого і зворотнього ходу:

$$\gamma_t(i,j) = \frac{P_\lambda(X_t = i, X_{t+1} = j, Y_0^T = y_0^T)}{P_\lambda(Y_0^T = y_0^T)} = \frac{\alpha_t(i) \cdot A_{ij} \cdot Bjy_{t+1}\beta_{t+1}(j)}{P_\lambda(Y_0^T = y_0^T)},$$
$$\gamma_t(i) = \frac{P_\lambda(X_t = i, Y_0^T = y_0^T)}{P_\lambda(Y_0^T = y_0^T)} = \frac{\alpha_t(i) \cdot \beta_t(j)}{P_\lambda(Y_0^T = y_0^T)}$$

За допомогою цих величин, оптимальну модель знаходять за ітераційним алгоритмом Баума-Велша [5]. Нехай $\lambda = (\pi, A, B)$ - початкова (апріорна) модель.

Алгоритм Баума-Велша є наступним:

1. Обчислюємо для поточної моделі λ величини

$$\alpha_t(i), \quad ,\beta_t(i), \quad \gamma_t(i), \quad t = 0, \ldots, T, \quad 1 \leqslant i \leqslant N,$$
$$\gamma_t(i,j), \quad t = 0, \ldots, T-1, \quad 1 \leqslant i,j \leqslant N.$$

2. Будуємо модель λ^* з параметрами

(а)

$$\pi_i^* = \gamma_0(i), \quad i = 1, \ldots, N$$

(б) Для $i = 1, \ldots, N, j = 1, \ldots, N$

$$A_{ij}^* = \frac{\sum_{t=0}^{T-1} \gamma_t(i,j)}{\sum_{t=0}^{T-1} \gamma_t(i)}$$

(в) Для $j = 1, \ldots, N, k = 1, \ldots, M$

$$B_{jk}^* = \frac{\sum_{0 \leqslant t \leqslant T, Y_t = k} \gamma_t(i)}{\sum_{t=0}^{T-1} \gamma_t(i)}$$

3. За переоціненою моделлю обчислюємо ймовірність $P_\lambda^*(Y_0^T = y_0^T)$. Якщо ймовірність $P_\lambda^*(Y_0^T = y_0^T)$ зросла порівняно з $P_\lambda(Y_0^T = y_0^T)$, то λ^* вважаємо поточною і повертаємося до п.1.

Зауваження. На практиці ітераційний алгоритм переоцінювання параметрів моделі зупиняємо, якщо $P_\lambda(Y_0^T = y_0^T)$ зростає менше, ніж деяке наперед задане порогове значення ε. Або ж можна наперед задати кількість ітерацій.

Зауваження. Якщо про модель немає якихось припущень, то вибираємо в якості початкових параметрів значення, що відповідають рівномірному розподілу на відповідних множинах: $\mu_i \sim \dfrac{1}{N}$, $A_{ij} \sim \dfrac{1}{N}$, $B_{jk} \sim \dfrac{1}{M}$. Рандомізація параметрів убезпечує від застрягання в точці локального максимуму. Щоб наблизитися до глобального максимуму, використовують також багатократну реалізацію алгоритму Баума-Велша з різних рандомізованих початкових точок.

Зауваження. Шкалювання. Всі три розв'язки основних задач для

ПММ потребують обчислень добутків ймовірностей. Ці добутки стрімко прямують до нуля при збільшенні часового параметра T. Тому довільна спроба реалізувати запропоновані алгоритми при великому T призведе до втрати значущості. Одним з рішень цієї проблеми в алгоритмі прямого і зворотного ходу є шкалювання, чи нормування відповідних добутків, а саме

$$\hat{\alpha}_t(i) = \frac{\alpha_t(i)}{\sum\limits_{j=0}^{N} \beta_t(j)}$$

$$\hat{\beta}_t(i) = \frac{\beta_t(i)}{\sum\limits_{j=0}^{N} \beta_t(j)}$$

В алгоритмі Вітербі для вирішення проблеми можна перейти до логарифмів. Оскільки $logmax = maxlog$, то

$$\log \delta_t(j) = \log \underset{i}{max}[\delta_{t-1}(i) \cdot A_{ij} \cdot B_{jy_t}] = \underset{i}{max}[\log \delta_{t-1}(i) + \log A_{ij} + \log B_{jy_t}]$$

Висновки до розділу 1

У цьому розділі ми дали основні означення ланцюга Маркова, ПММ, описали задачі та навели їх розв'язок, в рамках цієї моделі.

2 ДОСЛІДЖЕННЯ ВЛАСТИВОСТЕЙ ТА ОЦІНЮВАННЯ ПАРАМЕТРІВ АВТОРЕГРЕСІЙНОГО РІВНЯННЯ З МАРКІВСЬКИМИ ПЕРЕМИКАННЯМИ

У цьому розділі ми встановимо деякі властивості рівняння авторегресії з марківськими перемиканнями пов'язані з існуванням стаціонарного розв'язку цього рівняння та побудуємо оцінки невідомих параметрів цього рівняння.

2.1 Модель

Основу нашої моделі складає ланцюг Маркова — $\{X_n\}_{n\geqslant 1}$ на $E = \{1;2\}$ з початковим розподілом $\pi = (\pi_1, \pi_2)$ і матрицею перехідних ймовірностей

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

Спостережуваними є величини $Y = \{Y_n\}_{n\geqslant 0}$, початкова з них має стандартний нормальний розподіл $Y_0 \sim N(0{,}1)$, а інші пов'язані з попередніми формулою:

$$Y_{n+1} = \mu_{X_{n+1}} + a_{X_{n+1}} \cdot Y_n + \varepsilon_t^{X_{n+1}} = \tag{2.1}$$
$$= \begin{cases} \mu_1 + a_1 \cdot Y_n + \varepsilon_t^1, & X_{n+1} = 1, \\ \mu_2 + a_2 \cdot Y_n + \varepsilon_t^2, & X_{n+1} = 2, \end{cases}$$

де $\mu_1, \mu_2 = \text{const}$, $a_1, a_2 = \text{const}$, такі що $|a_i| < 1$, $\{X_n\}_{n\geqslant 1}$ — ланцюг Маркова, $\{\varepsilon_n^1\}_{n\geqslant 1}$, $\{\varepsilon_n^2\}_{n\geqslant 1}$ — незалежні від $(X_n)_{n\geqslant 1}$ та між собою послідовності незалежних однаково розподілених випадкових величин $\varepsilon_n^1 \sim N(0{,}\sigma_1^2)$; $\varepsilon_n^2 \sim N(0{,}\sigma_2^2)$, При фіксованих $Y_n = y_n$, $X_{n+1} = x$ умовний розподіл Y_{n+1} має вигляд:

$$\begin{aligned} p(y \mid x, y_n) = p_{Y_{n+1}|X_{n+1},Y_n}(y \mid x, y_n) = \\ = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(y-(a_x y_n + \mu_x))^2}{2\sigma_x^2}\right). \end{aligned} \tag{2.2}$$

З (2.2) бачимо, що з означення 1.3 ПММ умова 3) не виконується: кожне спостереження y залежить не лише від стану ланцюга Маркова в даний момент часу, а й від значення y у попередній момент часу — це є особливістю авторегресійних прихованих марківських моделей.

Тоді, нехай $\lambda = (\pi, A, a_1, a_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$ — фіксована модель.

Надалі позначатимемо $x = (x_1, \ldots, x_T)$.

Якщо вважати, що система функціонує в стаціонарному режимі, то вважаємо тоді, що початковий розподіл є стаціонарним. Стацінарний розподіл, це розподіл, що задовольняє наступні умови:

$$\begin{aligned} \pi A = \pi, \\ \sum_{i,j=1} \pi_i A_{ij} = \pi_j. \end{aligned} \tag{2.3}$$

Якщо врахувати, що матриця A має наступний вигляд:

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

то роз'язок рівняння (2.3) набуде вигляду

$$\begin{cases} \pi_1 = \dfrac{A_{21}}{A_{12} + A_{21}}, \\ \pi_2 = \dfrac{A_{12}}{A_{12} + A_{21}}. \end{cases} \tag{2.4}$$

Тоді для всіх t виконується $P(X_t = i) = \pi_i$.

Зауважимо, ще що

$$P(X_{t-1}=i \mid X_t=j)=\frac{P(X_{t-1}=i, X_t=j)}{P(X_t=j)}=\frac{P(X_{t-1}=i, X_t=j)}{\pi_j}=$$
$$=\frac{P(X_{t-1}=i)\cdot P(X_t=j \mid X_{t-1}=i)}{\pi_j}=\frac{\pi_i\cdot A_{ij}}{\pi_j}=\frac{A_{ji}\cdot A_{ij}}{A_{ij}}=A_{ji}.$$

2.2 Властивості рівняння авторегресії з марківськими перемиканнями

Маємо рівняння авторегресії з марківськими перемиканнями (2.1). При розгляді такого рівняння виникає питання чи існує його стаціонарний розв'язок.

Стаціонарність в широкому сенсі означає, шо математичне сподівання процесу не залежить від часу, тобто для всіх значень t виконується, що $MY_t = const$ та що коваріаційна функція залежить лише від різниці аргументів: $\mathrm{cov}(Y_t, Y_s) = K(t,s) = r(t-s)$.

Щоб перевірити необхідну умову існування стаціонарного розв'язку, перевіримо, що існує границя математичного сподівання, яка від n не залежить при n, що прямує до нескінченності: $\lim\limits_{n\to\infty} MY_n = \mathrm{const}$ та існує границя коваріаційної функції, яка залежить лише від різниці аргументів при n, що прямує до нескінченності: $\lim\limits_{n\to\infty} \mathrm{cov}(Y_n, Y_{n+s}) = r(s)$.

Теорема 2.1. *Якщо* $|a_i| < 1$*, то* $\exists \lim\limits_{t\to\infty} MY_t = \pi\cdot(I - a_d\cdot A)^{-1}\cdot\mu$*, де*

$$\pi=\begin{pmatrix}\pi_1 & \pi_2\end{pmatrix},\quad \mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\quad a_d=\begin{pmatrix}a_1 & 0\\ 0 & a_2\end{pmatrix},\quad A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}.$$

Доведення. Запишемо математичне сподівання для Y_t:

$$MY_t = P(X_t=1)\cdot M(Y_t \mid X_t=1) + P(X_t=2)\cdot M(Y_t \mid X_t=2)=$$
$$= \pi_1\cdot m_1(t) + \pi_2\cdot m_2(t). \tag{2.5}$$

В попередній формулі ми використовуємо позначення $m_1(t)$ і $m_2(t)$ для умовних математичних сподівань $M(Y_t \mid X_t = 1)$ та $M(Y_t \mid X_t = 2)$ відповідно. Використовуючи (2.1), обчислимо ці величини. Для $m_1(t)$ отримаємо:

$$\begin{gathered} m_1(t) = M(Y_t \mid X_t = 1) = M(\mu_{X_t} + a_{X_t} \cdot Y_{t-1} + \varepsilon_t^{X_t} \mid X_t = 1) = \\ = M(\mu_{X_t} \mid X_t = 1) + M(a_{X_t} \cdot Y_{t-1} \mid X_t = 1) + M(\varepsilon_t^{X_t} \mid X_t = 1) = \\ = \mu_1 + a_1 \cdot M(Y_{t-1} \mid X_t = 1) + 0 = \\ = \mu_1 + a_1 \cdot (P(X_{t-1} = 1 \mid X_t = 1) \cdot M(Y_{t-1} = 1 \mid X_{t-1} = 1) + \\ + P(X_{t-1} = 2 \mid X_t = 1) \cdot M(Y_{t-1} = 1 \mid X_{t-1} = 2)) = \\ = \mu_1 + a_1 \cdot (A_{11} \cdot m_1(t-1) + A_{12} \cdot m_2(t-1)). \end{gathered} \tag{2.6}$$

Аналогічно отримуємо, що $m_2(t)$ дорівнює:

$$m_2(t) = \mu_2 + a_2 \cdot (A_{21} \cdot m_1(t-1) + A_{22} \cdot m_2(t-1)), \tag{2.7}$$

Перейдемо до матричної форми запису отриманого рекурентного співвідношення:

$$\begin{gathered} m(t) = \begin{pmatrix} m_1(t) \\ m_2(t) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} m_1(t-1) \\ m_2(t-1) \end{pmatrix} = \\ = \mu + a_d \cdot A \cdot m(t-1). \end{gathered}$$

Використовуючи вищенаведене рекурентне співвідношення матимемо:

$$\begin{gathered} m(t) = \mu + a_d \cdot A \cdot m(t-1) = \mu + a_d \cdot A \cdot (\mu + a_d \cdot A \cdot m(t-2)) = \\ = (I + a_d \cdot A) \cdot \mu + (a_d \cdot A)^2 \cdot m(t-2) = \\ = (I + a_d \cdot A) \cdot \mu + (a_d \cdot A)^2 \cdot (\mu + a_d \cdot A \cdot m(t-3)) \\ = (I + a_d \cdot A + (a_d \cdot A)^2) \cdot \mu + (a_d \cdot A)^3 \cdot m(t-3) = \ldots = \end{gathered}$$

$$= (I + a_d \cdot A + (a_d \cdot A)^2 + ... + (a_d \cdot A)^{t-2}) \cdot \mu + (a_d \cdot A)^{t-1} \cdot m(1)$$

Врахувавши, що $m_i(1) = M(Y_1 \mid X_1 = i) = \mu_i$ матимемо:

$$m(t) = (I + a_d \cdot A + (a_d \cdot A)^2 + ... + (a_d \cdot A)^{t-2}) \cdot \mu + (a_d \cdot A)^{t-1} \cdot \mu =$$
$$= (I + a_d \cdot A + (a_d \cdot A)^2 + ... + (a_d \cdot A)^{t-2}(a_d \cdot A)^{t-1}) \cdot \mu$$

Оскільки $|a_i| < 1$ та матриця A є стохастична, то $||a_d \cdot A|| < 1$, тоді матриця $(I - a_d \cdot A)$ є оборотною та

$$m(t) = (I - a_d \cdot A)^{-1} \cdot (I - (a_d \cdot A)^t) \cdot \mu$$

Звідси випливає існування границі матриці умовних математичних сподівань $M(Y_t \mid X_t)$

$$\lim_{t\to\infty} m(t) = (I - a_d \cdot A)^{-1} \cdot (I - 0) \cdot \mu = (I - a_d \cdot A)^{-1} \cdot I \cdot \mu =$$
$$= (I - a_d \cdot A)^{-1} \cdot \mu = m$$

Звернемося з даним результатом до (2.5):

$$\lim_{t\to\infty} MY_t = \lim_{t\to\infty} (\pi_1 \cdot m_1(t) + \pi_2 \cdot m_2(t)) = \lim_{t\to\infty} \left(\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \cdot \begin{pmatrix} m_1(t) \\ m_2(t) \end{pmatrix} \right) =$$
$$= \lim_{t\to\infty} \pi \cdot m(t) = \pi \cdot (I - a_d \cdot A)^{-1} \cdot \mu.$$

$\square$

Наступною нашою задачею буде встановлення існування границі $\mathrm{cov}(Y_n, Y_{n+s})$, але спочатку переконаємося, що існує границя другого моменту $M(Y_t^2)$ при t прямує до нескінченності. Матиме місце наступна теорема.

Теорема 2.2. *Якщо* $|a_i| < 1$, *то існує границя* $\lim\limits_{t\to\infty} M(Y_t^2)$ *і*

$$\lim_{t\to\infty} M(Y_t^2) = [\pi \cdot (I - a_d \cdot A)^{-1} \cdot (d + 2b \cdot A \cdot m)], \text{ де}$$

$$\pi = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}, \quad d = \begin{pmatrix} \mu_1^2 + \sigma_1^2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix}, \quad a_d = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$b = \begin{pmatrix} \mu_1 \cdot a_1 & 0 \\ 0 & \mu_2 \cdot a_2 \end{pmatrix}, \quad m = (I - a_d \cdot A)^{-1} \cdot \mu, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

Доведення. Запишемо математичне сподівання для Y_t^2:

$$M(Y_t^2) = P(X_t = 1) \cdot M(Y_t^2 \mid X_t = 1) + P(X_t = 2) \cdot M(Y_t^2 \mid X_t = 2) = \\ = \pi_1 \cdot s_1(t) + \pi_2 \cdot s_2(t), \tag{2.8}$$

В попередній формулі ми використовуємо позначення $s_1(t)$ і $s_2(t)$ для умовних математичних сподівань $M(Y_t^2 \mid X_t = 1)$ та $M(Y_t^2 \mid X_t = 2)$ відповідно. Використовуючи (2.1), обчислимо $s_1(t)$ і $s_2(t)$. Для $s_1(t)$ отримаємо:

$$s_1(t) = M(Y_t^2 \mid X_t = 1) = M((\mu_{X_t} + a_{X_t} \cdot Y_{t-1} + \varepsilon_t^{X_t})^2 \mid X_t = 1) = \\ = M(\mu_{X_t}^2 + a_{X_t}^2 \cdot Y_{t-1}^2 + (\varepsilon_t^{X_t})^2 + 2\mu_{X_t} \cdot a_{X_t} \cdot Y_{t-1} \mid X_t = 1) + \\ + M(2\mu_{X_t} \cdot \varepsilon_t^{X_t} + 2a_{X_t} \cdot Y_{t-1} \cdot \varepsilon_t^{X_t}) \mid X_t = 1) = \\ = M(\mu_{X_t}^2 \mid X_t = 1) + M(a_{X_t}^2 \cdot Y_{t-1}^2 \mid X_t = 1) + M((\varepsilon_t^{X_t})^2 \mid X_t = 1) + \\ + 2M(\mu_{X_t} \cdot a_{X_t} \cdot Y_{t-1} \mid X_t = 1) + 2M(\mu_{X_t} \cdot \varepsilon_t^{X_t} \mid X_t = 1) + \\ + 2M(a_{X_t} \cdot Y_{t-1} \cdot \varepsilon_t^{X_t} \mid X_t = 1) = \mu_1^2 + a_1^2 \cdot M(Y_{t-1}^2 \mid X_t = 1) + \\ + M((\varepsilon_t^1)^2) + 2\mu_1 \cdot a_1 \cdot M(Y_{t-1} \mid X_t = 1) + 2\mu_1 \cdot M(\varepsilon_t^1) + \\ + 2a_1 \cdot M(Y_{t-1} \mid X_t = 1) \cdot M(\varepsilon_t^1)$$

Оскільки

$$M(\varepsilon_t^1) = 0,$$

$$M((\varepsilon_t^1)^2) = D\varepsilon_t^1 + (M\varepsilon_t^1)^2 = \sigma_1^2 - 0^2 = \sigma_1^2$$

та

$$\begin{gathered} M(Y_{t-1}^2 \mid X_t = 1) = P(X_{t-1} = 1 \mid X_t = 1) \cdot M(Y_{t-1}^2 = 1 \mid X_{t-1} = 1) + \\ + P(X_{t-1} = 2 \mid X_t = 1) \cdot M(Y_{t-1}^2 = 1 \mid X_{t-1} = 2) = \\ = A_{11} \cdot s_1(t-1) + A_{12} \cdot s_2(t-1) \end{gathered}$$

врахувавши (2.6) і (2.7), отримаємо таку рівність:

$$\begin{gathered} s_1(t) = \mu_1^2 + a_1^2 \cdot (A_{11} \cdot s_1(t-1) + A_{12} \cdot s_2(t-1)) + \sigma_1^2 + \\ + 2\mu_1 \cdot a_1 \cdot (A_{11} \cdot m_1(t-1) + A_{12} \cdot m_2(t-1)) = \\ = (\mu_1^2 + \sigma_1^2) + a_1^2 \cdot (A_{11} \cdot s_1(t-1) + A_{12} \cdot s_2(t-1)) + \\ + 2\mu_1 \cdot a_1 \cdot (A_{11} \cdot m_1(t-1) + A_{12} \cdot m_2(t-1)). \end{gathered}$$

Аналогічно отримуємо $s_2(t)$:

$$\begin{gathered} s_2(t) = (\mu_2^2 + \sigma_2^2) + a_2^2 \cdot (A_{21} \cdot s_1(t-1) + A_{22} \cdot s_2(t-1)) + \\ + 2\mu_2 \cdot a_2 \cdot (A_{21} \cdot m_1(t-1) + A_{22} \cdot m_2(t-1)). \end{gathered}$$

Перейдемо до матричної форми запису отриманого рекурентного співвідношення:

$$s(t) = \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = \begin{pmatrix} \mu_1^2 + \sigma_1^2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix} + \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}^2 \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \times$$

$$\times \begin{pmatrix} s_1(t-1) \\ s_2(t-1) \end{pmatrix} + 2 \begin{pmatrix} \mu_1 \cdot a_1 & 0 \\ 0 & \mu_2 \cdot a_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} m_1(t-1) \\ m_2(t-1) \end{pmatrix} =$$

$$= d + a_d^2 \cdot A \cdot s(t-1) + 2b_d \cdot A \cdot m(t-1)$$

Використовуючи вищенаведене рекурентне співвідношення матимемо:

$$s(t) = d + a_d^2 \cdot A \cdot s(t-1) + 2b_d \cdot A \cdot m(t-1) = d+$$
$$+a_d^2 \cdot A \cdot (d + a_d^2 \cdot A \cdot s(t-2) + 2b_d \cdot A \cdot m(t-2))+$$
$$+2b_d \cdot A \cdot m(t-1) = d + a_d^2 \cdot A \cdot d + (a_d^2 \cdot A)^2 \cdot s(t-2)+$$
$$+2a_d^2 \cdot A \cdot b_d \cdot A \cdot m(t-2) + 2b_d \cdot A \cdot m(t-1) =$$
$$= (I + a_d^2 \cdot A) \cdot d + 2b_d \cdot A \cdot (m(t-1) + a_d^2 \cdot A \cdot m(t-2))+$$
$$+(a_d^2 \cdot A)^2 \cdot s(t-2) = ... = (I + a_d^2 \cdot A + (a_d^2 \cdot A)^2 + ...+$$
$$+(a_d^2 \cdot A)^{t-1}) \cdot d + 2b_d \cdot A \cdot (m(t-1) + a_d^2 \cdot A \cdot m(t-2)+$$
$$+(a_d^2 \cdot A)^2 \cdot m(t-3) + ... + (a_d^2 \cdot A)^{t-1} \cdot m(0)) + (a_d^2 \cdot A)^t \cdot s(0)$$

Врахувавши, що $Y_0 \sim N(0{,}1)$ і не залежить від ланцюга Маркова $\{X_t\}_{t \geqslant 0}$

$$s_i(0) = M(Y_0^2 \mid X_0 = i) = MY_0^2 = DY_0 + (MY_0)^2 = 1 + 0^2 = 1,$$
$$m_i(0) = M(Y_0 \mid X_0 = i) = MY_0 = 0,$$

матимемо:

$$s(t) = (I + a_d^2 \cdot A + (a_d^2 \cdot A)^2 + ...+$$
$$+(a_d^2 \cdot A)^{t-1}) \cdot d + 2b \cdot A \cdot (m(t-1) + a_d^2 \cdot A \cdot m(t-2)+$$
$$+(a_d^2 \cdot A)^2 \cdot m(t-3) + ... + (a_d^2 \cdot A)^{t-2} \cdot m(1)) + (a_d^2 \cdot A)^t.$$

Оскільки $|a_i| < 1$ та матриця A є стохастична, то $||a_d^2 \cdot A|| < 1$, тоді матриця $(I - a_d^2 \cdot A)$ є оборотною та

$$(I + a_d^2 \cdot A + (a_d^2 \cdot A)^2 + ... + (a_d^2 \cdot A)^{t-1}) \cdot d =$$
$$= (I - a_d^2 \cdot A)^{-1} \cdot (I - (a_d^2 \cdot A)^t) \cdot d.$$

Застосуємо теорему Тьопліца до суми $\sum_{k=1}^{t-1}(a_d^2 \cdot A)^{t-1-k} \cdot m(k)$. Зауважимо, що:

$$\lim_{t\to\infty} m(t) = m,$$

$$\lim_{t\to\infty}(a_d^2 \cdot A)^t = 0,$$

$$\lim_{t\to\infty}(I - a_d^2 \cdot A) \cdot \sum_{k=0}^{t-1}(a_d^2 \cdot A)^k = I.$$

Отже за теоремою Тьопліца

$$\lim_{t\to\infty}(I - a_d^2 \cdot A) \cdot \sum_{n=1}^{t-1}(a_d^2 \cdot A)^{t-1-n} \cdot m(n) = m,$$

тобто:

$$\lim_{t\to\infty} \sum_{n=1}^{t-1}(a_d^2 \cdot A)^{t-1-n} \cdot m(n) = (I - a_d^2 \cdot A)^{-1} \cdot m.$$

Звідси випливає

$$\lim_{t\to\infty} s(t) = (I - a_d \cdot A)^{-1} \cdot (I - 0) \cdot d + 2b_d \cdot A \cdot (I - a_d^2 \cdot A)^{-1} \cdot m =$$
$$= (I - a_d \cdot A)^{-1} \cdot d + 2b_d \cdot A \cdot (I - a_d^2 \cdot A)^{-1} \cdot m =$$
$$= (I - a_d \cdot A)^{-1} \cdot (d + 2b_d \cdot A \cdot m) = s.$$

Звернемося з даним результатом до (2.8):

$$\lim_{t\to\infty} MY_t^2 = \lim_{t\to\infty}(\pi_1 \cdot s_1(t) + \pi_2 \cdot s_2(t)) = \lim_{t\to\infty}\left(\begin{pmatrix}\pi_1 & \pi_2\end{pmatrix} \cdot \begin{pmatrix} s_1(t) \\ s_2(t)\end{pmatrix}\right) =$$
$$= \lim_{t\to\infty}(\pi \cdot s(t)) = \pi \cdot ((I - a_d \cdot A)^{-1} \cdot (d + 2b_d \cdot A \cdot m)).$$

□

Тепер можемо знайти границю коваріаційної функції.

Теорема 2.3. *Якщо* $|a_i| < 1$, *то існує границя* $\lim\limits_{t\to\infty} \mathrm{cov}(Y_t,Y_{t+k}) = [\pi\cdot((I-a_d)^{-1}\cdot(I-a_d^k)\mu_d\cdot A^k\cdot m+(a_d\cdot A)^k\cdot s)-(\pi\cdot(I-a_d\cdot A)^{-1}\cdot\mu)^2]$, *де*

$$\pi = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}, \quad a_d = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \quad \mu_d = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$m = (I - a_d\cdot A)^{-1}\cdot\mu, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

$$s = (I - a_d\cdot A)^{-1}\cdot d + 2b_d\cdot A\cdot(I - a_d^2\cdot A)\cdot m, \quad d = \begin{pmatrix} \mu_1^2+\sigma_1^2 \\ \mu_2^2+\sigma_2^2 \end{pmatrix},$$

$$b_d = \begin{pmatrix} \mu_1\cdot a_1 & 0 \\ 0 & \mu_2\cdot a_2 \end{pmatrix}.$$

Доведення. Запишемо математичне сподівання для $Y_t\cdot Y_{t+k}$:

$$M(Y_t\cdot Y_{t+k}) = P(X_{t+k}=1)\cdot M(Y_t\cdot Y_{t+k}\mid X_{t+k}=1) + P(X_{t+k}=2)\times \\ \times M(Y_t\cdot Y_{t+k}\mid X_{t+k}=2) = \pi_1\cdot r_1^k(t) + \pi_2\cdot r_2^k(t), \tag{2.9}$$

В попередній формулі ми використовуємо позначення $r_1^k(t)$ і $r_2^k(t)$ для умовних математичних сподівань $M(Y_t\cdot Y_{t+k}\mid X_{t+k}=1)$ та $M(Y_t\cdot Y_{t+k}\mid X_{t+k}=2)$ відповідно. Використовуючи (2.1), обчислимо $r_1^k(t)$ і $r_2^k(t)$. Для $r_1^k(t)$ отримаємо:

$$r_1^k(t) = M(Y_t\cdot Y_{t+k}\mid X_{t+k}=1) = \\ = M(Y_t\cdot(\mu_{X_t+k} + a_{X_t+k}\cdot Y_{t+k-1} + \varepsilon_{t+k}^{X_t+k})\mid X_{t+k}=1) = \\ = M(Y_t\cdot\mu_{X_t+k} + Y_t\cdot a_{X_t+k}\cdot Y_{t+k-1} + Y_t\cdot\varepsilon_{t+k}^{X_t+k}\mid X_{t+k}=1) = \\ = M(Y_t\cdot\mu_{X_t+k}\mid X_{t+k}=1) + M(Y_t\cdot a_{X_t+k}\cdot Y_{t+k-1}\mid X_{t+k}=1) + \\ + M(Y_t\cdot\varepsilon_{t+k}^{X_t+k}\mid X_{t+k}=1) = \mu_1\cdot M(Y_t\mid X_{t+k}=1) +$$

$$+a_1 \cdot M(Y_t \cdot Y_{t+k-1} \mid X_{t+k} = 1) + M(Y_t \mid X_{t+k} = 1) \times M(\varepsilon^1_{t+k})$$

Вводимо позначення:

$$m_1^0(t) = m_1(t);$$
$$m_1^k(t) = A_{11} \cdot m_1^{k-1}(t) + A_{12} \cdot m_2^{k-1}(t);$$
$$m_2^k(t) = A_{21} \cdot m_1^{k-1}(t) + A_{22} \cdot m_2^{k-1}(t);$$
$$m^k(t) = A \cdot m^{k-1}(t) = \dots = A^k \cdot m^0(t) = A^k \cdot m(t).$$

та можемо використати

$$M(Y_t \mid X_{t+k} = 1) = m_1^k(t),$$
$$M(Y_t \cdot Y_{t+k-1} \mid X_{t+k} = 1) = A_{11} \cdot r_1^{k-1}(t) + A_{12} \cdot r_2^{k-1}(t),$$
$$M(\varepsilon^1_{t+k}) = 0,$$

тоді отримаємо таку рівність:

$$r_1^k(t) = \mu_1 \cdot m_1^k(t) + a_1 \cdot (A_{11} \cdot r_1^{k-1}(t) + A_{12} \cdot r_2^{k-1}(t)).$$

Аналогічно отримуємо $r_2^k(t)$:

$$r_2^k(t) = \mu_2 \cdot m_2^k(t) + a_2 \cdot (A_{21} \cdot r_1^{k-1}(t) + A_{22} \cdot r_2^{k-1}(t)).$$

Перейдемо до матричної форми запису отриманого рекурентного співвідношення:

$$r^k(t) = \begin{pmatrix} r_1^k(t) \\ r_2^k(t) \end{pmatrix} = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^k \cdot \begin{pmatrix} m_1(t) \\ m_2(t) \end{pmatrix} +$$
$$+ \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} \cdot \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \cdot \begin{pmatrix} r_1^{k-1}(t) \\ r_2^{k-1}(t) \end{pmatrix} =$$

$$= \mu_d \cdot A^k \cdot m(t) + a_d \cdot A \cdot r^{k-1}(t).$$

Використовуючи вищенаведене рекурентне співвідношення, матимемо:

$$r^k(t) = \mu_d \cdot A^k \cdot m(t) + a_d \cdot A \cdot r^{k-1}(t) = \mu_d \cdot A^k \cdot m(t)+$$
$$+a_d \cdot A \cdot (\mu_d \cdot A^{k-1} \cdot m(t) + a_d \cdot A \cdot r^{k-2}(t)) = \mu_d \cdot A^k \cdot m(t)+$$
$$+a_d \cdot \mu_d \cdot A^k \cdot m(t) + (a_d \cdot A)^2 \cdot r^{k-2}(t) = (I + a_d) \cdot \mu_d \cdot A^k \cdot m(t)+$$
$$+(a_d \cdot A)^2 \cdot r^{k-2}(t) = (I + a_d) \cdot \mu_d \cdot A^k \cdot m(t) + (a_d \cdot A)^2 \times$$
$$\times(\mu_d \cdot A^{k-2} \cdot m(t) + a_d \cdot A \cdot r^{k-3}(t)) = (I + a_d) \cdot \mu_d \cdot A^k \cdot m(t)+$$
$$+a_d^2 \cdot \mu_d \cdot A^k \cdot m(t) + (a_d \cdot A)^3 \cdot r^{k-3}(t) = (I + a_d + a_d^2) \cdot \mu_d \cdot A^k \cdot m(t)+$$
$$+(a_d \cdot A)^3 \cdot r^{k-3}(t) = ... = (I + a_d + a_d^2 + a_d^{k-1}) \cdot \mu_d \cdot A^k \cdot m(t)+$$
$$+(a_d \cdot A)^k \cdot r^0(t)$$

Врахувавши, що $r^0(t) = M(Y_t \cdot Y_{t+0} \mid X_{t+0} = 1) = M(Y_t^2 \mid X_t = 1) = s(t)$ матимемо:

$$r^k(t) = (I + a_d + a_d^2 + a_d^{k-1}) \cdot \mu_d \cdot A^k \cdot m(t)+$$
$$+(a_d \cdot A)^k \cdot s(t)$$

Оскільки $|a_i| < 1$ та матриця A є стохастична, то $||a_d \cdot A|| < 1$, тоді матриця $(I - a_d)$ є оборотною та

$$r^k(t)) = (I - a_d)^{-1} \cdot (I - a_d^k) \cdot \mu_d \cdot A^k \cdot m(t) + (a_d \cdot A)^k \cdot s(t)$$

Звідси випливає

$$\lim_{t\to\infty} r^k(t) = (I - a_d)^{-1} \cdot (I - a_d^k) \cdot \mu_d \cdot A^k \cdot m + (a_d \cdot A)^k \cdot s$$

Звернемося з даним результатом до (2.9):

$$\lim_{t\to\infty} M(Y_t \cdot Y_{t+k}) = \lim_{t\to\infty} (\pi_1 \cdot r_1^k(t) + \pi_2 \cdot r_2^k(t)) =$$

$$= \lim_{t\to\infty}\left(\begin{pmatrix}\pi_1 & \pi_2\end{pmatrix}\cdot\begin{pmatrix}r_1^k(t)\\ r_2^k(t)\end{pmatrix}\right) = \lim_{t\to\infty}(\pi\cdot r^k(t)) = \pi\cdot((I-a_d)^{-1}\times$$
$$\times(I-a_d^k)\cdot\mu_d\cdot A^k\cdot m + (a_d\cdot A)^k\cdot s).$$

Враховуючи отримані результати, можемо записати $\mathrm{cov}(Y_t,Y_{t+k})$:

$$\begin{gathered}\lim_{t\to\infty}\mathrm{cov}(Y_t,Y_{t+k}) = \lim_{t\to\infty}(M(Y_t,Y_{t+k}) - M(Y_t)M(Y_{t+k})) = \\ = \pi\cdot((I-a_d)^{-1}\cdot(I-a_d^k)\cdot\mu_d\cdot A^k\cdot m + (a_d\cdot A)^k\cdot s) - \\ -(\pi\cdot(I-a_d\cdot A)^{-1}\cdot\mu)\cdot(\pi\cdot(I-a_d\cdot A)^{-1}\cdot\mu) = \\ = \pi\cdot((I-a_d)^{-1}\cdot(I-a_d^k)\cdot\mu_d\cdot A^k\cdot m + (a_d\cdot A)^k\cdot s) - \\ -(\pi\cdot(I-a_d\cdot A)^{-1}\cdot\mu)^2.\end{gathered}$$

□

2.3 Побудова функції правдоподібності моделі

Як вже зауважувалося, оцінювання параметрів моделі зводиться до пошуку оцінки максимальної правдоподібності

$$\lambda^* = \arg\max_{\lambda} P_\lambda(Y=y), \tag{2.10}$$

де $Y=(Y_1,\ldots,Y_T)$, $y=(y_1,\ldots,y_T)$ [5].

Задача (2.10) розв'язується з використанням ітераційного алгоритму навчання Баума–Велша. А саме, задавши початкове наближення $\lambda^{(0)}$ моделі, знаходимо

$$\lambda^{(n+1)} = \arg\max_{\lambda} Q(\lambda^{(n)},\lambda), \tag{2.11}$$

де $Q(\lambda^{(n)},\lambda)$ — функція квазілогарифмічної правдоподібності, що задається наступною формулою:

$$Q(\lambda,\lambda^*) = \sum_{x=(x_1,\dots,x_T)} P_\lambda(Y=y,X=x)\cdot \ln P_{\lambda^*}(Y=y,X=x). \qquad (2.12)$$

Відомо (2.11), що $\lambda^{(n+1)} \underset{n\to\infty}{\to} \lambda^{**}$, де λ^{**} — точка, в якій досягається локальний максимум функції правдоподібності $P_\lambda(Y=y)$.

Функцію квазілогарифмічної правдоподібності (2.12) можна записати за допомогою функції повної правдоподібності $L_{\lambda,x} = P_\lambda(Y=y,X=x)$:

$$Q(\lambda,\lambda^*) = \sum_{x=(x_1,\dots,x_T)} L_{\lambda,x}\cdot \ln L_{\lambda^*,x}$$

Побудуємо функцію правдоподібності моделі $\lambda = (\pi, A, a_1, a_2, \mu_1, \mu_2, \sigma_1, \sigma_2)$

$$\begin{gathered}L_{\lambda,x} = p(y_0,y_1,\dots,y_T;x_1,\dots,x_T) = \\ = \pi_{x_1}A_{x_1x_2}\cdot\ldots\cdot A_{x_{T-1}x_T}p(y_0)\cdot p(y_1\mid x_1,y_0)\cdot\ldots\cdot p(y_T\mid x_T,y_{T-1}) = \\ = \pi_{x_1}\prod_{i=1}^{T-1} A_{x_ix_{i+1}}\cdot p(y_0)\cdot\prod_{i=1}^{T} p(y_i\mid x_i,y_{i-1})\times \\ \times p(y_1\mid x_1,y_0)\cdot\ldots\cdot p(y_T|x_T,y_{T-1}) = \\ = \pi_{x_1}\prod_{i=1}^{T-1} A_{x_ix_{i+1}}\cdot p(y_0)\cdot\prod_{i=1}^{T} p(y_i\mid x_i,y_{i-1}). \qquad (2.13)\end{gathered}$$

Враховуючи (2.2), матимемо

$$\prod_{i=1}^{T} p(y_i\mid x_i,y_{i-1}) = \prod_{i=1}^{T}\frac{1}{\sqrt{2\pi}\sigma_{x_i}}\exp\left(-\frac{(y_i-(a_{x_i}y_{i-1}+\mu_{x_i}))^2}{2\sigma_{x_i}^2}\right). \qquad (2.14)$$

Винісши з під добутку константи, що повторюються та використавши властивості показника степеня матимемо перетворення (2.14)

$$\prod_{i=1}^{T}\frac{1}{\sqrt{2\pi}\sigma_{x_i}}\exp\left(-\frac{(y_i-(a_{x_i}y_{i-1}+\mu_{x_i}))^2}{2\sigma_{x_i}^2}\right) =$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\prod_{i=1}^{T}\frac{1}{\sigma_{x_i}}\right)\cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{T}\frac{(y_i-(a_{x_i}y_{i-1}+\mu_{x_i}))^2}{{\sigma_{x_i}}^2}\right). \tag{2.15}$$

Оскільки Y_0 має стандартний нормальний розподіл $Y_0 \sim N(0{,}1)$, то враховуючи це та (2.2) отримаємо

$$p(y_0) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y_0^2}{2}}. \tag{2.16}$$

Підставививши результати (2.14), (2.15) та (2.16) у (2.13) отримаємо

$$L_{\lambda,x} = \pi_{x_1}\prod_{i=1}^{T-1}A_{x_ix_{i+1}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y_0^2}{2}}\left(\frac{1}{\sqrt{2\pi}}\right)^T\left(\prod_{i=1}^{T}\frac{1}{\sigma_{x_i}}\right)\times$$
$$\times\exp\left(-\frac{1}{2}\sum_{i=1}^{T}\frac{(y_i-(a_{x_i}y_{i-1}+\mu_{x_i}))^2}{{\sigma_{x_i}}^2}\right). \tag{2.17}$$

Звівши подібні множники отримаємо перетворену формулу (2.17)

$$L_{\lambda,x} = \pi_{x_1}\prod_{i=1}^{T-1}A_{x_ix_{i+1}}(2\pi)^{-\frac{T+1}{2}}\times$$
$$\times\left(\prod_{i=1}^{T}\sigma_{x_i}^{-1}\right)\exp\left(-\frac{1}{2}\left(y_0+\sum_{i=1}^{T}\frac{(y_i-(a_{x_i}y_{i-1}+\mu_{x_i}))^2}{{\sigma_{x_i}}^2}\right)\right). \tag{2.18}$$

2.4 Логарифмування функції прадоподібності

Прологарифмуємо функцію правдоподібності, тобто знайдемо $\ln L_{\lambda,x}$ скориставшися результатом (2.18).

$$\ln L_{\lambda,x} = \ln \pi_{x_1}\prod_{t=1}^{T-1}A_{x_tx_{t+1}}(2\pi)^{-\frac{T+1}{2}}\left(\prod_{t=1}^{T}\sigma_{x_t}^{-1}\right)\times$$
$$\times\exp\left(-\frac{1}{2}\left(y_0+\sum_{t=1}^{T}\frac{(y_t-(a_{x_t}y_{t-1}+\mu_{x_t}))^2}{{\sigma_{x_t}}^2}\right)\right). \tag{2.19}$$

Скориставшись властивостями ln від добутку в (2.19) отримаємо суму логарифмів:

$$\ln L_{\lambda,x} = \ln \pi_{x_1} + \sum_{t=1}^{T-1} \ln A_{x_t x_{t+1}} - \frac{T+1}{2} \ln 2\pi - \sum_{t=1}^{T} \ln \sigma_{x_t} - \\ -\frac{1}{2}\left(y_0 + \sum_{t=1}^{T} \frac{(y_t - (a_{x_t} y_{i-1} + \mu_{x_t}))^2}{{\sigma_{x_t}}^2}\right). \tag{2.20}$$

Оскільки величини з індексом x_i залежать від того чому дорівнює x_i, тобто $x_i = 1$ чи $x_i = 2$, то зручно буде скористатися індикаторною функцією $\mathbb{1}(x_i = k)$, яка прийматиме значення 1, якщо $x_i = k$ та 0 у протилежному випадку, де $k = \overline{1{,}2}$. Тоді (2.20) матиме вигляд:

$$\ln L_{\lambda,x} = -\frac{1}{2}((T+1)\ln 2\pi + y_0) + \ln \pi_1 \cdot \mathbb{1}(x_1 = 1) + \ln \pi_2 \cdot \mathbb{1}(x_1 = 2) + \\ + \sum_{i,j=1}^{2} \ln A_{ij} \sum_{t=1}^{T-1} \mathbb{1}(x_t = i, x_{t+1} = j) - \ln \sigma_1 \sum_{t=1}^{T} \mathbb{1}(x_t = 1) - \\ - \ln \sigma_2 \sum_{t=1}^{T} \mathbb{1}(x_t = 2) - \frac{1}{2} \sum_{t=1}^{T} \mathbb{1}(x_t = 1) \cdot \frac{(y_t - (a_1 y_{t-1} + \mu_1))^2}{\sigma_1^2} - \\ -\frac{1}{2} \sum_{t=1}^{T} \mathbb{1}(x_t = 2) \cdot \frac{(y_t - (a_2 y_{t-1} + \mu_2))^2}{\sigma_2^2}. \tag{2.21}$$

Введемо позначення для виразу що в (2.21) не залежить від параметрів моделі:

$$C = -\frac{1}{2}((T+1)\ln 2\pi + y_0). \tag{2.22}$$

Використавши заміну (2.22) у (2.21) отримаємо остаточний вираз для логарифма натурального функції правдоподібності моделі:

$$\ln L_{\lambda,x} = C + \ln \pi_1 \cdot \mathbb{1}(x_1 = 1) + \ln \pi_2 \cdot \mathbb{1}(x_1 = 2) + \\ + \sum_{i,j=1}^{2} \ln A_{ij} \sum_{t=1}^{T-1} \mathbb{1}(x_t = i, x_{t+1} = j) - \ln \sigma_1 \sum_{t=1}^{T} \mathbb{1}(x_t = 1) -$$

$$-\ln\sigma_2\sum_{t=1}^{T}\mathbb{1}(x_t=2)-\frac{1}{2}\sum_{t=1}^{T}\mathbb{1}(x_t=1)\cdot\frac{(y_t-(a_1y_{t-1}+\mu_1))^2}{\sigma_1^2}-$$
$$-\frac{1}{2}\sum_{t=1}^{T}\mathbb{1}(x_t=2)\cdot\frac{(y_t-(a_2y_{t-1}+\mu_2))^2}{\sigma_2^2}.$$

2.5 Побудова функції квазілогарифмічної правдоподібності

Введемо наступні позначення [7, p. 25–26]:

$$\alpha_t(i)=p(y_0,y_1,\ldots,y_t;X_t=i),\quad t=\overline{1,T},\ i=1{,}2, \tag{2.23}$$
$$\beta_t(i,y)=p(y_{t+1},\ldots,y_T\mid X_t=i,Y_t=y),\quad t=\overline{1,T-1},\ i=1{,}2. \tag{2.24}$$

В цих позначеннях ($\alpha_t(i)$ і $\beta_t(i,y)$) знайдемо коефіцієнти при змінних (або при їх ln) від яких залежить λ^* в функції $Q(\lambda,\lambda^*)$.

2.5.1 Коефіцієнт з $\ln\pi_i^*$

Коефіцієнт з $\ln\pi_i^*$ в функції $Q(\lambda,\lambda^*)$ буде таким:

$$\sum_{x=(x_1,\ldots,x_T)} L_{\lambda,x}\cdot\ln\pi_i^*\cdot\mathbb{1}(x_1=i)=$$
$$=\ln\pi_i^*\cdot\sum_{x=(x_1,\ldots,x_T)} L_\lambda(y_0,\ldots,y_T,x_1,\ldots,x_T)\cdot\mathbb{1}(x_1=i). \tag{2.25}$$

Оскільки функція правдоподібності множиться на індикаторну функцію $\mathbb{1}(x_1=i)$. яка прийматиме значення 1, лише коли $x_1=i$, то сума із (2.25) матиме наступний вигляд:

$$\sum_{x=(x_1,\ldots,x_T)} L_\lambda(y_0,\ldots,y_T,x_1,\ldots,x_T)\cdot\mathbb{1}(x_1=i)=$$
$$=L_\lambda(y_0,\ldots,y_T,x_1=i). \tag{2.26}$$

В свою чергу функція правдоподібності набуде вигляду:

$$L_\lambda(y_0,\ldots,y_T,x_1=i)=p(y_0,\ldots,y_T,x_1=i). \qquad (2.27)$$

Переставимо зміні від яких маємо залежність у (2.27) таким чином, щоб після ввести (2.23) і (2.24)

$$p(y_0,\ldots,y_T,x_1=i)=p(y_0,y_1,x_1=i,y_2,\ldots,y_T)=\alpha_1(i)\cdot\beta_1(i,y_1). \qquad (2.28)$$

Вираз (2.28) — це коефіцієнт при $\ln\pi_i^*$.

2.5.2 Коефіцієнт з $\ln\sigma_i^*$

Врахувавши, що функція правдоподібності множиться на індикаторну функцію $\mathbb{1}(x_t=i)$. яка прийматиме значення 1, лише коли $x_t=i$, то із подвійної суми залишиться лише одна сума і коефіцієнт з $\ln\sigma_i^*$ в функції $Q(\lambda,\lambda^*)$ буде таким:

$$\sum_{x=(x_1,\ldots x_T)} L_{\lambda,x}(-\sum_{t=1}^{T}\ln\sigma_i^*\cdot\mathbb{1}(x_t=i))=-\ln\sigma_i^*\times$$

$$\times\sum_{t=1}^{T}L_\lambda(y_0,\ldots,y_T,x_t=i). \qquad (2.29)$$

Сума із (2.29) матиме наступний вигляд:

$$\sum_{t=1}^{T}L_\lambda(y_0,\ldots,y_T,x_t=i)=\sum_{t=1}^{T}p(y_0,\ldots,y_T,x_t=i) \qquad (2.30)$$

Переставимо зміні від яких маємо залежність у (2.30) таким чином, щоб після ввести (2.23) і (2.24)

$$\sum_{t=1}^{T}p(y_0,\ldots,y_T,x_t=i)=\sum_{t=1}^{T}p(y_0,\ldots,y_t,x_t=i,y_{t+1},\ldots,y_T)=$$

$$= \sum_{t=1}^{T} \alpha_t(i) \cdot \beta_t(i,y_t) \tag{2.31}$$

(2.31) — це коефіцієнт при $\ln \sigma_i^*$

2.5.3 Коефіцієнт з μ_i^* та a_i^*

Оскільки μ_i^* і a_i^* входять разом у функцію правдоподібності, то й коефіцієнт з ними в функції $Q(\lambda,\lambda^*)$ буде спільним і матиме наступний вигляд:

$$\sum_{x=(x_1,\ldots x_T)} L_{\lambda,x}(-\frac{1}{2}\sum_{t=1}^{T} \mathbb{1}(x_t = i) \cdot \frac{(y_t - (a_i^* y_{t-1} + \mu_i{}^*))^2}{\sigma_i^{*2}}) =$$
$$= -\frac{1}{2} \cdot \sum_{t=1}^{T} \frac{(y_t - (a_i^* y_{t-1} + \mu_i^*))^2}{\sigma_i^{*2}} \times$$
$$\times \sum_{x=(x_1,\ldots x_T)} L_\lambda(y_0, \ldots, y_T, x_1, \ldots, x_T) \cdot \mathbb{1}(x_t = i) \tag{2.32}$$

Оскільки функція правдоподібності множиться на індикаторну функцію $\mathbb{1}(x_t = i)$. яка прийматиме значення 1, лише коли $x_1 = i$, то сума із (2.32) матиме наступний вигляд:

$$\sum_{x=(x_1,\ldots,x_T)} L_\lambda(y_0, \ldots, y_T, x_1, \ldots, x_T) \cdot \mathbb{1}(x_t = i) = L_\lambda(y_0, \ldots, y_T, x_t = i) \tag{2.33}$$

Підставивши (2.33) у (2.32) отримуємо:

$$-\frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - (a_i^* y_{t-1} + \mu_i^*))^2}{\sigma_i^{*2}} \cdot L_\lambda(y_0, \ldots, y_T, x_t = i) =$$
$$= -\frac{1}{2} \cdot \sum_{t=1}^{T} \frac{(y_t - (a_i^* y_{t-1} + \mu_i^*))^2}{\sigma_i^{*2}} \cdot p(y_0, \ldots, y_T, x_t = i) \tag{2.34}$$

Переставимо змінні від яких маємо залежність у (2.34) таким чином, щоб після ввести (2.23) і (2.24)

$$-\frac{1}{2}\cdot\sum_{t=1}^{T}\frac{(y_t-(a_i^*y_{t-1}+\mu_i^*))^2}{\sigma_i^{*2}}\cdot p(y_0,\dots,y_T,x_t=i)=$$

$$=-\frac{1}{2}\cdot\sum_{t=1}^{T}\frac{(y_t-(a_i^*y_{t-1}+\mu_i^*))^2}{\sigma_i^{*2}}\cdot p(y_0,\dots,y_t,x_t=i,y_{t+1},\dots,y_T)=$$

$$=-\frac{1}{2}\cdot\sum_{t=1}^{T}\frac{(y_t-(a_i^*y_{t-1}+\mu_i^*))^2}{\sigma_i^{*2}}\cdot \alpha_t(i)\cdot\beta_t(i,y_t) \tag{2.35}$$

Вираз (2.35) — це коефіцієнт з μ_i^* та a_i^*.

2.5.4 Коефіцієнт з $\ln A_{ij}^*$

Змінивши місцями суми та врахувавши, що функція правдоподібності множиться на індикаторну функцію $\mathbb{1}(x_t=i,x_{t+1}=j)$. яка прийматиме значення 1, лише коли водночас $x_t=i$ та $x_{t+1}=j$, то із подвійної суми залишиться лише одна сума і коефіцієнт з $\ln A_{ij}^*$ в функції $Q(\lambda,\lambda^*)$ буде таким:

$$\sum_{x=(x_1,\dots,x_T)} L_{\lambda,x}\sum_{i,j=1}^{2}\ln A_{ij}^*\sum_{t=1}^{T-1}\mathbb{1}(x_t=i,x_{t+1}=j)=$$

$$=\sum_{i,j=1}^{2}\ln A_{ij}^*\cdot\sum_{t=1}^{T-1}L_\lambda(y_0,\dots,y_t,x_t=i,x_{t+1}=j) \tag{2.36}$$

Сума із (2.36) матиме наступний вигляд:

$$\sum_{t=1}^{T-1}L_\lambda(y_0,\dots,y_t,x_t=i,x_{t+1}=j)=\sum_{t=1}^{T-1}p(y_0,\dots,y_T,x_t=i,x_{t+1}=j) \tag{2.37}$$

Переставимо зміні від яких маємо залежність у (2.37) таким чином, щоб після ввести (2.23) і (2.24)

$$\sum_{t=1}^{T-1} p(y_0,\ldots,y_T,x_t=i,x_{t+1}=j)=$$
$$=\sum_{t=1}^{T-1} p(y_0,\ldots y_t,x_t=i,x_{t+1}=j,y_{t+1},\ldots,y_T) \tag{2.38}$$

Врахувававши що функцію ймовірностей можна розкласти на добуток умовних функцій ймовірностей, матимемо:

$$p(y_0,\ldots y_t,x_t=i,x_{t+1}=j,y_{t+1},\ldots,y_T)=p(y_0,\ldots y_t,x_t=i)\times$$
$$\times p(x_{t+1}=j|x_t=i)\cdot p(y_{t+1}|x_{t+1}=j,y_t)\cdot p(y_{t+2},\ldots,y_t|x_{t+1}=j,y_{t+1}) \tag{2.39}$$

Скориставшись (2.23) і (2.24) та сенс A можемо визначити умовні ймовірності із (2.39):

$$p(y_0,y_1,\ldots,y_t;x_t=i)=\alpha_t(i) \tag{2.40}$$
$$p(x_{t+1}=j|x_t=i)=A_{ij} \tag{2.41}$$
$$p(y_{t+1}|x_{t+1}=i,y_t)=B_{iy_t}(y_{t+1})=\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{(y_{t+1}-(a_iy_t+\mu_i))^2}{2\sigma_i^2}} \tag{2.42}$$
$$p(y_{t+2},\ldots y_t|x_{t+1}=j,y_{t+1})=\beta_{t+1}(j,y_{t+1}) \tag{2.43}$$

Підставивши (2.40) (2.41) (2.42) і (2.43) у (2.39):

$$p(y_0,\ldots y_t,x_t=i)\cdot p(x_{t+1}=j|x_t=i)\cdot p(y_{t+1}|x_{t+1}=j,y_t)\times$$
$$\cdot p(y_{t+2},\ldots,y_t|x_{t+1}=j,y_{t+1})=\alpha_t(i)\cdot A_{ij}\cdot B_{jy_t}(y_{t+1})\cdot\beta_{t+1}(j,y_{t+1}) \tag{2.44}$$

Матимемо новий вигляд (2.37):

$$\sum_{t=1}^{T-1} L_\lambda(y_0,\ldots,y_t,x_t=i,x_{t+1}=j)=$$
$$=\sum_{t=1}^{T-1}\alpha_t(i)\cdot A_{ij}\cdot B_{jy_t}(y_{t+1})\cdot\beta_{t+1}(j,y_{t+1}) \tag{2.45}$$

(2.45) — це коефіцієнт при $\ln A^*_{ij}$

2.5.5 Збір функції квазілогарифмічної правдоподібності

Отже, функція $Q(\lambda,\lambda^*)$ має наступний вигляд:

$$Q(\lambda,\lambda^*) = -\frac{1}{2\sigma_1^{*2}}\sum_{t=1}^{T}\alpha_t(1)\beta_t(1,y_t)(y_t-(a_1^*y_{t-1}+\mu_1^*))^2-$$
$$-\frac{1}{2\sigma_2^{*2}}\cdot\sum_{t=1}^{T}\alpha_t(2)\beta_t(2,y_t)(y_t-(a_2^*y_{t-1}+\mu_2^*))^2+\sum_{i,j}\ln A^*_{ij}h_{ij}+$$
$$+\ln\pi_1^*f_1+\ln\pi_2^*f_2-\ln\sigma_1^*g_1-\ln\sigma_2^*g_2+C \tag{2.46}$$

У (2.46) ввели для зручності нові позначення:

$$h_{ij} = \sum_{t=1}^{T-1}\alpha_t(i)A_{ij}B_{jy_t}(y_{t+1})\beta_{t+1}(j,y_{t+1}),$$
$$B_{iy}(z) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp(-\frac{(z-(a_iy+\mu_i))^2}{2\sigma_i^2}),$$
$$f_1 = \alpha_1(1)\beta_1(1,y_1),$$
$$f_2 = \alpha_1(2)\beta_1(2,y_1),$$
$$g_1 = \sum_{t=1}^{T}\alpha_t(1)\beta_t(1,y_t),$$
$$g_2 = \sum_{t=1}^{T}\alpha_t(2)\beta_t(2,y_t),$$
$$C = -\frac{1}{2}((T+1)\ln(2\pi)+y_0) \tag{2.47}$$

2.6 Знаходження оцінок параметрів

2.6.1 Метод множників Лагранжа

Частина наших параметрів пов'язані лінійними залежностями, тому при їх знаходженні застосуємо метод множників Лагранжа.

Ідея методу множників Лагранжа при знаходженні розв'язку задачі нелінійного програмування, полягає в заміні початкової задачі дещо простішою. Для цього цільову функцію замінюють іншою, з більшою кількістю змінних і яка включає в себе умови, що подані як обмеження. Після такого перетворення подальше розв'язування задачі полягає в знаходженні екстремуму нової функції, на змінні якої не накладено ніяких обмежень. Тобто від початкової задачі пошуку умовного екстремуму переходимо до задачі відшукання безумовного екстремального значення іншої функції, яке визначається з допомогою необхідної умови існування екстремуму. Тобто, для розв'язування задачі необхідно знайти вирази частинних похідних нової цільової функції за кожною змінною і прирівняти їх до нуля. В результаті отримаємо систему рівнянь. Її розв'язок визначає так звані стаціонарні точки, серед яких є і шукані екстремальні значення функції.

2.6.2 Оцінки π_i^*

Згідно методу Лагранжа:

$$\pi_1^* + \pi_2^* = 1 \tag{2.48}$$

Запишемо частину функції правдоподібності у якій будуть лише доданки, що залежать від параметрів π_i^* та віднімемо θ, що множиться на 0, який виразимо із (2.48):

$$L(\pi_1^*,\pi_2^*,\theta) = f_1 \ln \pi_1^* + f_2 \ln \pi_2^* - \theta(\pi_1^* + \pi_2^* - 1) \tag{2.49}$$

Знайдемо часткові похідні від (2.54) за кожним параметром від якого зараз залежить $L(\pi_1^*,\pi_2^*,\theta)$

$$\begin{cases} \dfrac{\partial L}{\partial \pi_1^*} = \dfrac{1}{\pi_1^*} f_1 - \theta = 0, \\ \dfrac{\partial L}{\partial \pi_2^*} = \dfrac{1}{\pi_2^*} f_2 - \theta = 0, \\ \dfrac{\partial L}{\partial \theta} = \pi_1^* + \pi_2^* - 1 = 0. \end{cases} \tag{2.50}$$

Виразимо із (2.50) та (2.48) кожен параметр від якого зараз залежить $L(\pi_1^*,\pi_2^*,\theta)$

$$\begin{cases} \pi_1^* = \dfrac{f_1}{\theta} \\ \pi_2^* = \dfrac{f_2}{\theta} \\ \theta = f_1 + f_2 \end{cases} \tag{2.51}$$

Отримуємо оцінки параметрів π_i (з позначеннями із (2.47)):

$$\begin{cases} \pi_1^* = \dfrac{f_1}{f_1 + f_2} \\ \pi_2^* = \dfrac{f_2}{f_1 + f_2} \end{cases} \tag{2.52}$$

Якщо вважати, що система функціонує в стаціонарному режисмі, то вважаємо тоді, що початковий розподіл є стаціонарним.

2.6.3 Оцінки A_{i1}^* і A_{i2}^*

Згідно методу Лагранжа:

$$\forall i \sum_j A_{ij} = 1 \tag{2.53}$$

Запишемо частину функції правдоподібності у якій будуть лише доданки, що залежать від параметрів A_{ij}^* та віднімемо θ, що множиться на 0, який виразимо із (2.53):

$$L(A_{i1}^*,A_{i2}^*,\theta) = h_{i1}\ln A_{i1}^* + h_{i2}\ln A_{i2}^* - \theta(A_{i1}^* + A_{i2}^* - 1) \tag{2.54}$$

Знайдемо часткові похідні від (2.54) за кожним параметром від якого зараз залежить $L(A_{i1}^*,A_{i2}^*,\theta)$

$$\begin{cases} \dfrac{\partial L}{\partial A_{i1}^*} = \dfrac{1}{A_{i1}^*}h_{i1} - \theta = 0 \\ \dfrac{\partial L}{\partial A_{i2}^*} = \dfrac{1}{A_{i2}^*}h_{i2} - \theta = 0 \\ \dfrac{\partial L}{\partial \theta} = A_{i1}^* + A_{i2}^* - 1 = 0 \end{cases} \tag{2.55}$$

Виразимо із (2.55) та (2.53) кожен параметр від якого зараз залежить $L(A_{i1}^*,A_{i2}^*,\theta)$

$$\begin{cases} A_{i1}^* = \dfrac{h_{i1}}{\theta} \\ A_{i2}^* = \dfrac{h_{i2}}{\theta} \\ \theta = h_{i1} + h_{i2} \end{cases} \tag{2.56}$$

Отримуємо оцінки параметрів A_{ij} (з позначеннями із (2.47)):

$$\begin{cases} A_{i1}^* = \dfrac{h_{i1}}{h_{i1} + h_{i2}} \\ A_{i2}^* = \dfrac{h_{i2}}{h_{i1} + h_{i2}} \end{cases} \tag{2.57}$$

2.6.4 Оцінки μ_i^* і a_i^*

Запишемо частину функції правдоподібності у якій будуть лише доданки, що залежать від параметрів μ_i^* та a_i^*:

$$L(\mu_i^*, a_i^*) = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i^* y_{t-1} + \mu_i^*))^2 \tag{2.58}$$

Знайдемо часткові похідні від (2.58) за кожним параметром від якого зараз залежить $L(\mu_i^*, a_i^*)$

$$\begin{cases} \dfrac{\partial L}{\partial \mu_i^*} = -2\sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i^* y_{t-1} + \mu_i^*)) = 0, \\ \dfrac{\partial L}{\partial a_i^*} = -2\sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i^* y_{t-1} + \mu_i^*))y_{t-1} = 0. \end{cases} \tag{2.59}$$

Скоротимо обидві частини рівнянь із системи (2.59):

$$\begin{cases} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i^* y_{t-1} + \mu_i^*)) = 0 \\ \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i^* y_{t-1} + \mu_i^*))y_{t-1} = 0 \end{cases} \tag{2.60}$$

Розкриємо дужки у рівняннях системи (2.60):

$$\begin{aligned} &a_i^* \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} + \mu_i^* \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t) = \\ &\qquad = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t, \\ &a_i^* \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1}^2 + \mu_i^* \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} = \\ &\qquad = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t y_{t-1}. \end{aligned} \tag{2.61}$$

Запишемо систему (2.61) за допомогою матриць:

$$\begin{pmatrix} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t) \\ \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1}^2 & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \end{pmatrix} \cdot \begin{pmatrix} a_i^* \\ \mu_i^* \end{pmatrix} =$$

$$= \begin{pmatrix} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t \\ \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t y_{t-1} \end{pmatrix} \tag{2.62}$$

Розв'яжемо дану систему рівнянь за допомогою метода Крамера

Знайдемо визначник матриці на яку множиться матриця оцінок:

$$\Delta = \begin{vmatrix} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t) \\ \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1}^2 & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \end{vmatrix} =$$

$$= \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \cdot \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k)y_{k-1} - \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1}^2 \times$$

$$\times \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k) = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1}(\sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k)y_{k-1} -$$

$$- y_{t-1} \cdot \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k)) = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k) \times$$

$$\times (y_{k-1} - y_{t-1}) \tag{2.63}$$

Замінимо перший стовпець у матриці на яку множиться матриця оцінок на стовпець вільних членів та знайдемо її визначник:

$$\Delta_1 = \begin{vmatrix} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t) \\ \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t y_{t-1} & \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \end{vmatrix} =$$

$$= \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t \cdot \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k)y_{k-1} - \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_t y_{t-1} \times$$

$$\times\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k) = \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_t(\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_{k-1}-$$
$$-y_{t-1}\cdot\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)) = \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_t\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)\times$$
$$\times(y_{k-1}-y_{t-1}) \tag{2.64}$$

Замінимо перший стовпець у матриці на яку множиться матриця оцінок на стовпець вільних членів та знайдемо її визначник:

$$\Delta_2 = \begin{vmatrix} \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1} & \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_t \\ \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}^2 & \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_t y_{t-1} \end{vmatrix} =$$
$$= \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}\cdot\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_k y_{k-1} - \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}^2\times$$
$$\times\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_k = \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}(\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_k y_{k-1}-$$
$$-y_{t-1}\cdot\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_k) = \sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)y_k\times$$
$$\times(y_{k-1}-y_{t-1}) \tag{2.65}$$

Знайдемо a_i^* як частку Δ_1 і Δ:

$$a_i^* = \frac{\Delta_1}{\Delta} = \frac{\sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_t\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)\cdot(y_{k-1}-y_{t-1})}{\sum_{t=1}^{T}\alpha_t(i)\beta_t(i,y_t)y_{t-1}\sum_{k=1}^{T}\alpha_k(i)\beta_k(i,y_k)\cdot(y_{k-1}-y_{t-1})} \tag{2.66}$$

Знайдемо μ_i^* як частку Δ_2 і Δ:

$$\mu_i^* = \frac{\Delta_2}{\Delta} = \frac{\sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k)y_k \cdot (y_{k-1} - y_{t-1})}{\sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)y_{t-1} \sum_{k=1}^{T} \alpha_k(i)\beta_k(i,y_k) \cdot (y_{k-1} - y_{t-1})} \tag{2.67}$$

2.6.5 Оцінки σ_i^*

Тепер знаходимо оцінки σ_1^* і σ_2^* (підставивши оцінки μ_1^* і μ_2^*). У (2.46) ввели для зручності нові позначення і біля σ_i^* знаходитиметься наступна змінна:

$$g_i = \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t) \tag{2.68}$$

Запишемо частину функції правдоподібності у якій будуть лише доданки, що залежать від параметрів σ_i^*:

$$L(\sigma_i^*) = -\ln \sigma_i^* g_i - \frac{1}{2\sigma_i^{*2}} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i y_{t-1} + \mu_i^*))^2 \tag{2.69}$$

Знайдемо часткову похідну від (2.69) за параметром від якого зараз залежить $L(\sigma_i^*)$

$$\frac{\partial L}{\partial \sigma_i^*} = -\frac{1}{\sigma_i^*} g_i + \frac{1}{\sigma_i^{*3}} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i y_{t-1} + \mu_i^*))^2 = 0 \tag{2.70}$$

Перетворимо рівняння із (2.70):

$$\frac{1}{\sigma_i^{*3}} \sum_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i y_{t-1} + \mu_i^*))^2 = \frac{1}{\sigma_i^*} g_i \tag{2.71}$$

Скоротимо (2.71) та виразимо звідти σ_i^{*2} (з позначеннями із (2.68)):

$$\sigma_i^{*2} = \frac{\sum\limits_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - (a_i y_{t-1} + \mu_i^*))^2}{g_i} =$$

$$= \frac{\sum\limits_{t=1}^{T} \alpha_t(i)\beta_t(i,y_t)(y_t - a_i y_{t-1} - \mu_i^*)^2}{g_i} \tag{2.72}$$

Висновки до розділу 2

У розділі ми встановили деякі властивості рівняння авторегресії з марківськими перемиканнями пов'язані з існуванням стаціонарного розв'язку цього рівняння та побудували оцінки невідомих параметрів цього рівняння з використанням ітераційного ЕМ-алгоритму знаходження оцінки максимальної правдоподібності.

3 ЧИСЕЛЬНИЙ ЕКСПЕРИМЕНТ

Чисельний експеремент виконується за допомогою програмних можливостей мови Python. Зокрема бібліотек numpy, math та matplotlib. З бібліотеки numpy використовується модуль random для генерування випадкових величин (за якими саме правилами це буде зроблено визначають функції модуля) та модуль array для роботи з багатовимірними масивами. З бібліотеки math використовуються математичні функції (експонента, квадратний корінь, число π тощо). Бібліотека matplotlib використовується для побудови графіків. Додаток А містить код основних інструментальних програм для проведення експерементальних досліджень.

3.1 Генерування даних спостереження

Спершу ми обираємо кількість спостережуваних даних T та задаємо параметри моделі з якими ми будемо генерувати ці дані, а саме параметри: $(A_{ij})_{i,j=1,2}$, $\pi = (\pi_1, \pi_2)$, $\mu = (\mu_1, \mu_2)$, $a = (a_1, a_2)$, $\sigma^2 = (\sigma_1^2, \sigma_2^2)$. Тут є певні обмеження:

1) матриці $(A_{ij})_{i,j=1,2}$ та $\pi = (\pi_1, \pi_2)$ є стохастичними (усі елементи матриць є невід'ємними, а сума елементів рядків рівна одиниці);

2) елементи матриці $a = (a_1, a_2)$ є додатніми, але меншими за одиницю;

3) елементи матриці $\sigma^2 = (\sigma_1^2, \sigma_2^2)$ є додатніми та їх краще обирати меншими за 0,1 (тоді дані спостережень не будуть надто зашумленні, адже елементи матриці задають дисперсію шуму).

Далі генеруємо прихований марківський ланцюг довжини T. Для цього спершу нам потрібно згенерувати допоміжний вектор u випадкових величин, які рівномірно розподілені на інтервалі (0,1), довжини T

(робимо це за допомогою функції uniform модулю random із бібліотеки numpy). Перший елемент прихованого марківського ланцюга отримуємо при порівнянні першого елемента вектора u із першим елементом вектора початкового розподілу: якщо $u_1 \leqslant \pi_1$, то $x_1 = 1$ та в протилежному $x_1 = 2$. Наступний, другий елемент прихованого марківського ланцюга отримуємо при порівнянні другого елемента вектора u із першим елементом матриці перехідних ймовірностей, а рядок матриці визначається залежно від того яким був попередній, перший x: якщо $u_2 \leqslant A_{x_1 1}$, то $x_2 = 1$ та в протилежному $x_2 = 2$. Усі наступні x_i, заповнються аналогічно до x_2, при порівнянні відповідних u_i із першим елементом матриці перехідних ймовірностей, а рядок матриці визначається залежно від того яким був попередній x_{i-1}: якщо $u_i \leqslant A_{x_{i-1}1}$, то $x_i = 1$ та в протилежному $x_i = 2$.

Далі генеруємо вектор даних спостережень довжини $T + 1$. Для цього спершу нам потрібно згенерувати згенерувати дві незалежні від прихованого марківського ланцюга $(X_n)_{n \geqslant 1}$ та між собою послідовності незалежних однаково розподілених випадкових величин $\{\varepsilon_n^1\}_{n=1,T}$: $\varepsilon_n^1 \sim N(0,\sigma_1^2)$, $\{\varepsilon_n^2\}_{n=1,T}$: $\varepsilon_n^2 \sim N(0,\sigma_2^2)$ (робимо це за допомогою функції normal модулю random із бібліотеки numpy). Далі генеруємо початкову спостережувану величину Y_0 із стандартним нормальним розподілом (робимо це за допомогою функції normal модулю random із бібліотеки numpy) та генеруємо згідно (2.1) спостережуванні дані.

Виконуємо візуалізацію спостережуванних даних у вигляді графіку за допомогою бібліотеки matplotlib та функції scatter.

Після того як ми отримали повністю дані спостережень та прихований ланцюг Маркова, використовуючи обрані параметри можна обрахувати матрицю, що пов'язує приховані стани спостережень із самими спостереженнями (за формулою 2.42). Оскільки ця матриця буде непрямо переоцінюватися, якщо використати у ній оцінки параметрів, то записуємо її у вигляді функції, що можна буде використовувати у ітераційному алгоритмі.

3.2 Ітераційний алгоритм

Реалізацію ітераційного алгоритму можна розбити на декілька основних частин:

1) запис функцій для обрахунку коефіцієнтів прямого ходу (за формулою 2.23) та зворотного ходу (за формулою 2.24);

2) запис оцінок параметрів AR ПММ записаних у формулах (2.52), (2.57), (2.66), (2.67) та (2.72);

3) запис функції, що описує одну ітерацію;

4) запис функції, що змушує виконуватися попередньо задану функцію необхідну кількість разів.

Після виконання ітераційного алгоритму, отримуємо графік, що демонструє збіжність параметрів до своїх істинних значень.

3.3 Результати чисельного експерименту

Спершу ми здійснили пошук вектора початкового розподілу $\pi = (\pi_1, \pi_2)$ та матриці перехідних ймовірностей

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

У таблиці Б.1 з додатку Б наведено отримані значення матриці перехідних ймовірностей A для різних значень дисперсії шуму σ^2, кількості даних T та кількості ітерацій $iters$.

Як бачимо із таблиці, збіжність до істинних значень елементів матриці перехідних ймовірностей A чіткіша за мешних значень дисперсії шуму (навіть при меншій кількості ітерацій), кількість необхідних ітерацій в основному залежить від кількості даних спостережень.

Розглянемо детальніше один із випадків.

Обираємо наступну кількість спостережуваних даних: $T = 1000$. Початкові параметри моделі зазначені на рис. 3.1.

```
A= [[0.7, 0.3], [0.1, 0.9]]
pi= [0.3, 0.7]
mu= [1, 2]
sigma= [0.9, 1.6]
a= [0.4, 0.5]
```

Рисунок 3.1 – Початкові параметри моделі.

На рис. 3.2 зображені згенеровані дані при обраних параметрах. Тут червоний колір вказує, що в момент часу t режим був 1, а зелений — що в момент часу t режим був 2.

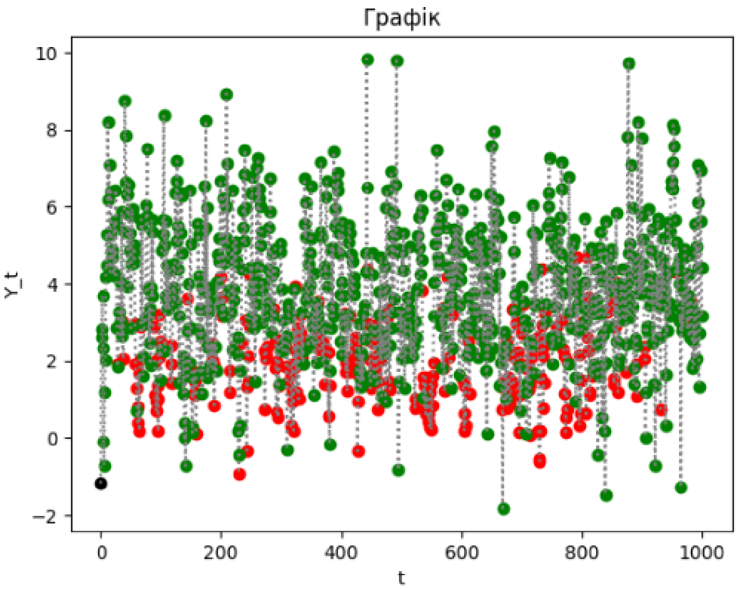

Рисунок 3.2 – Згенеровані дані.

На рис. 3.3 зображені початкові значення параметрів перед початком ітераційного алгоритму, які шукатимемо у ньому. Дані коефіцієнти зазвичають обирають близькими до значення $\frac{1}{N}$, де N — кількість станів прихованого ланцюга. Також при виборі початкових значень цих коефіцієнтів важливо пам'ятати, що вектор початкового розподілу π та матриця перехідних ймовірностей A є стохастичними за рядками.

```
A= [[0.501, 0.499], [0.499, 0.501]]
pi= [0.499, 0.501]
```

Рисунок 3.3 – Значення параметрів перед початком ітераційного алгоритму.

На рис. 3.4 зображені значення параметрів після закінчення ітераційного алгоритму, тобто після 1000 ітерацій.

```
PI= [0.0, 1.0]
A= [[0.7388585778071096, 0.26114142219289166], [0.09663716098064867, 0.9033628390193533]]
PIreal= [0. 1.]
Areal= [[0.7, 0.3], [0.1, 0.9]]
```

Рисунок 3.4 – Значення параметрів після закінчення ітераційного алгоритму.

На рис. 3.5 графічно зображена збіжність елементів матриці A до своїх істинних значень поітераційно.

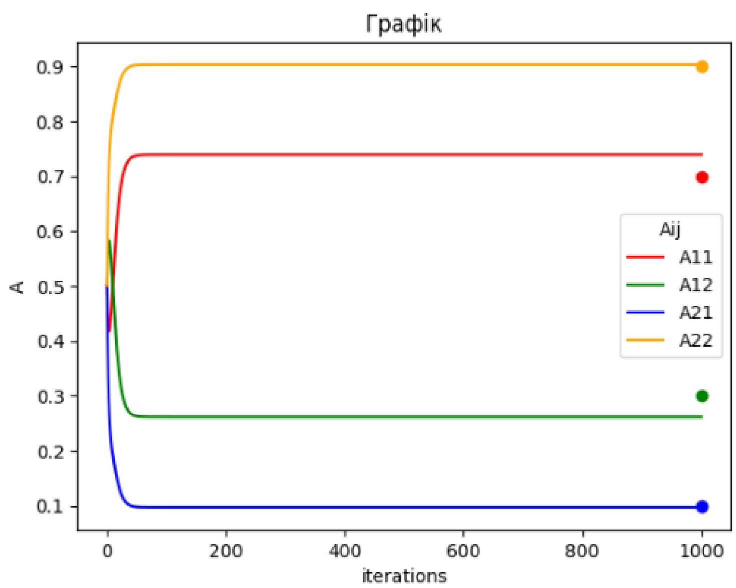

Рисунок 3.5 – Збіжність матриці перехідних ймовірностей.

З рис. 3.6 бачимо, що ланцюг розпочався із стану 1, тобто $x_1 = 2$ (графічно це видно з того, що на рис. 3.2 спостереження в момент часу $t = 1$ зафарбоване у зелений колір, як і значення, що дорівнює 1 у векторі π з рис. 3.6).

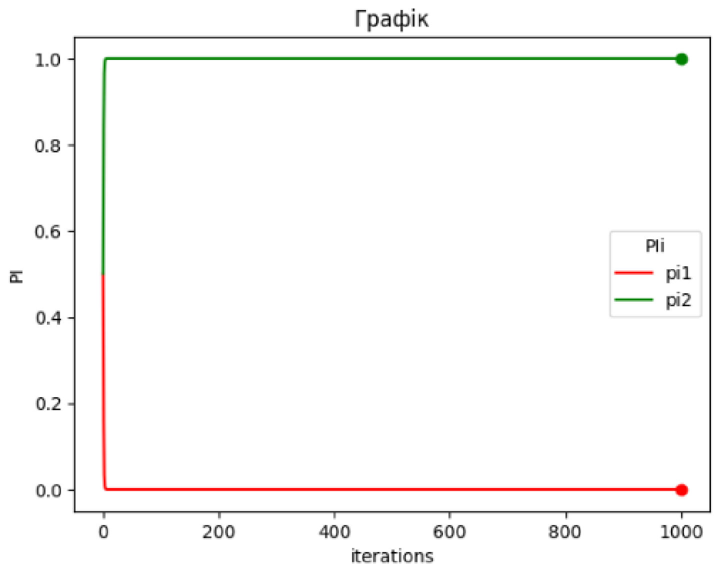

Рисунок 3.6 – Збіжність вектора початкового розподілу.

Розглянемо ще один випадок у якому $T = 100$, згенеровані дані зображені на рис 3.7.

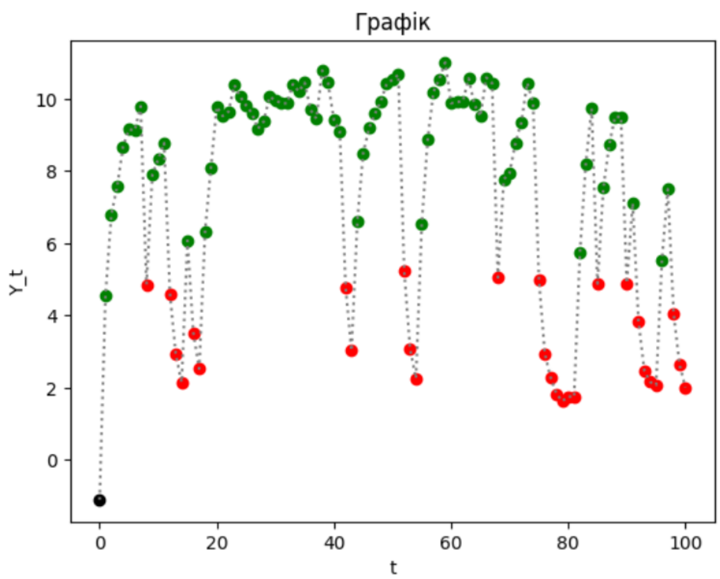

Рисунок 3.7 – Згенеровані дані.

У ньому оцінюємо вектор π, матрицю A та параметри μ. Результати виконання ітераційного алгоритму($iterations = 400$) у даному випадку дає результати наведені на рис. 3.8.

```
PI= [0.0, 1.0]
A= [[0.6296296295024403, 0.37037037049755983], [0.15277777777151813, 0.847222222228482]]
MU [1.0522016375535583, 4.882125247202028]
a= [0.4, 0.5]
SIGMA^2= [0.1, 0.4]
PIreal= [0. 1.]
Areal= [[0.7, 0.3], [0.1, 0.9]]
MUreal= [1, 5]
aReal= [0.4, 0.5]
SIGMAreal^2= [0.1, 0.4]
```

Рисунок 3.8 – Значення параметрів після закінчення ітераційного алгоритму.

Наступні три рисунки демонструють збіжність цих величин до своїх істинних значень.

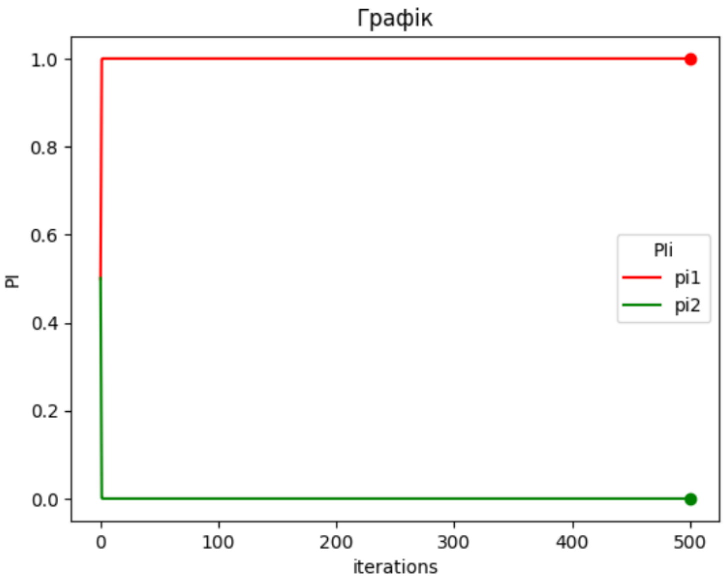

Рисунок 3.9 – Збіжність вектора початкового розподілу.

Як нескладно помітити у цьому випадку збіжність елементів матриці A не така хороша, як коли інші параметри «фіксовані».

При дослідженні ми дійшли до висновку, щоб змогти водночас відшукати істинні значення якнайбільшої кількості параметрів рівняння

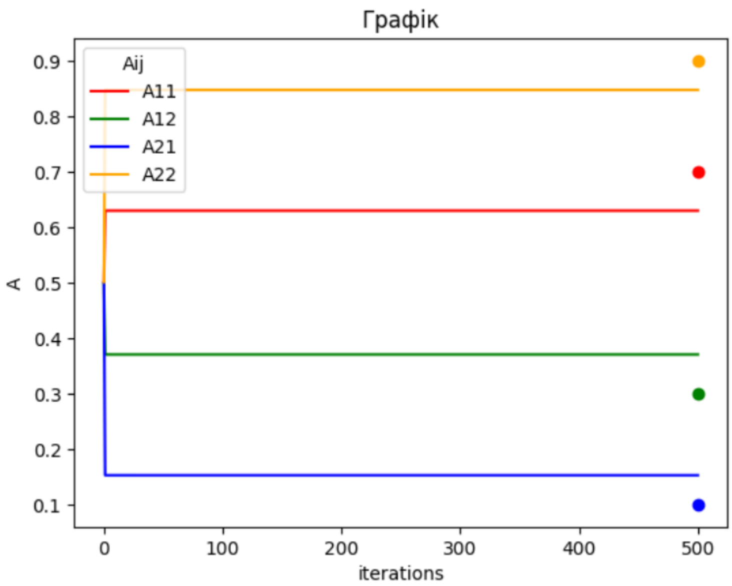

Рисунок 3.10 – Збіжність матриці перехідних ймовірностей.

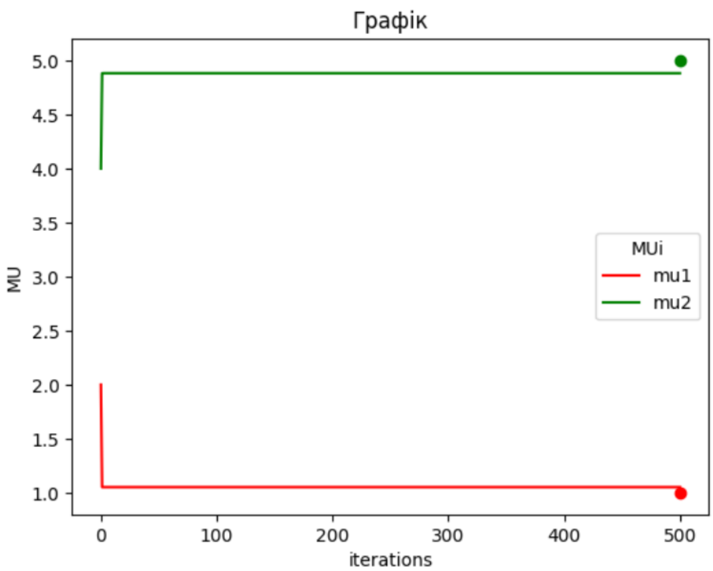

Рисунок 3.11 – Збіжність параметра μ.

необхідно перед застосуванням ітераційного алгоритму провести додаткову кластеризацію методом k–середніх параметрів рівняння a, μ та σ^2, що й планується зробити при подальшій роботі над дослідженням.

Висновки до розділу 3

У розділі описано чисельний експеримент виконаний за допомогою програмних можливостей мови Python, описано реалізацію ітераційного алгоритму та проаналізовано результати отримані при дослідженні. Також описано напрямок подальших досліджень, а саме проведення додаткової кластеризації методом k–середніх параметрів рівняння a, μ та σ^2 перед застосуванням ітераційного алгоритму, щоб пришвидшити збіжність усіх параметрів до їх істинного значення. Даний розділ містить два додатки, з текстами основних програм та з таблицею сформованих результатів, щодо знаходження матриці A.

ВИСНОВКИ

Для дослідження було проведено пошук та аналіз опублікованих джерел за тематикою дослідження.

Завдяки дослідженим джерелам з тематики ПММ вдалося більш детально ознайомитися із ПММ, які мають широке застосування в задачах розпізнавання мовлення, вирівнювання послідовностей в біоінформатиці, виявлення атак в криптографії тощо, та їх модифікацією AR ПММ. Було наведено основні означення ланцюга Маркова та ПММ, а також описано задачі, які існують в рамках цієї моделі, та наведено їх розв'язок.

У досліджені описано модель, встановлено деякі властивості рівняння авторегресії з марківськими перемиканнями пов'язані з існуванням стаціонарного розв'язку цього рівняння, а також побудовано оцінки параметрів прихованого марківського ланцюга, що керує перемиканнями у цьому рівнянні.

Програмна ралізація рівняння авторегресії з марківськими перемиканнями виконана за допомогою мови Python, у ній згенеровано дані рівняння авторегресії з марківськими перемиканнями та імплементовано ітераційний алгоритм навчання Баума—Велша.

В подальшій роботі над дослідження планується провести додаткову кластеризацію методом k–середніх параметрів рівняння a, μ та σ^2 перед застосуванням ітераційного алгоритму, щоб пришвидшити збіжність усіх параметрів до їх істинного значення.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Terence C. Mills. *Time series techniques for economists.* Англ. Cambridge University Press, 1990, 377 p.

[2] А. Н. Ширяев. *Основы стохастической финансовой математики.* Рос. Т. Том 1. Факты. Модели. 2 т. МЦНМО, 2016. 440 с.

[3] James D. Hamilton. *A new approach to the economic analysis of nonstationary time series and the business cycle.* Англ. Econometrica. 1989.

[4] О.В. Мочук та І.І. Ніщенко. *Рівняння авторегресії з марковськими перемиканнями.* XXI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми фізики, математики та інформатики». 2023.

[5] T. Koski. *Hidden Markov models for bioinformatics.* Англ. Kluwer Academic Publishers, 2001.

[6] Rabiner L.R. *A tutorial on hidden Markov models and selected applications in speech recognition.* Англ. Proceedings of the IEEE 77: 257-286, 1989.

[7] Mikael Nilsson. *First order hidden markov model. Theory and implementation issues.* Англ. Research Report Department of Signal Processing Blekinge Institute of Technology, Sweden, 2005.

[8] V. Barbu та N. Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications.* Англ. 2008, 377 p.

ДОДАТОК А ТЕКСТИ ПРОГРАМ

У цьому додатку наведені тексти основних інструментальних програм для проведення експериментальних досліджень. Перелік необхідних бібліотек мови Python, що використовуються, наведений нижче:

```
import numpy as np
import math as m
import matplotlib.pyplot as plt
from numpy import random
from numpy import array
from numpy import inf
from numpy.ma.core import log
```

А.1 Генерування даних

Задання кількості даних та параметрів з якими їх генеруємо

```
T = 400

pi = [0.3, 0.7]
A= [[0.7, 0.3], [0.1, 0.9]]
mu = [1, 4]
sigma = [1, 4]
a = [0.4, 0.5]
```

Генеруємо u вектор допоміжних величин, які рівномірно розподілені на проміжку (0,1) довжини T

```
u=[]

for k in range(T):
    u.append(np.random.uniform(0, 1))
```

Генеруємо ланцюг Маркова довжини T

```
x=[]

if u[0]<=pi[0]:
    X_1=1
else:
    X_1=2
x.append(X_1)

i=X_1

for j in range(1, T):
  if u[j]<=A[i-1][0]:
    X_j=1
  else:
    X_j=2
  x.append(X_j)
  i=X_j
```

Генеруємо два вектори шумів e довжини T

```
e1=[]
e2=[]

for l in range(T):
    e1.append(np.random.normal(0, sigma[0]))
    e2.append(np.random.normal(0, sigma[1]))
```

Генеруємо дані спостереження: величину y_0 та вектор y наступних T спостережень, пов'язаних з марківським процесом (за формулою (2.1))

```
y_0=np.random.normal(0, 1)

y=[]
```

```

if (x[0]==1):
  y_1=mu[0]+a[0]*y_0+e1[0]
else:
  y_1=mu[1]+a[1]*y_0+e2[0]
y.append(y_1)

for n in range(1,T):
  if (x[n]==1):
    y_n=mu[0]+a[0]*y[n-1]+e1[n]
  else:
    y_n=mu[1]+a[1]*y[n-1]+e2[n]
  y.append(y_n)
```

Виводимо графічно дані спостереження

```
t = []

for i in range(0, T+1):
  t.append(i)

t_1=[]
y_1=[]
t_2=[]
y_2=[]

for k in range(T):
  if x[k]==1:
    t_1.append(t[k+1])
    y_1.append(y[k])
  else:
    t_2.append(t[k+1])
    y_2.append(y[k])

Y=[]
Y.append(y_0)
for k in range(T):
  Y.append(y[k])

plt.scatter(t_1, y_1, color='red')
plt.scatter(t_2, y_2, color='green')
plt.scatter(0, y_0, color='black')
```

```
plt.plot(t, Y, linestyle = 'dotted', color='grey')
plt.title("Graph")
plt.xlabel("t")
plt.ylabel("Y_t")
plt.show()
```

Заповнення матриці, що пов'язує приховані стани спостережень із самими спостереженнями(за формулою (2.2))

```
def matrixB(sigma, a, mu):
  B0=1/(m.sqrt(2*m.pi))*m.exp(-1*y_0**2/2)

  B=np.zeros((N, T))

  for i in range(N):
    B[i][0]=m.exp(-1*(y[0]-(a[i]*y_0+mu[i]))**2/(2*sigma[i]))/(m.sqrt(2*m.pi*sigma[i]))

    for j in range(1, T):
      B[i][j]=m.exp(-1*(y[j]-(a[i]*y[j-1]+mu[i]))**2/(2*sigma[i]))/(m.sqrt(2*m.pi*sigma[i]))

  return(B0, B)
```

A.2 Ітераційний алгоритм

Введення стартових параметрів для алгоритму та запам'ятовування істинних

```
Areal=A
PIreal=pi
Breal=B
MUreal=mu
SIGMAreal=sigma
aReal=a

A=[[0.501, 0.499], [0.499, 0.501]]
pi=[0.499, 0.501]
```

```

pi1=[]
pi2=[]
pi1.append(pi[0])
pi2.append(pi[1])

A11=[]
A12=[]
A21=[]
A22=[]
A11.append(A[0][0])
A12.append(A[0][1])
A21.append(A[1][0])
A22.append(A[1][1])

mu1=[]
mu2=[]
mu1.append(mu[0])
mu2.append(mu[1])

a1=[]
a2=[]
a1.append(a[0])
a2.append(a[1])

sigma1=[]
sigma2=[]
sigma1.append(sigma[0])
sigma2.append(sigma[1])

minIters=500
iters=0
```

Функція знаходження коефіцієнтів прямого ходу

```
def matrixALPHA(pi, A, B0, B):
  c=np.zeros(T)
  alpha=np.zeros((T, N))

  #alpha_0(i)
  for i in range(N):
    alpha[0][i]=pi[i]*B0*B[i][0]
```

```
    # alpha[0][i]=pi[i]*B[i][0]
    c[0]=c[0]+alpha[0][i]

  #SHKALalpha_0(i)
  c[0]=1/c[0]
  for i in range(N):
    alpha[0][i]=c[0]*alpha[0][i]

  #alpha_t(i)
  for t in range(1, T):
    for i in range(N):
      for j in range(N):
        alpha[t][i]=alpha[t][i]+alpha[t-1][j]*A[j][i]
      alpha[t][i]=alpha[t][i]*B[i][t]
      c[t]=c[t]+alpha[t][i]

    #SHKALalpha_t(i)
    c[t]=1/c[t]
    for i in range(N):
      alpha[t][i]=c[t]*alpha[t][i]

  return(c, alpha)
```

Функція знаходження коефіцієнтів зворотного ходу

```
def matrixBETA(c, A, B):
  beta=np.zeros((T, N))

  #beta_T-1(i)
  for i in range(N):
    beta[T-1][i]=c[T-1]

  #beta_t(i)
  for t in range(T-2, -1, -1):
    for i in range(N):
      for j in range(N):
        beta[t][i]=beta[t][i]+A[i][j]*B[j][t+1]*beta[t+1][j]

      #SHKALbeta_t(i)
      beta[t][i]=c[t]*beta[t][i]

  return(beta)
```

Додаткова функція для знаходження параметрів A_{ij} та π_i

```
def matixsGAMMA(alpha, beta, A, B):
  gamma_i=np.zeros((T, N))
  gamma_ij=np.zeros((T-1, N, N))

  for t in range(T-1):

    denom=10**(-300)
    for i in range(N):
      for j in range(N):
        denom=denom+alpha[t][i]*A[i][j]*B[j][t+1]*beta[t+1][j]
    for i in range(N):
      for j in range(N):
        gamma_ij[t][i][j]=(alpha[t][i]*A[i][j]*B[j][t+1]*beta[t+1][j])/denom
        gamma_i[t][i]=gamma_i[t][i]+gamma_ij[t][i][j]

  gamma_T-1(i)
  denom=0
  for i in range(N):
    denom=denom+alpha[T-1][i]
  for i in range(N):
    gamma_i[T-1][i]=alpha[T-1][i]/denom

  return(gamma_i, gamma_ij)
```

Додаткова функція для знаходження параметрів μ та a у випадку без шкалювання

```
def ProdShkalCoef(c):
  P = 0

  for t in range(T):
    P=P*c[t]

  return(P)
```

Додаткова функція для знаходження параметрів μ та a

```
def wShkal(A, B, alpha, beta, P, sigma, a, mu):
  w = np.zeros((T, N))

  for t in range(T-1):
    for i in range(N):
      for j in range (N):
        w[t][i]=w[t][i]+A[i][j]*B[j][t+1]*beta[t+1][j]
      w[t][i]=alpha[t][i]*w[t][i]

  for i in range(N):
    w[T-1][i]=alpha[T-1][i]

  #notShkal
  # for t in range(T):
  #   w[t][i]=P*w[t][i]

  return(w)
```

Оцінка елементів вектора π

```
def re_estimatePI(gamma_i, pi):
  for i in range(N):
    pi[i]=gamma_i[0][i]

  return(pi)
```

Оцінка елементів матриці A

```
def re_estimateA(gamma_i, gamma_ij, A):
  for i in range(N):
    for j in range(N):
      numer=0
      denom=0

      for t in range(T-1):
        numer=numer+gamma_ij[t][i][j]
```

```
      denom=denom+gamma_i[t][i]

    A[i][j]=numer/denom
  return(A)
```

Оцінка параметрів μ та a

```
def re_estimateMUandA(w, a, mu):
  sum1=np.zeros(N)
  sum2=np.zeros(N)
  sum3=np.zeros(N)
  sum4=np.zeros(N)
  sum5=np.zeros(N)
  K=np.zeros(N)

  for i in range(N):
    sum2[i]=w[0][i]*y_0**2
    sum3[i]=+w[0][i]*y_0

    for t in range(1, T):
      sum1[i]=sum1[i]+w[t][i]
      sum2[i]=sum2[i]+w[t][i]*y[t-1]**2
      sum3[i]=sum3[i]+w[t][i]*y[t-1]

    K[i]=sum1[i]*sum2[i]-sum3[i]**2

  for i in range(N):
    sum4[i]=w[0][i]*y_0
    sum5[i]=w[0][i]*y_0*y[0]

    for t in range(1, T):
      sum4[i]=sum4[i]+w[t][i]*y[t]
      sum5[i]=sum5[i]+w[t][i]*y[t-1]*y[t]

    mu[i]=(sum2[i]*sum4[i]-sum3[i]*sum5[i])/K[i]
    a[i]=(sum1[i]*sum5[i]-sum3[i]*sum4[i])/K[i]

  return(mu, a)
```

Оцінка дисперсії шуму σ^2

```
def re_estimateSIGMA(alpha, beta, sigma, a, mu):
  g=np.zeros(N)
  for i in range(N):
    for t in range(T):
      g[i]=g[i]+alpha[t][i]*beta[t][i]

  for i in range(N):
    sigma[i]=alpha[0][i]*beta[0][i]*(y[0]-(a[i]*y_0+mu[i]))**2
    for t in range(1, T):
      sigma[i]=sigma[i]+alpha[t][i]*beta[t][i]*(y[t]-(a[i]*y[t-1]+mu[i]))**2
    sigma[i]=sigma[i]/g[i]

    return(sigma)
```

Функція однієї ітерації

```
def ONEiters(pi, A, B0, B, sigma, a, mu):
  #Calculation of coefficients
  c, alpha = matrixALPHA(pi, A, B0, B)
  beta = matrixBETA(c, A, B)
  gamma_i, gamma_ij = matixsGAMMA(alpha, beta, A, B)
  P = ProdShkalCoef(c)
  w = wShkal(A, B, alpha, beta, P, sigma, a, mu)

  #Reevaluation of the model
  pi = re_estimatePI(gamma_i, pi)
  A = re_estimateA(gamma_i, gamma_ij, A)
  mu, a = re_estimateMUandA(w, a, mu)
  sigma = re_estimateSIGMA(alpha, beta, sigma, a, mu)
  B0, B = matrixB(sigma, a, mu)

  #Filling data for graphs
  pi1.append(pi[0])
  pi2.append(pi[1])
  A11.append(A[0][0])
  A12.append(A[0][1])
  A21.append(A[1][0])
  A22.append(A[1][1])
  mu1.append(mu[0])
```

```
  mu2.append(mu[1])
  a1.append(a[0])
  a2.append(a[1])
  sigma1.append(sigma[0])
  sigma2.append(sigma[1])

  return(pi, A, B)
```

Імплеменетація ітераційного алгоритму

```
def Implem(iters, pi, A, B):
  B0, B = matrixB(sigma, a, mu)
  while (iters<minIters):
    pi, A, B = ONEiters(pi, A, B0, B, sigma, a, mu)
    iters=iters+1

  print("PI=", pi, "\nA=", A, "\nMU", mu, "a=", a,  "\nSIGMAreal^2=", sigma)
  print("PIreal=", PIreal, "\nAreal=", Areal, "\nMUreal=", MUreal, "\naReal=", aReal, "\
    nSIGMAreal=", SIGMAreal)

  return(iters)
```

Графік збіжності елементів вектора π

```
iterations = range(iters+1)
plt.plot(iterations, pi1, color='red', label='pi1')
plt.plot(iterations, pi2, color='green', label='pi2')
plt.scatter(iters, PIreal[0], color='red')
plt.scatter(iters, PIreal[1], color='green')
plt.title("Graph")
plt.xlabel("iterations")
plt.ylabel("PI")
plt.legend(title='PIi')
plt.show()
```

Графік збіжності елементів матриці A

```
iterations = range(iters+1)
plt.plot(iterations, A11, color='red', label='A11')
plt.plot(iterations, A12, color='green', label='A12')
plt.plot(iterations, A21, color='blue', label='A21')
plt.plot(iterations, A22, color='orange', label='A22')
plt.scatter(iters, Areal[0][0], color='red')
plt.scatter(iters, Areal[0][1], color='green')
plt.scatter(iters, Areal[1][0], color='blue')
plt.scatter(iters, Areal[1][1], color='orange')
plt.title("Graph")
plt.xlabel("iterations")
plt.ylabel("A")
plt.legend(title='Aij')
plt.show()
```

Графік збіжності параметрів μ

```
iterations = range(iters+1)
plt.plot(iterations, mu1, color='red', label='mu1')
plt.plot(iterations, mu2, color='green', label='mu2')
plt.scatter(iters, MUreal[0], color='red')
plt.scatter(iters, MUreal[1], color='green')
plt.title("Graph")
plt.xlabel("iterations")
plt.ylabel("MU")
plt.legend(title='MUi')
plt.show()
```

Графік збіжності параметрів a

```
iterations = range(iters+1)
plt.plot(iterations, a1, color='red', label='a1')
plt.plot(iterations, a2, color='green', label='a2')
plt.scatter(iters, aReal[0], color='red')
plt.scatter(iters, aReal[1], color='green')
plt.title("Graph")
plt.xlabel("iterations")
plt.ylabel("a")
plt.legend(title='a_i')
plt.show()
```

Графік збіжності дисперсії шуму σ^2

```
iterations = range(iters+1)
plt.plot(iterations, sigma1, color='red', label='sigma1')
plt.plot(iterations, sigma2, color='green', label='sigma2')
plt.scatter(iters, SIGMAreal[0], color='red')
plt.scatter(iters, SIGMAreal[1], color='green')
plt.title("Graph")
plt.xlabel("iterations")
plt.ylabel("SIGMA^2")
plt.legend(title='SIGMA_i^2')
plt.show()
```

ДОДАТОК Б ВЕЛИКА ТАБЛИЦЯ

Табл. Б.1: Знайдені значення матриці A

Істинна A	$\begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}$		
T	iters		
	500	1000	2500
σ^2	$\begin{pmatrix} 0.01 & 0.04 \end{pmatrix}$		
125	$\begin{pmatrix} 0.722 & 0.278 \\ 0.114 & 0.886 \end{pmatrix}$	$\begin{pmatrix} 0.76 & 0.24 \\ 0.051 & 0.949 \end{pmatrix}$	$\begin{pmatrix} 0.714 & 0.286 \\ 0.073 & 0.927 \end{pmatrix}$
250	$\begin{pmatrix} 0.714 & 0.286 \\ 0.097 & 0.903 \end{pmatrix}$	$\begin{pmatrix} 0.677 & 0.323 \\ 0.107 & 0.893 \end{pmatrix}$	$\begin{pmatrix} 0.759 & 0.241 \\ 0.112 & 0.888 \end{pmatrix}$
500	$\begin{pmatrix} 0.673 & 0.327 \\ 0.089 & 0.911 \end{pmatrix}$	$\begin{pmatrix} 0.714 & 0.286 \\ 0.099 & 0.901 \end{pmatrix}$	$\begin{pmatrix} 0.71 & 0.29 \\ 0.096 & 0.904 \end{pmatrix}$
σ^2	$\begin{pmatrix} 0.1 & 0.4 \end{pmatrix}$		
125	$\begin{pmatrix} 0.714 & 0.286 \\ 0.135 & 0.865 \end{pmatrix}$	$\begin{pmatrix} 0.734 & 0.266 \\ 0.062 & 0.938 \end{pmatrix}$	$\begin{pmatrix} 0.654 & 0.346 \\ 0.065 & 0.935 \end{pmatrix}$
250	$\begin{pmatrix} 0.685 & 0.315 \\ 0.064 & 0.936 \end{pmatrix}$	$\begin{pmatrix} 0.709 & 0.291 \\ 0.082 & 0.918 \end{pmatrix}$	$\begin{pmatrix} 0.682 & 0.318 \\ 0.061 & 0.939 \end{pmatrix}$

Продовження на наступній сторінці

Табл. Б.1: Знайдені значення матриці A (продовження)

Істинна A	$\begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}$		
T	iters		
	500	1000	2500
σ^2	$\begin{pmatrix} 0.1 & 0.4 \end{pmatrix}$		
500	$\begin{pmatrix} 0.779 & 0.221 \\ 0.059 & 0.941 \end{pmatrix}$	$\begin{pmatrix} 0.691 & 0.309 \\ 0.061 & 0.939 \end{pmatrix}$	$\begin{pmatrix} 0.794 & 0.206 \\ 0.085 & 0.915 \end{pmatrix}$
σ^2	$\begin{pmatrix} 0.4 & 0.9 \end{pmatrix}$		
125	$\begin{pmatrix} 0.634 & 0.365 \\ 0.086 & 0.914 \end{pmatrix}$	$\begin{pmatrix} 0.747 & 0.253 \\ 0.159 & 0.841 \end{pmatrix}$	$\begin{pmatrix} 0.781 & 0.219 \\ 0.059 & 0.941 \end{pmatrix}$
250	$\begin{pmatrix} 0.708 & 0.292 \\ 0.077 & 0.923 \end{pmatrix}$	$\begin{pmatrix} 0.653 & 0.347 \\ 0.058 & 0.942 \end{pmatrix}$	$\begin{pmatrix} 0.736 & 0.264 \\ 0.09 & 0.91 \end{pmatrix}$
500	$\begin{pmatrix} 0.786 & 0.214 \\ 0.077 & 0.923 \end{pmatrix}$	$\begin{pmatrix} 0.705 & 0.295 \\ 0.076 & 0.924 \end{pmatrix}$	$\begin{pmatrix} 0.78 & 0.22 \\ 0.083 & 0.917 \end{pmatrix}$
σ^2	$\begin{pmatrix} 0.9 & 1.6 \end{pmatrix}$		
125	$\begin{pmatrix} 0.75 & 0.25 \\ 0.175 & 0.825 \end{pmatrix}$	$\begin{pmatrix} 0.842 & 0.158 \\ 0.095 & 0.905 \end{pmatrix}$	$\begin{pmatrix} 0.726 & 0.274 \\ 0.161 & 0.839 \end{pmatrix}$

Продовження на наступній сторінці

Табл. Б.1: Знайдені значення матриці A (продовження)

Істинна A	$\begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.9 \end{pmatrix}$		
T	iters		
	500	1000	2500
σ^2	$\begin{pmatrix} 0.9 & 1.6 \end{pmatrix}$		
250	$\begin{pmatrix} 0.509 & 0.491 \\ 0.164 & 0.836 \end{pmatrix}$	$\begin{pmatrix} 0.801 & 0.199 \\ 0.077 & 0.923 \end{pmatrix}$	$\begin{pmatrix} 0.714 & 0.286 \\ 0.209 & 0.791 \end{pmatrix}$
500	$\begin{pmatrix} 0.726 & 0.274 \\ 0.124 & 0.876 \end{pmatrix}$	$\begin{pmatrix} 0.604 & 0.396 \\ 0.131 & 0.869 \end{pmatrix}$	$\begin{pmatrix} 0.748 & 0.252 \\ 0.134 & 0.866 \end{pmatrix}$